DAGSTUHL
REPORTS

Volume 12, Issue 9, September 2022

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
https://www.dagstuhl.de/dagpub/2192-5283

Publication date
April, 2023

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://dnb.d-nb.de.

Digital Object Identifier: 10.4230/DagRep.12.9.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:
- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
https://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 22361

Challenges and Opportunities of Democracy in the Digital Society

Abraham Bernstein[*1], **Anita Gohdes**[*2], **Cristina Sarasua**[†3], **Steffen Staab**[*4], **and Beth Simone Noveck**[*5]

1 **Universität Zürich, CH.** ernstein@ifi.uzh.ch
2 **Hertie School – Berlin, DE.** gohdes@hertie-school.org
3 **Universität Zürich, CH.** sarasua@ifi.uzh.ch
4 **University of Stuttgart, DE.** steffen.staab@ipvs.uni-stuttgart.de
5 **New York University, US.** noveck@thegovlab.org

Abstract

Digital technologies amplify and change societal processes. So far, society and intellectuals have painted two extremes of viewing the effects of the digital transformation on democratic life. While the early 2000s to mid-2010s declared the "liberating" aspects of digital technology, the post-Brexit events and the 2016 US elections have emphasized the "dark side" of the digital revolution. Now, explicit effort is needed to go beyond tech saviorism or doom scenarios.

To this end, we organized the Dagstuhl Seminar 22361 "Challenges and Opportunities of Democracy in the Digital Society" to discuss the future of digital democracy.

This report presents a summary of the seminar, which took place in Dagstuhl in September 2022. The seminar attracted scientific scholars from various disciplines, including political science, computer science, jurisprudence, and communication science, as well as civic technology practitioners.

Seminar September 4–9, 2022 – http://www.dagstuhl.de/22361

2012 ACM Subject Classification Computing methodologies; Human-centered computing; Information systems; Applied computing → Law; Social and professional topics → Political speech

Keywords and phrases co-design, democratic regulation, large-scale decision-making, large-scale deliberation, society

Digital Object Identifier 10.4230/DagRep.12.9.1

1 Executive Summary

Abraham Bernstein (Unversity of Zurich, CH)
Anita Gohdes (Hertie School, DE)
Cristina Sarasua (Unversity of Zurich, CH)
Steffen Staab (University of Stuttgart, DE)

In this Dagstuhl Seminar, we aimed to have interdisciplinary discussions on the *challenges* and *opportunities* of online platforms, online participation, and online deliberation, including experts in politics, law, technology, governance, and policy-making.

[*] Editor / Organizer
[†] Editorial Assistant / Collector

In order to achieve a thorough integration of perspectives, we started the first day of the seminar with several keynote talks by scholars from political science, computer science, communication and law. The keynote speakers were Fabrizio Gilardi (Universität Zürich), Anna de Liddo (Open University), Pablo Aragón (Wikimedia Foundation), Eleni Kyza (Cyprus University of Technology), and Felix Uhlmann (Universität Zürich). After these talks, the seminar organized a brainstorming session to identify key discussion topics related to democracy in the digital society. Based on these discussion topics, the participants worked on six breakout sessions: *Goals*, *Actors*, *Narratives and Bias*, *Structure*, *Technology*, and *Success Metrics*. Additionally, throughout the seminar, Markus Brill (Technical University of Berlin), Abraham Bernstein (Universität Zürich), Róbert Bjarnason (Citizens Foundation), Gefion Thürmer (King's College London), Gianluca Demartini (University of Queensland), and Harith Alani (Open University) gave short presentations on various topics including computational social choice, diversity in news recommender systems, citizen science, and misinformation.

The remainder of this report provides the abstracts of the talks and the group discussions.

2 Table of Contents

3 Overview of Keynote Talks

3.1 Problem Definition in the Digital Democracy

Fabrizio Gilardi (Universität Zürich, CH)

Digital technology is widely perceived to cause important problems, such as fake news, hate speech, and political polarization, which call for policy responses. However, there is no consensus on the specific nature or intensity of those problems nor, therefore, what kinds of actions would be appropriate. On the one hand, there is nothing special about this lack of agreement. The contestation of the nature of problems and solutions is a key feature of politics, in any area. On the other hand, the tension is particularly significant in the area of digital technology: policy-makers often struggle to fully understand the issues, and problem definition is subject to a high degree of political contestation enabled by digital technology itself. The talk discussed these questions and illustrated them through three specific analyses: (1) the emergence of content moderation as a political issue, (2) the effects of decentralised social media on user sharing behavior, and (3) the role of media coverage for platform policy change.

3.2 Harnessing the Power of Constructive Disagreement to Enable Healthier Public Deliberation

Anna De Liddo (The Open University – Milton Keynes, GB)

There are no simple solutions to complex societal challenges. Whether it is climate change or dealing with the devastating impacts of the coronavirus pandemic, the questions these problems pose to humanity have no single correct answer. Addressing them requires the collaboration of governments, charities, businesses and individuals. However, at a time when society seems to be dominated by dogma and discord, building consensus on what action to take may seem like the biggest hurdle to overcome. We cannot overlook the role of the Internet in fomenting divisions. Fake news and social media bubbles filter our reality and have the power to entrench us on one side of the argument, preventing us from understanding the views of others. However, research on Collective Intelligence also notes that technology can be a powerful tool to help us find common ground, even in cases where it seems we could not be further apart.

For more than a decade, I have argued for a new kind of collective intelligence, elsewhere referred to as Contested Collective Intelligence (CCI), which is mediated by new technologies for dialogue and argumentation and specifically aims to help people make sense of and co-create innovative solutions to complicated challenges. The CCI tools that we have developed harness the power of technology to enable people around the world to construct shared understandings, even when, at first glance, they disagree. Combining advanced computational methods, such as natural language processing and machine learning, with intuitive multimedia interactions, these user-friendly tools harness and structure online conversations to identify stated and unstated points of agreement within a discussion group and in this way help the group better understand and address complex problems.

In this talk, I provided a brief overview of Contested Collective Intelligence research and presented two of our CCI tools: bcause.app and democraticreflection.org.

3.3 Decidim: Technopolitical Networks for Participatory Democracy

Pablo Aragón (Wikimedia Foundation – Barcelona, ES)

strategicDecidim (`decidim.org`) is a digital platform for participatory democracy, built entirely and collaboratively as free open source software. More specifically, Decidim is a web environment that allows anyone to create and configure a technopolitical network. The platform can be deployed by any organization (local/regional/national governments, universities, nonprofits) to host large-scale citizen participatory processes for strategic planning, participatory budgeting, public consultations and collaborative policy-making. The project was launched in 2017 in Barcelona and, 5 years later, there are hundreds of active instances around the world. In this talk, key lessons from research and practice with Decidim were shared, including the impact of its deliberative platform design and the technopolitical principles that guide the participatory development of the project.

3.4 Challenges and Opportunities of E-democracy from the Perspective of Communication Studies

Eleni Kyza (Cyprus University of Technology – Limassol, CY)

This talk was organized in three parts. The first part revolved around communication studies and the calls for a need to re-define core definitions and operationalizations of communication studies, so that the pivotal role of AI technologies is acknowledged and examined. Towards this end, there is increasing discussion about examining the agentic role of AI, in addition to the traditional anthropocentric conceptualization of the study of communication. In the second part of my talk, I presented some examples from a recently concluded work from the Horizon 2020 project Co-Inform (Co-Creating Misinformation Resilient Societies, proposal 770302). Our focus in this work was on investigating how a co-created browser plugin influenced citizens' perception of misinformation and their subsequent actions. As part of this talk, I also briefly discussed the media and information literacy implications of such work for learning, education, and the design of such interventions. I concluded the talk with a summary of open areas of inquiry, informed by the Co-Inform work on how to combat misinformation on social media.

3.5 What is Democracy from a Legal Perspective and What Can Computers Do for It – Again, from a Legal Perspective?

Felix Uhlmann (Universität Zürich, CH)

Democracy is more than a decision by the majority. It is also the rule of law, encompassing fundamental rights such as freedom of expression, property, access to courts etc. on the individual level and institutions with defined competencies such as parliament, government etc. on the institutional level. Democracy is also a process. The involvement of citizens in the legislative process, the transparency of the debate as well as on financing are essential for the functioning of a modern democracy. Switzerland has opted for quite a radical system allowing popular initiatives to amend more or less any article of the Swiss constitution.

Access to large quantities of data, artificial intelligence or both elements combined may fundamentally influence democracy. They may enhance the consultation process initiated from the authorities as well as bring together like-minded people to draft and launch a popular initiative. These possibilities are still unexplored both by private and state actors.

4 Overview of Short Talks

4.1 Computational Social Choice and Digital Democracy

Markus Brill (TU Berlin, DE)

The successful design of digital democracy systems presents a multidisciplinary research challenge. In this short presentation, I explained what computational social choice is and I argued why tools and techniques from this field are relevant for the design of online decision-making platforms and other digital democracy systems.

4.2 Escaping the Echo Chamber: The Quest for the Normative News Recommender Systems

Abraham Bernstein (Universität Zürich, CH)

Recommender systems and social networks are often faulted to be the cause for creating Echo Chambers – settings where people mostly encounter news that match their preferences or those that are popular among similar users, resulting in their isolation inside familiar but insulated information silos. Echo chambers, in turn, have been attributed to be one cause for the polarization of society, which leads to the increased difficulty to promote tolerance, build consensus, and forge compromises.

To escape these echo chambers, we propose to change the focus of recommender systems from optimizing prediction accuracy only to considering measures for social cohesion. The talk also succinctly presented some results from an empirical study investigating if such a recommender system would actually have the desired results (see [1] for details).

References

1 Lucien Heitz, Juliane A. Lischka, Alena Birrer, Bibek Paudel, Suzanne Tolmeijer, Laura Laugwitz, and Abraham Bernstein. Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism*, pages 1–21, 2022.

4.3 Citizens Foundation. Connecting Governments & Citizens

Róbert Bjarnason (Citizens Foundation Iceland – Reykjavik, IS)

Many believe technology has undermined our faith in debate online; instead, our work shows non-profit innovations in technology restoring trust in democratic deliberation and institutions. Our partners, like Reykjavík city, the State of New Jersey, the Scottish Parliament, and World Bank, have used our solutions, making better decisions in thousands of projects in 45 countries since 2008. Your Priorities offers open-source idea generation and deliberation. Connecting governments and citizens by bringing people together to debate and prioritize innovative ideas to improve their communities. The Better Reykjavik project was started in 2010 and has now become institutionalized in Reykjavik. It is an example of mass online community participation with 70,000 citizens engaging out of a population of 120,000; over 40,000 registered users submitted 11,000 ideas and 25,000 debate points. Another example is the Scottish Parliament using Your Priorities to engage with citizens in Scotland. The challenge addressed is that the Scottish Parliament committees must better understand the needs of Scottish citizens concerning various subjects.

4.4 Addressing Misinformation through Innovation, Arts, and Citizen Science

Gefion Thürmer (King's College London, GB)

Misinformation can be addressed in different ways. The MediaFutures project does so by engaging with start-ups and artists (separately and together) who create products and artworks that improve fact-checking, increase media literacy, or raise awareness in society. The Action and Impetus projects support citizen science initiatives. They seek ways to engage citizens in the entire scientific process, from asking questions through collecting data to confronting policy makers with their results – which in turn raises scientific literacy and awareness.

4.5 The Anatomy of an AI System for Misinformation Detection, and Where Humans Fit in It

Gianluca Demartini (The University of Queensland, AU)

Information warfare instruments have recently been used to weaponize misinformation to foster propaganda and to reach political goals by influencing populations at scale. In this talk, we discussed how human-in-the-loop AI technology can support expert fact-checking efforts that have been increasing substantially due to the rise in the spread of misinformation. We first described the general fact-checking process and then discuss at which steps AI and humans can help. We looked at how Twitter has been crowdsourcing fact-checking, as an example of fact-checking on social media. Finally, we reflected on the human bias dimension in fact-checking, and at how the concept of truth may change over time and over different definitions of truthfulness.

4.6 Have You Been Misinformed?

Harith Alani (The Open University – Milton Keynes, GB)

As long as there has been information, there has been misinformation. During the last few years, a lot of attention has been paid to developing tools that can detect which information is reliable and which is likely to be fake or misinforming. However, we are still learning how, when, and where such advanced technologies or the work of fact-checkers around the world can help in stopping misinformation from spreading. My goal in this talk was to demonstrate that we also hold false or unreliable beliefs and argue that we need technologies that can assess the information we and others share over time. Additionally, I discussed the benefits, challenges, and risks of using automated methods for correcting people when they share misinformation.

5 Working Groups

5.1 Goals

Lynda Hardman (CWI – Amsterdam, NL & Utrecht University, NL) and Abraham Bernstein (Universität Zürich, CH)

We discussed possible audiences for publications inspired by the seminar. These include policymakers, funding agencies, research evaluators, academics, citizens (in general) and activists.[1] We recommended that addressing policymakers and academics would have the higher priority.

The goals we discussed were aimed at the stages of the democratic process for which (computational) support could be provided. At a high level, the goals we identified were:

1. Citizens
   - Inform, as a basis for forming an opinion
   - Ensure that citizens (and other actors) have appropriate and verifiable information for their decision making
   - Deliberate, as a basis for forming an opinion
   - Ensure that citizens have sufficient opportunities to expose themselves to different opinions and deliberations (incentives for) engagement (with both politicians and administrations/civil servants)
2. Procedural/Institutional
   - Improve the democratic processes through appropriate technology
   - Protect individuals from repercussions due to their political activity

In addition to developing the goals of technological support for democratic processes, we identified the common goals with a complementary initiative: the "Vienna Manifesto on Digital Humanism".[2]

5.2 Actors

Marco Steenbergen (Universität Zürich, CH), Fynn Bachmann (Universität Zürich, CH), Eleni Kyza (Cyprus University of Technology – Limassol, CY) and Cristina Sarasua (Universität Zürich, CH)

With floundering trust in politics and declining electoral participation rates in many of the world's democracies, the call for efforts to re-engage citizens grows ever louder. Over the past two decades, this has resulted in a large number of different participatory formats, including participatory budgeting, citizen juries, and minipublics. In nearly all cases, the number of participants in those formats has been limited. Digitalization offers new opportunities to create large-scale deliberations and co-creation projects for citizens. Scale is important because public acceptance of new participatory forms is enhanced when large segments of the public are represented. There have already been excellent experiments with large-scale technnology-aided deliberation (see the keynote by Pablo Aragón).

Just because one builds platforms, however, does not mean that all problems of participation are resolved. It is very clear that unequal participation from different groups in society remains an urgent problem. These problems start with recruitment but do not end there. Citizens may not be active participants, they may not be heard because they are less well equipped expressing their views, and they may ultimately drop out, perhaps more frustrated than they started. A major challenge in the digital age is to engage citizens, to retain them, and to ensure that they find an effective voice in deliberative and co-creative platforms. We

[1] Representatives of organisations active in promoting higher citizen involvement in democratic decision-making had been invited to the seminar but were unfortunately unable to join.
[2] Vienna Manifesto on Digital Humanism `https://dighum.ec.tuwien.ac.at/dighum-manifesto/`

think here of automatic moderation, which detects when certain individuals have fallen silent or have made inputs that have gone unanswered. We can also think of assistive technologies that help citizens sort and formulate their ideas and arguments in order to enhance their impact.

While citizens are key actors in democratic society, we cannot lose sight of other actors, including political parties, representatives, and bureaucrats. We still find that parties and representatives have a tenuous grip of digital technology. Opportunities are missed and risks under-estimated. Enhancing digital skills within the representative and executive institutions is essential if democracy is to flourish in the digital age. Here, the principles laid out in the goals should be a guiding light.

5.3 Narratives and Bias

Fabrizio Gilardi (Universität Zürich, CH), Anita Gohdes (Hertie School of Governance – Berlin, DE), Farane Jalali (Max Planck Institute for Informatics (MPI), DE), Jörn Lamla (University of Kassel, DE), Catarina Pereira (Universität Zürich, CH) and Miklovana Tuci (Universität Zürich, CH)

The group discussed the relevance of narratives about digital technology and democracy, and how they are lined to perceived dangers such as misinformation, echo chambers, and the reproduction of bias through algorithmic decision-making. Importantly, there may be a significant mismatch between the discourse around those dangers (e.g. in the media) and their actual importance. An example is "fake news," which were widely discussed as a main driver of political events such as Brexit and the election of Donald Trump. Verification of those claims by independent researchers was initially very difficult due to problems of access to social media data, which platforms (in particular Facebook) obstructed with arguments linked to privacy and data protection. When reliable research findings became available, they tended to disconfirm many of the ideas around fake news. For example, several independent studies, using different data and methods, concluded that the consumption and sharing of fake news is – on average – very low. Instead, they are concentrated among a specific subgroup of people who tend to be older and very conservative. While these findings do not imply that fake news are not a problem, they point to different policy responses than the initial narrative did.

Against this background, the group discussed specific issues driving the power of narratives about digital technology and democracy and inhibiting research that could challenge them, namely, the ground truth problem. The ground truth problem means that agreeing on categories (e.g., what is "fake news"?) is inherently contested, which makes it very difficult to monitor them. Moreover, there is often a denominator problem, that is, the lack of a benchmark to assess the prevalence of behaviors. By contrast, the media have no trouble finding examples: there is a lot of everything on the internet. Furthermore, both ground truth and denominators are moving targets, because of potentially rapid change in underlying problems (e.g., COVID and the Ukraine war) as well as the relevance and user base of different platforms (e.g., the rise of TikTok).

The group discussion concluded with a brainstorming session regarding possible solutions as broader recommendations.

5.4 Structures in Digital Democracy

Steffen Staab (University of Stuttgart, DE), Markus Brill (TU Berlin, DE), Martin Emmer (FU Berlin, DE), Jörn Lamla (University of Kassel, DE), Libor Pavlícek (Charles University – Prague, CZ), and Gefion Thürmer (King's College London, GB)

A common way to break down the policy life cycle results in the following five stages:
1. Agenda setting
2. Policy formulation
3. Decision making
4. Policy implementation
5. Policy evaluation

Usually, this process is a loop, as after policy evaluation the policy may be terminated and a new agenda might be set or discussion about the existing policy may continue. The reader may note that the process does not need to be strictly sequential and stages may be skipped or repeated.

The working group discussed whether and how the digitalization of democracy means changing these existing stages by rethinking and restructuring process steps, or whether the overall process is stable and the speed and the agility of the process change. We addressed this question by discussing the past, the present and the desired future execution of policy life cycle stages as follows:

Considering stage 1, digitalization has profoundly changed the means for agenda setting. In the past, media gatekeepers have hugely influenced agenda setting and were not easily swayed by activists. Nowadays, traditional media still play a major role, but social media enlarge the set of voices that are heard, which may help minorities, but which increasingly often leads to information overload of the public. As a consequence, few topics are high on the agenda and many lower ranked topics tend to be forgotten. Specific tools like Decidim may help to spread the valuation of importance more evenly among a broader range of topics and, hence, their regular use should help in the future. A core objective for the future would be to "make uncertainty productive". Current discussions often lead to overly simplistic solutions, where only few aspects are considered because of information overload. If future tools and processes could help to manage a multiplicity of uncertain dimensions, agenda items could be managed in ways that would improve the joint satisfaction of multiple goals.

Considering stage 2, in the past policy formulation could be characterized by tight control by executive and legislative powers who oversaw control of relevant knowledge. Political parties and lobby groups were highly active and only in a minority of cases activism and street protests would contribute. Today, there is a tendency towards increased transparency. Social media-supported activism has led to new kinds of protests that increased the number of actors, while political parties exhibit declining numbers of members. Because of digitalization, knowledge can and is more broadly shared allowing the broad public to know about details of formulated policies.

With regard to the future, we argue in favor of increased transparency about lobbyism and specifically transparency of the policy formulation process. Because of increased complexity of regulations, it is not only important what is written, but who wrote it. Overseeing implications of policies seems only possible by stronger collaborations between politicians and professionals in order to achieve evidence-based policy formulation. Many questions remain open about such co-design of policies: who has access? Who has the right to be heard by their representative? How is consultation fed back to the public commenters?

Considering stage 3, decision-making may change more slowly than other aspects of the policy life cycle. Thereby, the success of novel means of e-voting or liquid democracy does not only depend on the process, but on the assumptions underlying this process. When voting is strictly confidential (as it is in Germany), there does not seem to exist a technology that makes e-voting as sound proof as traditional voting with paper ballots. This is different, if votes can be made public, because then it becomes easy to monitor the soundness of an electronic voting process.

Considering stage 4, in the past policy implementation focused on guiding people by giving or taking money and imposing court orders. Today, we observe a digital turn in connection with a behavioral turn trying to nudge people using behavioral politics and behavioral economics or appealing to citizens as an alternative or complementary means to imposing laws.

For the future, we would wish for governmental processes to be streamlined by digitalization for efficiency and effectiveness as well as an administration that seamlessly serves the citizens rather than imposing heavy administrative burdens on citizens and organizations. A key ingredient to such support may be the real-time observation of individuals and organizations, which, however, bears the dangers of ubiquitous surveillance and manipulation, as they have become obvious in experiments with social credit systems. It remains an open question how to take advantage of benefits while avoiding these heavy drawbacks – and whether this is possible at all.

Little progress has been seen regarding stage 5, policy evaluation. In the future, we might perhaps see more comprehensive consultation and feedback by the public. It would be desirable to formulate key performance indicators that can be reported automatically already during policy formulation. Also, open data might help to judge appropriateness of policies.

Finally, these structures need to be understood and elaborated in terms of overarching or meta perspectives of these process stages. As illustrated by the overarching objective of (non-)anonymity of voting, these meta perspectives may deeply affect the working of the digitalization-enhanced policy life cycle.

5.5 Technology: Challenges and Opportunities of Technology for Democracy

Anna De Liddo (The Open University – Milton Keynes, GB), Harith Alani (The Open University – Milton Keynes, GB) and Pablo Aragón (Wikimedia Foundation – Barcelona, ES)

Within the technology group, we first examined what technology can do for democracy. We tried to ask ourselves which democratic challenges we think are particularly well suited to be addressed by a technology-mediated approach, and we asked what are the main challenges and opportunities we see emerging in the near future in the field of technology research for democracy. These challenges and opportunities are summarised below. We propose them as key points of focus for the fields of computer, social and political sciences interested in the study, design and application of technology for democracy.

5.5.1 Challenges

1. **Fair and inclusive technology design**. How can we make fairness and inclusion the key values not only behind technology design but also behind tech governance, regulation and applications?
2. **Testing, experimentation and evaluation at scale and in the wild**. Advancing scientific knowledge on technology for democracy requires solid testing in, often, controlled environments. But democracy experiments are often very costly or impossible to carry out in the wild. How can we enable testing, experimentation and evaluation at different (smaller) scales that leads to solid scientific insights?
3. **Online social dynamics and influence of existing digital practices**. In a blended socio-technical system as our society is nowadays, digital practice have proved to radically influence real world practice, organisations and even power. On that account, it is a fundamental question to ask: How can we study the influence of the digital world on the analog world and vice versa? How do we account for, monitor, and counteract negative influences, while leveraging the positive ones?
4. **New economic models for neutral/open sustainability**. Digitally mediated democratic research, technologies and experiments need political endorsement and financial support to become self sustainable and survive the research project lifespan. But such support and endorsement can hardly come from economic systems of models which brake fundamental values such as equity, inclusion, fairness and democracy. Hence a key challenge for this research to thrive is: what Economic models, based on democratic values, can be designed and devised to support, enable research in technology for democracy?

5.5.2 Opportunities

5. **Re-engage people in public and civic life**. Citizen disengagement from politics and public life is one of the fundamental reasons for the systemic failure of our democracy (see, for example, the result of the last Italian general elections, which saw only slightly more than 50 percent of Italians go to the polls). The public has difficulty relating to politics and often believes that it is only a context of victories and defeats that cannot be influenced by individual action (much less individual voting). New technologies can be used to make politics more interesting, fairer, more engaging, and even more fun. Improving participation can be the first step and the way to solve the dangerous problem of political disempowerment.
6. **Improve minority representation in public choices**. New technologies to improve representation in decision-making contexts can be key to enabling minority voices to be represented in democratic realities.
7. **Improve trust in politics by enhancing fairness, transparency and accountability**. Citizens are increasingly concerned about transparency and accountability in the functioning of any democratic institution, but at the same time they have become increasingly aware of the issue of data privacy. New decentralized technologies and distributed ledger systems can provide transparency without endangering individual privacy.
8. **Improve sharing of resources and expertise across geographical barriers**. The Open Science, Open Data, and Open Education movements have demonstrated that technology can improve access to science, information, and education for communities around the world that lack the infrastructure, money, or human resources to obtain them. These technologies are inherently suited to democratize access to knowledge across geographic, economic, and social barriers.

9. **Informing new democratic models, social justice and redistribution of power**. Successful examples of digital democracy projects have shown that technology can be used to disrupt power, protest and mobilize the masses. These examples show the opportunities that technology can offer to bring democracy to places where it is not taken for granted.
10. **Increasing the reach of participation**. Social media have shown that ICT tools can radically broaden participation in many social processes such as commerce, education, work and socialization. Building on this potential, we ask: What does new technology need to grow democracy?

5.5.3 Definition of Technologies in/for Democracy? (Technology as a Medium and Actor)

What are democratic technologies? We have been reflecting on the difference between "Technologies in Democracy", that is, technologies used to promote, mediate, or participate in democratic processes in general, and "Technologies for Democracy", that is, tools designed, repurposed, or adapted specifically to improve democratic processes. "Technologies for democracy" are intended to influence and change the way democratic processes take place. This distinction is useful for classifying different types of technologies and studying the contexts in which they are applied. We examined the current classification of democratic technologies in the HCI and CSCW fields and found the focus on civic technologies as tools for enhancing democratic participation particularly inspiring. Civic technology has been defined by Saldivar et al. "as technology (primarily information technology) that facilitates democratic governance among citizens." [1] Government- and citizen-centered definitions of civic technology for democratic participation are useful in focusing attention on the role of governments in democratic technology research.

A government-centered view of civic technologies, for example, would include all ICT tools used "by cities for service delivery, civic engagement, and data analysis to inform decision-making (Living Cities 2012)" (Saldivar et al. 2019). These could be ordinary social media or data collection and integration services. While a citizen-centric definition of civic tech presents it as "platforms and applications that enable citizens to connect and collaborate with each other and with government [2]" [1].

Both definitions mention government involvement, but the second only tangentially. This means that democratic technology research also takes place outside existing government structures and institutions. A citizen-centered view of democratic participation tools looks at democratic governance rather than formal democratic institutions, and thus includes a variety of democratic practices that emerge and flourish outside formal institutions.

However, an important question to ask is: To what extent should civic technologies aim to influence and change the way governance processes take place in order to be classified as democratic technologies (or technologies for democracy)?

The need for future research to clarify existing definitions and classifications of democratic technologies is identified.

5.5.4 Stories of Success, Failure and Disruption

Reflecting on the most recent stories of success, failure and disruption of technology applications in real-world democratic contexts, we have come to two main reflections, which we offer for further discussion and insight.

- **We cannot solve technology when the problem stems from the economic model**

A famous case in which technology was used to undermine (rather than improve) democracy was Cambridge Analytica's use of Facebook data to influence the 2016 U.S. election. Every effort on the part of the company (Facebook) to "fix" the problem seems to have gone in vain, as the use of social media data and the spread of disinformation is now an unstoppable social phenomenon that has become "a problem" in its own right and, most importantly, as the economic model behind the technology remains profit.[3] We ask ourselves: Do democratic technologies have a "hidden" requirement, a design value of any democratic tool, namely the need for such tools to be based on economic models that are democratic in themselves and do not pursue the profit of the few?

- **Digital democracy seems to work when technology is used to distribute power**.
  Successful digital democracy projects with democratic technology at their core seem to emerge in power structures that are open and, in some ways, conducive to changing existing democratic structures and practices. This cultural and political engagement, together with concrete public and financial support from official local institutions, seems to have been a key success factor for digital democracy projects such as "Better Reykjavik", "e-Democracia", and Decidim.Barcelona.[4] These projects have grown and sustained due to the strong support of local councils and official government institutions. These institutions have provided approval, continued funding, and a platform to proactively change democratic practices at the local level. We ask: to what extent can democratic technologies and digital democracy projects succeed without such a framework and institutional support?

5.5.5 Fields of Interest

Finally, we have attempted to list a number of issues and research questions that require urgent attention from the interdisciplinary research community on democratic technologies or technologies in/for democracy, as we conceptualized them above. The list is by no means conclusive and is intended as a source of inspiration for future research.

- Explainability/Intervention
  - Explainability: Identifying why a claim or argument that is misleading or incorrect is key to changing perception and opinion. It remains unclear how to do this effectively.
  - Measuring success/impact of intervention should not be limited to changing the opinion of the target individual. Audience matters, too.
- Collective Intelligence
  - Computational social choice can provide mechanisms for the aggregation of individual preferences.
  - In the field of deliberation, it is not clear what the best way to aggregate/summarize discussions is.
  - Visual analytics to support deliberation processes
  - Can AI improve CI and vice versa?
    * Machine-in-the-loop vs. human-in-the-loop

[3] `https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation`
[4] `https://www.theguardian.com/public-leaders-network/2017/feb/23/democracy-digital-lessons-brazil-iceland-spain`

- Misinformation
  - Multilinguality
  - Multimedia: current tech is very text focused.
  - Bias: most models are trained on misinformation in particular topics / platforms.
  - Long-term impact: tools/methods to capture and process the long term impact of collective misinformation are needed (e.g., tracking misinformation towards the EU over many years).
  - Accountability: holding politicians/influencers accountable
  - Economic disincentivization of disinformation
  - Changing policy, business model, and practice towards promoting "good" information and behavior.
- Polarization
  - Tech to (a) measure polarization on given topics, and (b) identify the main sources that are feeling this polarization (e.g., newsmedia, politicians, groups, bots?)
  - Capturing cross-polar argumentation
  - Detecting ideas/arguments bridging polarized scenarios
  - To what extent diversity and disagreement can help reduce polarization and build common ground?
- Sensemaking and critical thinking
  - Technology for slowing down
  - Slow tech for a better digital democracy [5]
  - Technology for sensemaking
  - Technology for stability vs agility
- Experimentation
  - Resource inequality (big tech vs civic tech)
  - Experimentation design challenges (costs, large-scale, in the wild)
  - Who is responsible for designing experiments?
  - Inclusiveness of experimentation and testing of relevant tech

References

1 Jorge Saldivar, Cristhian Parra, Marcelo Alcaraz, Rebeca Arteta, and Luca Cernuzzi. Civic technology for social innovation. *Comput. Support. Cooperative Work.*, 28(1-2):169–207, 2019.

2 Manik Suri. From crowd-sourcing potholes to community policing: Applying interoperability theory to analyze the expansion of 'open311'. *Berkman Center Research Publication*, (2013-18), 2013.

5.6 Success Metrics

Marco Steenbergen (Universität Zürich, CH), Harith Alani (The Open University – Milton Keynes, GB), Abraham Bernstein (Universität Zürich, CH), Cristina Sarasua (Universität Zürich, CH) and Gefion Thürmer (King's College London, GB)

Democracy serves many purposes. It lends legitimacy to authority and decisions, it solves problems, and it engages citizens, to name a few. These purposes should be reflected in our metrics. Depending on the function, however, not all metrics are equally relevant. For

[5] https://medium.com/qleek/the-slow-tech-manifesto-1b39fbcd1c48

instance, creating respectful discourse of all claims, no matter how outlandish, is important when the goal is engagement. On the other hand, it likely introduces many inefficiencies when the goal is to solve an urgent problem. All functions should find a place somewhere in democracy, but they do not necessarily have to be satisfied all at once.

When building technologies that assist democratic processes, it is important to keep an eye on function: what specific goal should this solution cater to. The measure of success is then defined in terms of that goal. This functionalist logic is increasingly found in normative theories of (deliberative) democracy.

That said, we can specify a series of general metrics, a subset of which may be crucial for a specific purpose. We list them here in several rubrics.

1. **Representation:** Is a technology capable of bringing in all of the stakeholders in a topic? Metrics include a comparison of those formally partaking to the public. Beyond this, one can also look at *active* participation and reciprocation (are all the arguments of all stakeholders heard?).
2. **Respect:** Does a technology enhance respectful interaction among citizens? In an age of affective polarization across political parties and groups, respectful discourse often seems in short supply. A minimum standard for technology is that it cools down discourse to the point that hate speech is eliminated. Higher bars can be set, however, such as the degree of perspective-taking: even if person A continues to disagree with B, can they at least understand and respect where B is coming from?
3. **Output:** Does a co-creation or deliberative process yield an output that is helpful, for instance, to policy-makers? Here, one can think of the legal quality of a proposed law or the factual accuracy of an epistemic discussion. Digital technologies should help citizens and policy-makers to reach high-quality outcomes, as judged by experts.
4. **Legitimacy:** Are the outcomes of human-in-the-loop or computer-in-the-loop deliberative and co-creative processes acceptable to those who did not participate? Metrics here include survey-based measures of acceptance.

Participants

- Harith Alani
The Open University – Milton Keynes, GB
- Pablo Aragón
Wikimedia Foundation – Barcelona, ES
- Fynn Bachmann
Universität Zürich, CH
- Abraham Bernstein
Universität Zürich, CH
- Markus Brill
TU Berlin, DE
- Anna De Liddo
The Open University – Milton Keynes, GB
- Martin Emmer
FU Berlin, DE
- Fabrizio Gilard
Universität Zürich, CH
- Anita Gohdes
Hertie School of Governance – Berlin, DE
- Lynda Hardman
CWI – Amsterdam, NL & Utrecht University, NL
- Farane Jalali
Max Planck Institute for Informatics (MPI), DE
- Eleni Kyza
Cyprus University of Technology – Limassol, CY
- Jörn Lamla
University of Kassel, DE
- Libor Pavlícek
Charles University – Prague, CZ
- Catarina Pereira
Universität Zürich, CH
- Cristina Sarasua
Universität Zürich, CH
- Steffen Staab
University of Stuttgart, DE
- Marco Steenbergen
Universität Zürich, CH
- Gefion Thürmer
King's College London, GB
- Miklovana Tuci
Universität Zürich, CH
- Felix Uhlmann
Universität Zürich, CH

Remote Participants

- Róbert Bjarnason
Citizens Foundation Iceland – Reykjavik, IS
- Soon Ae Chun
CUNY College of Staten Island, US
- Gianluca Demartini
The University of Queensland – Brisbane, AU
- Karsten Donnay
Universität Zürich, CH
- Andreas Jungherr
University of Bamberg, DE
- Valeria Vuk
Universität Zürich, CH

Report from Dagstuhl Seminar 22362

Model-Driven Engineering of Digital Twins

Loek Cleophas[1], Thomas Godfrey[2], Djamel Eddine Khelladi[3], Daniel Lehner[4], Benoit Combemale*[5], Bernhard Rumpe*[6], and Steffen Zschaler*[7]

1 **Eindhoven University of Technology, NL & Stellenbosch University, SA.** l.g.w.a.cleophas@tue.nl
2 **King's College London, GB.** thomas.1.godfrey@kcl.ac.uk
3 **CNRS, IRISA, Univ. Rennes, FR.** djamel-eddine.khelladi@irisa.fr
4 **Johannes Kepler Universität, Linz, AT.** daniel.lehner@jku.at
5 **University & IRISA – Rennes, FR.** benoit.combemale@irisa.fr
6 **RWTH Aachen, DE.** rumpe@se-rwth.de
7 **King's College London, GB.** steffen.zschaler@kcl.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22362 "Model-Driven Engineering of Digital Twins".

Digital twins are an emerging concept with the potential for revolutionising the way we interact with the physical world. Digital twins can be used for improved analysis and understanding of complex systems as well as for control and transformation of these systems. Digital twins are themselves complex software systems, posing novel software-engineering challenges, which have so far not been sufficiently addressed by the software-engineering research community.

The seminar aimed as a key outcome to contribute to a solid research roadmap for the new Software Engineering subdiscipline of Model-Based Development of Digital Twins. This paper is an intermediate result, which is thought to be further discussed in the research community that has also been built using this seminar.

Seminar September 4–9, 2022 – http://www.dagstuhl.de/22362

2012 ACM Subject Classification Software and its engineering → Software notations and tools

Keywords and phrases Software Engineering, Model-Driven Engineering, Digital Twins, Model Management, Data Management, Models@runtime

Digital Object Identifier 10.4230/DagRep.12.9.20

* Editor / Organizer

Model-Driven Engineering of Digital Twins, *Dagstuhl Reports*, Vol. 12, Issue 9, pp. 20–40

Editors: L. Cleophas, T. Godfrey, D. Khelladi, D. Lehner, B. Combemale, B. Rumpe, and S. Zschaler

Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Summary

Loek Cleophas
Thomas Godfrey
Djamel Eddine Khelladi
Daniel Lehner
Benoit Combemale
Bernhard Rumpe
Steffen Zschaler

Digital twins are an emerging concept with the potential for revolutionising the way we interact with the physical world. Early versions of digital twins have already been applied successfully in almost all known areas, including engineering and development areas, but also scientific domains and cultural, social, and economic domains.

Digital twins are leveraging the digitalization of increasingly more application domains and are intensively using various models of different types in a descriptive, predictive, and prescriptive way. Furthermore, the the ever-increasing availability of data, the last improvements of sensor technologies, and the reliable connnectivity enable direct inspection and manipulation of real-world systems, both for physical systems and objects as well as social systems and organisations respectively their processes.

The concept has seen strong interest in industry, where there is a desire to control increasingly complex systems of systems, ensuring they behave as expected and to control their adaptation to the environment or any deviations with the initial plan. Digital twins can be used for improved analysis and understanding of complex systems (in silico experimentation) as well as for control and transformation of these systems. Digital twins are themselves complex software systems, posing novel software-engineering challenges, which have so far not been sufficiently addressed by the software-engineering research community.

There is a need for solid foundations to ensure the development of tools and methods according to well-established principles. We believe that Model-Driven Engineering (MDE), will be a key technology for the successful systematic engineering of Digital Twins. In this Dagstuhl Seminar, the goal was to bring together both practitioners and researchers to

- (i) reflect on the concept of Digital Twins and the software-engineering challenges posed,
- (ii) identify relevant existing MDE approaches and technologies that can help tackle the challenge of systematically engineering digital twins, and
- (iii) define an academia–industry research roadmap for systematic engineering of digital twins based on MDE.

As the intended primary goal of the seminar is to create a community and establish a research roadmap, we have been discussing the following topics:

- Challenges faced in real-world development of Digital Twins.
- Opportunities offered by MDE.
- Active exploration of collaboration opportunities.

The following paper reflects the discussions and some of the outcomes, however, we also identified that the overall topic is not only relevant, but also highly innovative, which is why this paper does only reflect an intermediate status of discussions and results, but the community will vividly go on to solve the challenges identified and addressed in the rest of this paper.

One key outcome of the seminar and its continuing community activities will be to contribute to a solid research roadmap for the new Software Engineering sub-discipline of Model-Based Development of Digital Twins.

Definitions to set the stage

There are two core terms that need appropriate definitions, namely model driven engineering (MDE) and digital twin. While MDE (and with it the terms model and modeling language) are relatively straightforward, there are different variants of definitions for digital twins that served as a starting base.

▶ **Definition 1** (Model Driven Engineering (MDE)). *Model Driven Engineering* is a state-of-art software engineering approach for supporting the increasingly complex construction and maintenance of large-scale systems [4, 3, 2]. In particular, MDE allows domain experts, architects, developers to build languages and their tools that play an important role in all phases of the development process [7].

As digital twins are currently a relatively young and in particular evolving area of research, it is not surprising, that there is a variety of different definitions available. (E.g. [1] identifies more than hundred different definitions).

At the beginning and during the seminar we identified the following definitions to be particularly of interest.

▶ **Definition 2** (Digital Twin (DT)). A *digital twin* (DT) is a comprehensive digital representation of an actual system, service or product (the Physical Twin, PT), synchronized at a specified frequency and fidelity [5]. The digital twin includes the properties, conditions and behavior of the physical entity through models and data, and is continuously updated with real-time system data [6]. The exchange of data between the digital and the physical twins takes place through bidirectional data connections.

An alternative definition can be found at the website of the software engineering institute of RWTH Aachen University[1] together with some additional discussions:

▶ **Definition 3** (Digital Twin (DT)). A *digital twin* of a system consists of

- a set of models of the system and
- a set of digital shadows, both of which are purposefully updated on a regular basis,
- provides a set of services to use both purposefully with respect to the original system, and
- can send
  - information about the environment and
  - control commands to the original system.

▶ **Definition 4** (Digital Twin System (DTS)). Based on these definitions, a *Digital Twin System* can be defined as the combination of an actual system and the DTs of this actual system.

[1] `https://www.se-rwth.de/essay/Digital-Twin-Definition/`

2 Table of Contents

Timeslot	Monday	Tuesday	Rest of the week
09:00-10:00	Opening and Seminar Objectives Participant introductions	Presentations: Futuristic scenarios for DT use 10+5min each	Plenary session
10:00-10:30	Coffee Break		
10:30-12:15	Presentation and discussion: Digital Twin Terminology	Presentations: Futuristic scenarios for DT use 10+5min each	Groups discussion
12:15-13:30	Lunch		
13:30-15:30	Presentations: Existing DTs in different contexts 10+5min each	Discussion: “Requirements elicitation”: what do we need to build to create a full DT?	Groups discussion
15:30-16:00	Coffee Break		
16:00-17:30	Discussion: Properties of existing DTs, commonalities, differences, challenges	Discussion: Planning for rest of the week (schedule and breaking down)	Groups discussion
18:00-19:15	Dinner		
19:30-20:30	Discussion: Planning for rest of the week (main objectives)	*Organizer's meeting*	Presentations from participants and debrief
Evenings	Time for demonstrations, discussions, pool, …		

Figure 1 Schedule of the seminar.

3 Schedule of the seminar and presentations

This section presents the schedule of the dagstuhl week and the various presentations we had during the seminar.

3.1 Schedule of the week at Dagstuhl

Figure 1 depicts the schedule of the seminar week. It started by opening the seminar week and then with presentations and discussions about the terminology, context, and properties of digital twins. We then discussed the planning of the rest of the week. In the second day, we had presentations of real world scenarios of digital twins. After that, we discussed about requirements for digital twins and finalized the planning of the rest of the week. Finally, for the rest of the week, we split in three groups to work in parallel, with a plenary session in the morning and the evening to debrief each other about what was done and will be done.

4 Overview of Talks

This section presents the various presentations we had during the seminar. It details the author and the talk's title and abstract.

4.1 Learning Digital Twin Models

Shaukat Ali (Simula Research Laboratory – Oslo, NO)

Given that operational cyber-physical systems (CPS) produce continuous data, a complementary approach to model-based engineering is to learn digital twins models with machine learning techniques and providing functionalities such as predictions and anomaly detection.

This talk will start with presenting an opinion on the next generation of digital twins (Quantum Digital Twins), where some aspects of digital twins will be implemented as quantum software and executed on quantum computers, e.g., for simulating the physical environment that can be realistically simulated with quantum-mechanical principles.

Followed by this opinion, the talk will present some recent works on learning digital twins from historical data and continuous updates of digital twins with continuous data from operational CPS. Various machine learning techniques were applied, such as generative adversarial networks, curriculum learning, and transfer learning to learn digital twins. The digital twins were built for use cases from the transportation domain and water distribution/treatment plants. These digital twins were focused on anomaly detection and waiting time predictions.

4.2 Theory building and sociotechnical digital twin: MDE Requirements

Balbir Barn (Middlesex University – London, GB)

Depending upon the flavour of digital twin (and that is an entirely different problem), a digital twin of a real world artefact is designed for a purpose. Such a purpose can can range from trying to understand an "as is" situation, analysing multiple options in a decision making scenario through to tractable transformation of the real world artefact. Underpinning all these purposes is a form of theory building and assumptions of underlying theories.

A theory in its most sparse understanding is a statement of what causes what, and why, and under what circumstances. A theory can be a contingent statement or a proven statement. We use theories all the time. Peter Naur referred to programming as form of theory construction. Decision makers in organisations use theory every day. They make decisions on some basis of cause and effect, often without being specific about their reasoning. Naturally, theories are empirical or can remain conceptual explanations.

The most widely accepted notion of a Digital Twin (DT) is one where there is a cyber-physical component. DTs of civil structures such as bridges are one such example. Such DTs rely on well understood physics based laws which provide empirically tested theories. New generations of DT belong to a socio-technical context where there is heady mix of human action, systems integration and emergent behaviour that cannot readily be assumed to follow a particular predicted route. Such Socio-technical DTs when they are constructed need input from social science, psychology and other inter-disciplinary fields. Theories for justifying concepts, relationships and rules of such a STDT are therefore obtained from these non-IT fields. How do we embed and use these theories?

The challenge for model driven engineering of STDT is therefore further attenuated and more engineering support is needed. For example, we need a modelling language that is sufficiently rich to capture working descriptions of key social science theories. We will need libraries of these theories together with working examples. We will need theory integration environments and accompanying methods that support theory building. We want to be able to run a model and view the execution of the theory and its explanation.

The benefits of being able to demonstrably prove that an established theory underpins a component of a DT specification model provides external validity through a reference bench mark of a well understood theory that has been critiqued at scale and over time. Model validity concerns are therefore mitigated. Essentially this is a research challenge and I am not yet aware of significant progress in this area with respect to Digital Twins.

4.3 Digital twins Automotive

Ion Barosan(TU Eindhoven, NL)

Advancements in the automotive industry is being highly dependent on Software technology as the industry is now changing from manual driving to different levels of autonomous driving. In order to develop autonomous driving vehicles, digital twins are now used to achieve the desired features and functionality. The approach to use digital twins saves huge resources as it accelerates product development and perform virtual simulation forecasts before production. Another use case is using the twin in tandem with the physical environment providing functionalities of monitoring, fault detection to name a few. The Automotive Technology research lab at Eindhoven University of Technology, Netherlands is performing research and development of such systems. As a part of this lab, this project focuses on design and implementation of a digital twin software framework for autonomous articulated vehicles within a distribution center. The goal is to develop such a system, that replicates 'the real world moving of autonomous trucks in a distribution center' within a Digital Twin. Additionally, the virtual system should be able to gather data from various sensors in order to enhance the development of the physical truck. This data can then be used by users to further study or visualize. In order to develop this complex system, a software system representing the digital twin called the TruckLab-DTF has been developed. Using the TruckLab-DTF product development teams working on achieving autonomy in distribution centers can apply insights from the virtual twin and physical system directly to their development. Additionally, requirements can be verified in the digital twin early in the design phase, saving time and money. Thus, the use of digital twin with this domain is multi and this project is an initial step towards the realization of much larger "system of systems." In extension, designed software can be also used as an educational technology tool for vehicle dynamics and control-based courses.

4.4 Digital twins = models@run.time + Simulations

Nelly Bencomo (Durham University, GB)

Digital twins (DT) have emerged as a promising paradigm for run-time modelling and performability prediction of cyber-physical systems (CPS) that can be applied in multiple domains. Different definitions and industrial applications of DT have materialised, going from purely visual three-dimensional models to predictive tools, many of them focusing on data-driven evaluations. We want to focus on a conceptual framework based on autonomic systems to host DT run-time models based on a structured and systematic approach. A model at run-time can be defined as an abstract representation of a system, including its structure and behaviour, which exist alongside the running system. Run-time models provide support for decision-making and reasoning based on design-time knowledge. However, they can also offer themselves as a run-time abstraction based on information that may emerge at run-time and was not foreseen before execution. New techniques based on machine learning (ML) and Bayesian inference offer great potential to support the update of run-time models during execution. Run-time models can be updated using these new techniques to provide better-informed decision-making based on evidence collected at run-time. The syncing of the real and the digital twin: Models@runt.time and the MAPE-K loop can provide the structured basis of the software architecture presented and how the required casual connection of run-time models is realised to sync the real and the digital twins. This process keeps the running system inside an envelope of good behaviour.

4.5 Digital Twin for DevOps Process Improvement

Francis Bordeleau (ETS – Montreal, CA)

DevOps emerged in the last decade as the prominent approach to increase productivity and system quality in the software industry. It advocates for automation and monitoring at all stages of software development and operations, and aims for shorter development cycles, increased frequency of deployment, and more reliable releases. Its adoption by industry leaders (e.g. Amazon, Facebook, Google, and Netflix) has resulted in spectacular progress. However, evolving/improving the software process remains a main challenge and many companies are struggling with the implementation and evolution of software processes. The lack of a systematic approach makes continuous improvement an ad hoc journey in which decisions are based on intuition rather than facts. To enable the systematic improvement of DevOps processes, we propose a digital twin approach that addresses two main challenges: 1) the continuous monitoring and measurement of DevOps process according to specific objectives to detect issues so that improvement actions can be taken; and 2) the evaluation of various modification alternatives to reach a specified DevOps process improvement objective; this allows to take decisions on a scientific basis rather than in an ad hoc manner.

4.6 Model Driven Adaptable Digital Twins

Tony Clark (Aston University – Birmingham, UK)

A series of architecture models for digital twins of increasing sophistication is presented, leading to a digital twin that adapts in order to control a real-world asset. A technology is in development that supports the construction of adaptable digital twins. An overview of a simple use of the technology is presented.

4.7 Model-Driven Engineering for Enterprise Digital Twins: Opportunities and Challenges

Benoit Combemale (IRISA, University Rennes, FR)

Enterprises are rapidly evolving as complex socio-technical systems that require continuous adaptation regarding a dynamic and uncertain environment. Various stakeholders are involved and deliver continuous knowledge through heterogeneous digital models (systems, processes, organizations, etc.) manipulated with various scientific and engineering environments. This modeling continuum open new opportunities for adaptable and efficient enterprises, but also raise new modeling challenges. In this talk, I explore the use of model-driven engineering to develop and operate enterprise digital twins. The talk covers the current state of the art, and provide concrete implementations for smart trade-off analysis and decision making. Finally, I discuss open challenges and draw a roadmap for the community.

4.8 Conceptual Modelling for Risk Modelling

Georg Grossmann (University of South Australia – Mawson Lakes, AU)

The modelling of risks and predicting the impact of risks is crucial in physical asset management. Digital Twins can support the risk modelling and prediction by providing a holistic view of the assets but there is no existing standard on how to model risks. The challenges of modeling risks comprise the support of different abstraction levels, integration of different views of risks and supporting the evolution of risk models over time through Digital Twins. We provide an overview of those challenges with examples from energy industry.

4.9 Digital Twins for Cyber-Physical Systems

Gabor Karsai (Vanderbilt University – Nashville, US)

The talk will give a very brief summary of existing Digital Twin efforts in the US, and then introduces some ideas about the specific challenges digital twins have to solve in the domain of Cyber-Physical Systems. A specific approach for addressing the model integration problem is discussed, as well as how DT-s can assist in autonomous system operations. Finally, five fundamental challenges are posed.

4.10 The MB.OS Approach: How can Complexity be Managed in a Software-Defined World?

Oliver Kopp (Mercedes-Benz AG – Stuttgart, DE)

Product complexity is ubiquitous – and it is constantly increasing. Instead of working against it, we should accept it as challenge and make complexity manageable.

Why are software methods becoming increasingly important in this context? Because that is exactly where they come in: Complexity becomes controllable by software methods. For the automotive business, this means that we must use systems engineering methods to manage product variance holistically, based on a central data model having an end-to-end scope over lifetime. To do this, we follow a data-centric approach that maps every aspect: Both in the problem space (requirements, features, regulatory specifications) including the product configurations and in the solution space (types, options, functions, components) including the corresponding type configurations.

In this talk an overview of our concept “Typebased Product Line Engineering” is presented: How can we move from a reactive, quantity-based product documentation approach to a proactive product documentation approach based on individual and concrete configurations? The role of a Digital Twin is outlined – especially in the context of a vehicle that can always be upgraded with new features via over-the-air updates throughout its lifetime in customer hands.

Acknowledgement: The talk was originally presented by Christian Seiler, Mercedes-Benz AG at the Digital Product Forum 2022. This work is partially based on the research project SofDCar (19S21002), which is funded by the German Federal Ministry for Economic Affairs and Climate Action.

4.11 OASIS TOSCA: A meta and meta-meta model for modeling and managing of structured applications

Oliver Kopp (Mercedes-Benz AG – Stuttgart, DE)

Deployment denotes the installation and the running of necessary components of an application. There are mulitple tools available to do so: Terraform nad Ansible form two examples. They all have different meta-models to capture the desired application topology (also called "infrastrucutre-as-code"). OASIS TOSCA ("Topology and Orchestration Specification for Cloud Applications") is a standard including both a meta-model and a meta-meta model to model the application topology in a standardized and vendor-neutral way. This talk outlines aspects of TOSCA and presents Eclipse Winery as one tool to model applicaton topologies using TOSCA.

Acknowledgement: This work is partially based on the research project SofDCar (19S21002), which is funded by the German Federal Ministry for Economic Affairs and Climate Action.

4.12 JabRef as a Literature Management Tool and a Software Engineering Training Tool

Oliver Kopp (Mercedes-Benz AG – Stuttgart, DE)

JabRef is a literature management software based on BibTeX written in Java. The project management is mainly done using GitHub's features. The talk will fist outlined JabRef's functionalities and then dives into open source software development. We showed how the JabRef team supports newcomers to find new issues and to craft a code contribution – and how this helps to use JabRef as teaching object for training Software Engineering on intermediate level.

4.13 Towards Dynamic Self-adaptive DT Architectures

Daniel Lehner (Johannes Kepler Universität Linz, AT)

With the emergence of Digital Twins, more and more architectures are developed to use these Digital Twins in particular context to serve a specific purpose. However, the components of such architectures are usually targeted to its particular context and purpose, with limited adaptability. This makes it hard to change an architecture once the purpose of the underlying system changes, or reuse existing components in developing new architectures. To solve this challenge, we propose a feature model of Digital Twins that enables to put together new software architectures that leverage these Digital Twins. The component-based realization of the individual features on the software and modeling language level should enable an efficient plug and use of Digital Twin architectures in the future.

4.14 Engineering/Working with Digital Twins

Bernhard Mitschang (Universität Stuttgart, DE)

From a data-centric point of view digital twins are just data structures/models that keep relevant information of real-world artifacts (machines, products, environments etc.). Thus "working with digital twins (cf.my title)" simply means adapting/changing/extending/merging/intersecting/snipping/... these structures/models. My main questions raised – and perhaps even answered during this seminar – are:

- Is this really so simple?
- What does it mean to merge two digital twin models and what is the impact and consequences thereof?
- Don't we need to consider (domain) semantics?

4.15 On the Conceptual Modeling of Behavior: Dynamic Reclassification of Entities

Alfonso Pierantonio (University of L'Aquila, IT)

The concept of classification as realized in most traditional object-oriented computer languages has certain limitations that may inhibit its application to modeling more complex phenomena. This is likely to prove problematic as modern software becomes increasingly more integrated with the highly dynamic physical world. In this paper, we first provide a detailed description of these limitations, followed by an outline of a novel approach to classification designed to overcome them. The proposed approach replaces the static multiple-inheritance hierarchy approach found in many object-oriented languages with multiple dynamic class hierarchies each based on different classification criteria. Furthermore, to better deal with ambiguous classification schemes, it supports potentially overlapping class membership within any given scheme. Also included is a brief overview of how this approach could be realized in the design of advanced computer languages.

4.16 Simulation to digital twin

Fiona A. C. Polack (University of Hull, GB)

We have already explored engineering a simulation so that it can be trusted (in the non-strict sense), and so that the basis for the trust can be interrogated and challenged: the principled simulation approach has been used in e.g. immune system simulation and robotics. In 2000s we proposed MDE as one way to automate development so that effort could focus on abstraction, design and interpretation of results, plus use of the simulation for its purpose. Digital Twin development might perhaps learn from engineering principled simulation, particularly in areas such as purpose, results interpretation/use, fitness/validation, management of trustworthiness. Again MDE could allow "people" to focus on these non-automatable but critical areas.

4.17 Digital Twins – Challenges from a Modelling Perspective

Matthias Riebisch (Universität Hamburg, DE)

This talk describes challenges for the development of the digital twin from a modelling perspective.

4.18 Federation of Digital Twins to conform with Manufacturing Systems

Matthias Riebisch (Universität Hamburg, DE)

Manufacturing systems are structured as multi-layered systems, with different goals and constraints for the layers. Digital Twins must reflect this by being split into parts what we call Federated Digital Twin. For its parts, and for the bridges between them, appropriate decisions on coverage, structure, technology, etc. are required.

4.19 Digital Ecosystems and Digital Twins

Markus Stumptner (University of South Australia – Mawson Lakes, AU)

Digital Twins need to accommodate the interactions of the System under Study (SuS) to be able to collect data, stimuli and outputs. This requires interaction with the (software) ecosystem surrounding the SuS, determined by relevant functions, and defining a landscape of models and their interactions.

4.20 The importance of being Uncertain

Antonio Vallecillo (University of Málaga, ES)

A fundamental characteristic of software models is their ability to represent the relevant characteristics of the system under study, at the appropriate level of abstraction. We now live in the age of cyber-physical systems, smart applications and the Internet of things, which require some forms of interaction with the physical world. Uncertainty is an inherent property of any system that operates in a real environment or that interacts with physical elements or with humans but, unfortunately, the explicit representation, management and analysis of uncertainty has not received much attention by the software modeling community. In this talk we analyze the impact of measurement uncertainty on the behavior of the system using examples, and describe the traditional ways of dealing with it, namely using fixed and confidence (adaptive) intervals. We then discuss the need to consider all attributes as random variables and the importance of properly comparing them.

4.21 Digital Z (model/shadow/twin/passport/ . . .) "twinning" for and by Systems Engineering

Hans Vangheluwe (University of Antwerp, BE)

Throughout their life-cycle (design, production, assembly, operation and optimization, maintenance, re-purposing, disposal), smart, adaptive systems will include ecosystems of Digital Zs (models/shadows/twins) to achieve a plethora of goals such as condition monitoring, fault diagnosis, predictive maintenance, optimization, etc. . This builds on many existing techniques, architectures and standards from real-time simulation, co-simulation, systems and control theory, IoT, knowledge management, machine learning, experiment/validity frames, surrogate modelling, etc. To satisfy system goals, a federated knowledge repository (graph) (a "Modelverse") containing both "linguistic" and "ontological" information, is used as a starting point for inferencing. This leads to the (product line: goals to realizations) construction of new Digital Z "experiments" which, when deployed, in turn, yield new data/knowledge which is merged (during or at the end of an experiment) with the information already present in the Modelverse. As multiple concurrent inferencing and experiment processes may exist, the Modelverse acts as a blackboard. Conceptually, one Digital Z "experiment" is created per Property of Interest (to be monitored, satisfied or optimized).

Multi-Paradigm Modelling (MPM) principles are used throughout: model explicitly, most appropriate formalism(s), abstraction(s) for architectures and views, requirements and designs, with workflows modelled explicitly too.

4.22 Reflections about Digital Twins

Andreas Wortmann (Universität Stuttgart, DE)

Digital twins are starting to appear everywhere. In this talk, I reflect upon definitions, purposes, and tools related to engineering and operating digital twins based on the largest cross-domain mapping study on the topic to date.

4.23 Digital Twins for Learning Healthcare Systems

Steffen Zschaler (King's College London, GB)

I report on our experience building agent-based simulations and, further, digital twins for learning healthcare systems, the role that MDE has to play in this, and the challenges for future work.

4.24 Are Digital Twins going to rule the world?

Mark van den Brand (TU Eindhoven, NL)

Digital Twins play an important role in engineering complex (high-tech) systems. They allow for real-time analysis of engineered systems in order to detect anomalies, to predict maintenance, and to optimise behaviour. The use of model-driven techniques accelerates the development of digital twins. The focus of the Software Engineering and Technology group from the TU/e focuses on the efficient development of Digital Twins, by means of model reuse and advance orchestration of the reused models.

The use of models used in the engineering process leads to the question whether the digital twin can replace the supervisory control of the engineered system. What are the requirements to facilitate this:

- Supervisory control in the cloud by means of the engineered system
- Fast communication channels between physical entity and the engineered system (5G and beyond)
- Sensors, actuators and edge computation on the engineered system.

5 Problem Space Summary

5.1 Contexts

In this group, the participants focused on identifying the different contexts in which a Digital Twin may be produced, the goals and purposes of those Digital Twins, and their qualities and properties. We motivated our definitions based on the range of talks given during the seminar, as outlined in section 4. The talks included a variety of contexts for digital twins, including a variety of assets for which a digital twin is constructed, a variety of environments in which the digital twin is embedded, and a variety of purposes for which the digital twin was constructed. Below, we provide a short summary of the key discussion points within the group, and our findings:

From the range of talks demonstrated in section 4, we have seen that there are many different assets for which a digital twin can be developed. We came to define the 'context' of a digital twin as the asset type, as well as the asset's environment. Properties of the environment may include whether there is human intervention in the environment, whether the system is cyber-controlled or not, and whether the asset was engineered or is natural.

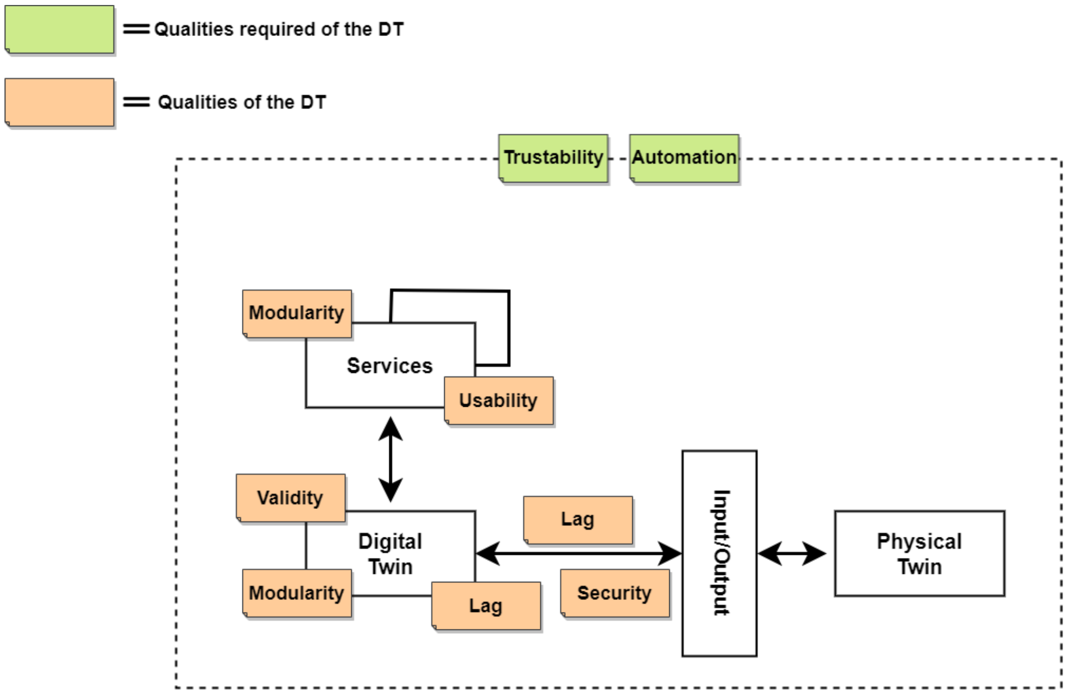


Figure 2 Digital Twin schema annotated with qualities.

5.2 Qualities of Digital Twins

In the problem space group, we then discussed the "qualities" of digital twins. I.e. the desirable properties of the system towards some measure or mitigation. We discussed the qualities of both the digital twin as an artifact, as well as the qualities that the digital twin must achieve for the user. Qualities apply to all engineered aspects including I/O and data flows between the physical and digital twins as well as the digital twin itself. We do not, however, comment on the qualities of the 'physical twin' as this is beyond the scope of the concerns of a digital twin developer.

During our discussions, we reached the opinion that qualities in the digital twin context are closely linked to qualities in the systems/software engineering context more generally. We, therefore, did not endeavor to produce a comprehensive list of qualities but instead aimed to identify some of the qualities of relevance to (different) digital twins.

In figure 2 we annotate a digital twin architecture diagram with the most relevant quality properties. These include:

- Validity: A measure to show that the digital twin is a "good enough" digital representation of the physical entity/twin
- Fidelity: How much reliance can be placed on the outputs (data, control signals, observations, etc.) of the digital twin.
- Trustworthiness: Closely linked to properties of fidelity and traceability. The degree to which stakeholders trust the validity of the digital twin as a whole.
- Modularity: The capacity to allow for the substitution of components such as services to fit the needs of the user in a given context
- Usability: The human-computer interaction properties of the system, and the standard of user experience the system provide
- Synchronisation: The properties of 'lag' between the digital and physical twin

- Security: Including considerations on risk analysis, data protection, and data biases

The goal of this group's discussions was to identify the fundamental scope and properties of digital twins. These properties were then used to motivate the requirements for the design space group as discussed in Section 6.

6 Design Space

The discussions in this group were targeting the design space aspect of using MDE techniques for Engineering Digital Twins. As a first discussion point, a conceptual model of digital twins was developed in order to gain a common understanding and terminology of the individual aspects and components of digital twins among the participants. The aim of this conceptual model was that it can be instantiated in various ways for specific domains, including cases where some of the components are missing.

A list of instantiations has been created for use cases provided by the participants in the domains on Anomaly Detection, Predictive Maintenance, Optimization, Diagnostics, and Policy Planning. The conceptual model was validated in these particular use cases, and in addition, for each use case, a design pattern and a list of open challenges was derived. In a final round of discussions, the results from the individual use cases were consolidated to come up with (i) a method for developing digital twins based on the individual design patterns, and (ii) a list of open challenges for designing digital twins using MDE techniques. The found challenges include (i) supporting experts in validating a DT system, (ii) capturing assumptions and the scope of a DT effectively, (iii) specification languages for DT systems, and (iv) design technologies for modeling privacy and security aspects of DTs.

7 Solution Space

The discussions in this group were targeting the necessary components needed for the MDE of Digital Twin Systems (DTSes). A model-based approach requires models that (1) may be used in the digital twin of the physical twin and (2) form the basis for generation and engineering tasks.

The distinction between descriptive, predictive, and prescriptive roles of models was discussed, as well as the fact that models are heterogeneous and composition covering various integration aspects is needed. Metamodels form the basis for model transformations from the various model kinds to other models or artefacts. MDE provides support for composition, consistency, an overall management of models. Overall, MDE provides a number of solutions for the challenges to be addressed in development of digital twins, but it is also evident that general MDE techniques and methods need to be adapted for the specific challenges of digital twin development.

Regarding connectivity, the group discussed how communication protocol choices depend on needs i.t.o. coupling, fidelity, service level, and quality of service; both for communication inside the DT as well as to its environment.

The importance of data storage, and of abstracting from data for purposes depending on the services the data is required for, were addressed. Suitable data pipelines, and integration with external services such as maintenance and operations management are important aspects regarding data and its management. Of course, models themselves need to be stored and processable so can be considered a form of data as well.

Finally, concerning services, we distinguished between internal and external services, with the latter for example covering data visualization and interaction with the digital twin. Important from the services point of view as well are data exchange, the availability of suitable execution and simulation engines, and data analytics. For service generation and operation, (meta)models also play an important role.

The group identified a list of challenges around the solution space for MDE of digital twins, concerning (1) integration and data exchange between different components; (2) consistency of models, data, and metamodels under evolution scenarios; (3) modeling environment interoperability (possibly including wrapping); (4) runtime support for time series data; (5) further standardization to allow interoperability and reuse of the artefacts involved in MDE based digital twins; (6) variability in DTs when considering product lines; (7) wrapping of tools patterns for the description and MDE of DTs; and (8) flexible definition of new services.

8 Conclusion

In this paper we have documented a fruitful week of discussions and community building in the the Dagstuhl Seminar **22362 – Model-Driven Engineering of Digital Twins**. This paper threat describes preliminary results of these discussions. We strongly expect that the discussions we started will continue in fruitful collaborations, potentially EU or national research programs addressing the question how to define digital twins in a more efficient and essential way and finally also lead to further events, e.g. such as a scientific conference about model-based engineering of digital twins.

As a short summary: we have identified that MDE does provide a number of solutions for the challenges to be addressed in development of digital twins, but it is also evident, that a specific adaptation of the general MDE techniques and methods is needed to address the specific challenges of digital twin development. And because the domain of digital twins is relevant, it is absolutely worthwhile addressing these challenges with MDE techniques.

Finally, we really want to thank all the local assistance, the organizational assistance and the strategic organizational people for making Dagstuhl this wonderful place to meet and discuss and put innovations forward. Thanks alot, we really appreciate it, especially inn these pandemic times.

References

1 Manuela Dalibor, Nico Jansen, Bernhard Rumpe, David Schmalzing, Louis Wachtmeister, Manuel Wimmer, and Andreas Wortmann. A cross-domain systematic mapping study on software engineering for Digital Twins. *Journal of Systems and Software*, November 2022.

2 John Hutchinson, Mark Rouncefield, and Jon Whittle. Model-driven engineering practices in industry. In *Proceedings of the 33rd International Conference on Software Engineering*, pages 633–642. ACM, 2011.

3 John Hutchinson, Jon Whittle, Mark Rouncefield, and Steinar Kristoffersen. Empirical assessment of mde in industry. In *Proceedings of the 33rd International Conference on Software Engineering*, pages 471–480. ACM, 2011.

4 Stuart Kent. Model driven engineering. In Michael Butler, Luigia Petre, and Kaisa Sere, editors, *Integrated Formal Methods*, pages 286–298, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

5 MDT. Digital twin consortium, “glossary of digital twins,”. `https://www.digitaltwinconsortium.org/glossary/index.html`, 2021.

6 Paula Munoz, Javier Troya, and Antonio Vallecillo. Using uml and ocl models to realize high-level digital twins. In *2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, pages 212–220. IEEE, 2021.

7 Juha-Pekka Tolvanen and Steven Kelly. Metaedit+: defining and using integrated domain-specific modeling languages. In *The 24th ACM SIGPLAN conference companion on OOPSLA*, pages 819–820, 2009.

Participants

- Shaukat Ali
Simula Research Laboratory – Oslo, NO
- Balbir Barn
Middlesex University – London, GB
- Ion Barosan
TU Eindhoven, NL
- Nelly Bencomo
Durham University, GB
- Francis Bordeleau
ETS – Montreal, CA
- Tony Clark
Aston University – Birmingham, GB
- Loek Cleophas
TU Eindhoven, NL
- Benoit Combemale
University & IRISA – Rennes, FR
- Thomas Godfrey
King's College London, GB
- Georg Grossmann
University of South Australia – Mawson Lakes, AU
- Gabor Karsai
Vanderbilt University, US
- Djamel Khelladi
CNRS – IRISA – Rennes, FR
- Oliver Kopp
Mercedes-Benz AG – Stuttgart, DE
- Daniel Lehner
Johannes Kepler Universität Linz, AT
- Bernhard Mitschang
Universität Stuttgart, DE
- Paula Muñoz Ariza
University of Mĺaga, ES
- Alfonso Pierantonio
University of L'Aquila, IT
- Fiona A. C. Polack
University of Hull, GB
- Matthias Riebisch
Universität Hamburg, DE
- Bernhard Rumpe
RWTH Aachen, DE
- Holger Schlingloff
HU Berlin, DE
- Markus Stumptner
University of South Australia – Mawson Lakes, AU
- Antonio Vallecillo
University of Málaga, ES
- Mark van den Brand
TU Eindhoven, NL
- Hans Vangheluwe
University of Antwerp, BE
- Andreas Wortmann
Universität Stuttgart, DE
- Steffen Zschaler
King's College London, GB

Report from Dagstuhl Seminar 22371

Algebraic and Analytic Methods in Computational Complexity

Markus Bläser*[1], Valentine Kabanets*[2], Ronen Shaltiel*[3], and Jacobo Torán*[4]

1 **Universität des Saarlandes – Saarbrücken, DE.** `mblaeser@cs.uni-saarland.de`
2 **Simon Fraser University – Burnaby, CA.** `kabanets@cs.sfu.ca`
3 **University of Haifa, IL.** `ronen@cs.haifa.ac.il`
4 **Universität Ulm, DE.** `jacobo.toran@uni-ulm.de`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 2237 "Algebraic and Analytic Methods in Computational Complexity".

Computational Complexity is concerned with the resources that are required for algorithms to detect properties of combinatorial objects and structures. It has often proven true that the best way to argue about these combinatorial objects is by establishing a connection (perhaps approximate) to a more well-behaved algebraic setting.

Beside algebraic methods, analytic methods have been used for quite some time in theoretical computer science. These methods can also be used to solve algebraic problems as show by many recent examples in the areas of derandomization, coding theory or circuit lower bounds. These new directions were in the focus of the Dagstuhl Seminar and reflect the developments in the field since the previous Dagstuhl Seminar 18391.

This Dagstuhl Seminar brought together researchers who are using a diverse array of algebraic and analytic methods in a variety of settings. Researchers in these areas are relying on ever more sophisticated and specialized mathematics and this seminar played a role in educating a diverse community about the latest new techniques, spurring further progress.

Seminar September 11–16, 2022 – http://www.dagstuhl.de/22371

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory; Theory of computation → Circuit complexity; Theory of computation → Problems, reductions and completeness; Theory of computation

Keywords and phrases (de-)randomization, algebra, circuits, coding, computational complexity

Digital Object Identifier 10.4230/DagRep.12.9.41

1 Executive Summary

Markus Bläser (Universität des Saarlandes – Saarbrücken, DE)
Valentine Kabanets (Simon Fraser University – Burnaby, CA)
Ronen Shaltiel (University of Haifa, IL)
Jacobo Torán (Universität Ulm, DE)

Introduction

The seminar on algebraic methods in computational complexity has traditionally taken place every two years in Dagstuhl for many years. In these meetings, we try to bring together leading researchers in a very active and broad area of theoretical computer science, having

* Editor / Organizer

the algebraic methods as a unifying thread. Researchers in these areas are relying on ever more sophisticated and specialized mathematics and this seminar can play an important role in educating a diverse community about the latest new techniques, spurring further progress. For the year 2022, we added a new direction that focused besides the algebraic aspect also on methods from analysis. The seminar brought together more than 40 researchers covering a wide spectrum of complexity theory. We had 24 talks, most of them lasting about 45 minutes, leaving ample room for discussions. In the following we describe the major topics of discussion in more detail.

Some areas of focus

Computational complexity is a fundamental and active subarea of theoretical computer science that has produced some of the most well known results in theoretical computer science in recent years. Here we discuss a few broad themes which highlight the importance of algebra as well as analytic methods in computational complexity, and which represent some focus areas of our present seminar.

Circuit complexity

Boolean circuits are one of the most fundamental model of computation. Due to its combinatorial nature, they seem more amenable to formal analysis than the uniform models such as Turing machines. The classical lower bound techniques of Razborov and Smolensky are algebraic: they work by first approximating $\mathrm{AC}^0[p]$ circuits (constant-depth circuits with AND, OR, NOT, and counting modulo prime p gates) by low-degree polynomials, and then proving that certain functions (like Majority) are not well correlated with such polynomials. The Fourier expansion of a Boolean function and its representation as a real multilinear polynomial as well as other analytic tools have been added in the last years to the bag of tools used for the analysis of Boolean circuits. In the seminar, we talked about recent results in circuit complexity.

Andrej Bogdanov talked about property testing. He constructed a natural tester that tells if a function from $\{0,1\}^n$ to some Abelian group is linear (or far from linear).

Frederic Green proved a new correlation bound for certain exponential sums over characteristic 5.

William Hoza presented the construction of a Boolean function F on n bits such that F can be computed by a uniform depth-$(d+1)$ AC^0 circuit with $O(n)$ wires, but F cannot be computed by any depth-d TC^0 circuit with $n^{1+\gamma}$ wires, where $\gamma = 2^{-\Theta(d)}$ and $d = o(\log\log n)$.

Michal Koucký dealt with a classical problem, the simulation of Turing machines by circuits. He gave a new simple proof for the classical result that Turing machines running in time $t(n)$ and space $s(n)$ can be simulated by Boolean circuits of size $O(t(n)\log s(n))$ and of depth $O(t(n))$.

Meena Mahajan presented relations between the minimum rank of a decision tree computing a Boolean function and other complexity measures of the function, as well as a new composition theorem in terms of rank and decision tree depth.

In his talk, *Rocco Servedio* establish a new quantitative version of the Gaussian correlation inequality. It gives a lower bound on the correlation of two centrally symmetric convex sets based on their "common influential directions".

A new family of sampling tasks was presented by *Rahul Santhanam*. He showed that any non-trivial algorithmic solutions to tasks from this family imply new uniform lower bounds such as "NP not in uniform ACC^0" or "NP does not have uniform depth-2 threshold circuits".

Algebraic complexity

A class of circuits especially suited for the use of algebraic techniques is that of *arithmetic circuits*. These are circuit models that compute polynomial functions by using gates performing arithmetic operations (additions, subtractions, multiplications, divisions, etc.) Two fundamental complexity measures for arithmetic circuits are the *size* and the *depth* or *product depth*.

Prerona Chatterjee considered the question of proving lower bounds against non-commutative circuits better than $\Omega(n \log n)$. She showed a quadratic lower bound against the n-variate central symmetric polynomial.

Arkadev Chattopadhyay talked about connections between communication complexity measures and monotone arithmetic circuit lower bounds. He constructed a (set-multilinear) monotone polynomial that can be computed by depth-3 multilinear formulas in sub-cubic size but requires exponential size to be computed by monotone arithmetic circuits. Second, he proved the existence of a polynomial over n variables in VNP, for which $2^{\Omega(n)}$ size ϵ-sensitive lower bounds hold if $\epsilon = 2^{-O(n)}$.

Barrier results in the group-theoretic approach to bounding the exponent of matrix multiplication was the topic of the talk by *Chris Umans*. He showed that finite groups of Lie type cannot prove $\omega = 2$ and presented a further barrier result. Then he gave constructions in the continuous setting, which can potentially evade these two barriers.

Pascal Koiran studied the decomposition of multivariate polynomials as sums of powers of linear forms. He presented a randomized algorithm for the following problem: Given a homogeneous polynomial of degree d as a blackbox, decide whether it can be written as a linear combination of dth powers of linearly independent complex linear forms.

Nutan Limaye proved in her talk that there exist monomial symmetric polynomials that are hard for the class VNP.

Pseudorandomness and derandomization

The theory of pseudorandomness studies explicit constructions and applications of "random-like" objects of combinatorial or algebraic type. The common feature of such objects is that it is easy to construct one by random sampling, but a very important problem is to get efficient *deterministic* constructions.

Eric Allender proved that Kolmogorov complexity characterizes statistical zero knowledge. Every decidable promise problem has a non-interactive statistical zero-knowledge proof system if and only if it is randomly reducible to a promise problem for Kolmogorov-random strings.

Random walks on expanders are a useful tool in complexity theory. *Gil Cohen* explained how the inherent cost can be reduced from exponential to linear by applying a permutation after each random step.

Sylvester-Gallai type problems have found applications in polynomial identity testing and coding theory. *Rafael Oliveira* discussed such problems and their relation to algebraic computation, and presented a theorem that radical Sylvester-Gallai configurations for cubic polynomials must have small dimension.

Ryan O'Donnell explained how to contruct high-dimensional expanders from Chevalley groups.

Motivated by applications from cryptography, *Noga Ron-Zewi* studied a new interactive variant of PCPs, so-called interactive oracle proofs. She showed that for this model the overhead in the encoding can be made arbitrarily small and the prover complexity overhead can be made constant.

In his talk, *Amon Ta-Shma* gave an alternative construction of the lossless condenser by Guruswami, Umans and Vadhan. Instead of Parvaresh-Vardy codes, the new construction is based on multiplicity codes.

A Chor-Goldreich source is a sequence of random variables where each has min-entropy, even conditioned on the previous ones. *David Zuckerman* showed how to extend this notion in several ways, most notably allowing each random variable to have Shannon entropy conditioned on previous ones. He then proved new pseudorandomness results for Shannon-CG sources.

Border complexity and invariant theory

Many problems in algebraic complexity theory can be written as an orbit closure problem. We are given a vector space V and a group G acting on it. The orbit Gv of an element $v \in V$ is the set $\{gv \mid g \in G\}$ and its closure is the usual closure in the Zariski topology. For instance, we can formulate the tensor border rank problem in this language: Alder and Strassen proved that the question whether a tensor t has border rank $\leq r$ is equivalent to deciding whether t is in the orbit closure (under the standard action $\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n$) of the so-called unit tensor of size r. As second example is provided by Mulmuley and Sohoni who formulated a variant of the permanent versus determinant question as an orbit closure problem.

Peter Bürgisser gave an introduction to new algorithmic and analysis techniques that extend convex optimization from the classical Euclidean setting to a general geodesic setting. He pointed out the relevance of invariant and representation theory for for complexity theory and highlighted connections to different areas of mathematics, statistics, computer science, and physics.

Rohit Gurjar considered determinants of the matrices of the form $(\sum_i A_i x_i)$ where each A_i is rank one. He showed that this class of polynomials is closed under approximation.

Approximate complexity was also the topic of *Nitin Saxena's* talk. He proved that the border of bounded-top-fanin depth-3 circuits is relatively easy, since it can be computed by a polynomial-size algebraic branching program.

Counting and quantum complexity

In order to study the #P (non-)membership of some concrete problems, *Christian Ikenmeyer* started the development of a classification of the #P closure properties on affine varieties. He obtained oracle separations between counting classes, where the existence of the oracle is based on properties of the vanishing ideal of an affine variety.

Steve Fenner considered a problem in quantum computing, the construction of a "realistic" Hamiltonian for quantum fanout.

Conclusion

The talks of the seminar ranged over a broad assortment of subjects with the underlying theme of using algebraic and analytic techniques. It was a very fruitful meeting and it has hopefully initiated new directions in research. Several participants specifically mentioned that they appreciated the particular focus on a common class of techniques (rather than end results) as a unifying theme of the workshop. We look forward to our next seminar.

2 Table of Contents

3 Overview of Talks

3.1 Kolmogorov Complexity Characterizes Statistical Zero Knowledge

Eric Allender (Rutgers University – Piscataway, US)

Joint work of Eric Allender, Harsha Tirumala, and Shuichi Hirahara
Main reference Eric Allender, Shuichi Hirahara, Harsha Tirumala: "Kolmogorov Complexity Characterizes Statistical Zero Knowledge", ECCC TR22-127, 2022
URL https://eccc.weizmann.ac.il/report/2022/127/

We show that a decidable promise problem has a non-interactive statistical zero-knowledge proof system if and only if it is randomly reducible to a promise problem for Kolmogorov-random strings, with a superlogarithmic additive approximation term. This extends recent work by Saks and Santhanam (CCC 2022). We build on this to give new characterizations of Statistical Zero Knowledge (SZK), as well as the related classes NISZK_L and SZK_L.

3.2 Direct sum testing over Abelian groups

Andrej Bogdanov (The Chinese University of Hong Kong, HK)

Joint work of Andrej Bogdanov, Gautam Prakriya
Main reference Andrej Bogdanov, Gautam Prakriya: "Direct Sum and Partitionability Testing over General Groups", in Proc. of the 48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference), LIPIcs, Vol. 198, pp. 33:1–33:19, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.
URL http://dx.doi.org/10.4230/LIPIcs.ICALP.2021.33

I spoke about a natural tester that tells if a function from $\{0,1\}^n$ to some Abelian group like Z_3 is linear (or far from linear). More generally, the tester can be used to tell if a multivariate function $g(x_1, ..., x_n)$ admits a direct sum decomposition $f(x_1) + ... + f(x_n)$ for some f.

3.3 Optimization, Complexity and Invariant Theory

Peter Bürgisser (TU Berlin, DE)

Joint work of Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, Avi Wigderson
Main reference Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, Avi Wigderson: "Towards a Theory of Non-Commutative Optimization: Geodesic 1st and 2nd Order Methods for Moment Maps and Polytopes", in Proc. of the 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2019.
URL http://dx.doi.org/10.1109/focs.2019.00055

Invariant and representation theory studies symmetries by means of group actions and is a well established source of unifying principles in mathematics and physics. Recent research suggests its relevance for complexity and optimization through quantitative and algorithmic questions. The goal of the talk is to give an introduction to new algorithmic and analysis techniques that extend convex optimization from the classical Euclidean setting to a general geodesic setting. We also point out surprising connections to a diverse set of problems in different areas of mathematics, statistics, computer science, and physics.

3.4 A Quadratic Lower Bound Against Homogeneous Non-Commutative Circuits

Prerona Chatterjee (The Czech Academy of Sciences – Prague, CZ)

Joint work of Prerona Chatterjee, Pavel Hrubeš

Inspite of the various strong lower bounds against constant depth circuits and the depth reduction results in algebraic circuit complexity, the best lower bound known against general algebraic circuits remains $\Omega(n \log n)$ [Strassen, 1973; Baur-Strassen 1983]. Nothing better is known even in the more restrictive non-commutative setting where the product gates are considered to denote non-commutative multiplication. This is surprising since exponential lower bunds are known against algebraic formulas [Nisan 1991] and super polynomial lower bounds are known against homogenous formulas for polynomials computable even by ABPs [Tavenas, Limaye, Srinivasan, 2022]. A natural question is therefore to prove better lower bounds against non-commutative circuits. In this talk, we make progress in this question by showing a quadratic lower bound against the n-variate central symmetric polynomial. Further, the simplicity of the proof motivates us to ask whether a similar lower bound can be shown against general non-commutative circuits. This is ongoing work with Pavel Hrubes.

3.5 Monotone Arithmetic Lower Bounds Via Communication Complexity

Arkadev Chattopadhyay (TIFR – Mumbai, IN)

Joint work of Arkadev Chattopadhyay, Rajit Datta, Utsab Ghosal, Partha Mukhopadhyay
Main reference Arkadev Chattopadhyay, Rajit Datta, Partha Mukhopadhyay: "Lower bounds for monotone arithmetic circuits via communication complexity", in Proc. of the STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021, pp. 786–799, ACM, 2021.
URL http://dx.doi.org/10.1145/3406325.3451069

We make two novel connections between communication complexity measures and monotone arithmetic circuit lower bounds. The first connection exploits the corruption measure. We formulate a general method that constructs a set-multilinear polynomial P_f from a Boolean function f and uses the corruption bound of $f \circ \text{XOR}$ to imply a size lower bound on monotone arithmetic circuits computing P_f. Using this method, we construct [1] a (set-multilinear) monotone polynomial that can be computed by depth-3 multilinear formulas in sub-cubic size but require exponential size to be computed by monotone arithmetic circuits. It was not even known, before our work, if general formulas of arbitrary depth could provide exponential savings in size over monotone circuits.

The second connection uses the discrepancy measure from communication complexity to lower bound the size of monotone circuits computing a polynomial even in an ϵ-sensitive way. Very recently, Hrubes [3] showed that ϵ-sensitive monotone lower bounds, for arbitrary small positive ϵ, implies general circuit lower bounds. We formulate [2] a general recipe between discrepancy under a *universal* distribution and ϵ-sensitive bounds. Using this connection, we show the following:

- there exists a polynomial over n variables, crafted out of the Boolean inner-product function defined using expander graphs, that is in VNP and for which $2^{\Omega(n)}$ size ϵ-sensitive lower bounds hold if $\epsilon = 2^{-O(n)}$.
- the spanning tree polynomial, defined over the edge variables of a complete graph on n vertices, needs $2^{\Omega(n)}$ size to be computed by monotone circuits in an ϵ-sensitive way as long as $\epsilon = 2^{-O(n)}$. Recall that the number of variables of this spanning tree polynomial is $\Theta(n^2)$ and it is in VP.

This is based on two papers referenced below.

References

1 Arkadev Chattopadhyay, Rajit Datta, and Partha Mukhopadhyay, *Lower bounds for monotone arithmetic circuits via communication complexity*, STOC, 2021.
2 Arkadev Chattopadhyay, Rajit Datta, Utsab Ghosal, and Partha Mukhopadhyay, *Monotone complexity of spanning tree polynomial revisited*, ITCS, 2022.
3 Pavel Hrubes, *On ϵ-sensitive monotone computations*, Computational Complexity, 2020.

3.6 Random walks on rotating expanders

Gil Cohen (Tel Aviv University, IL)

Joint work of Gil Cohen, Gal Maor

Random walks on expanders are extremely useful in TOC. Unfortunately though, they have an inherent cost. E.g., the spectral expansion of a Ramanujan graph deteriorates exponentially with the length of the walk (when compared to a Ramanujan graph of the same degree). In this talk, we will see how this exponential cost can be reduced to linear by applying a permutation after each random step. These permutations are tailor-made to the graph at hand, requiring no randomness. Our proof is established using the powerful framework of finite free probability and interlacing families that was introduced, around ten years ago, by Marcus, Spielman and Srivastava in their seminal works on the existence of bipartite Ramanujan graphs of every size and every degree, and in their solution to the Kadison-Singer problem.

3.7 A "Realistic" Hamiltonian for Quantum Fanout

Stephen A. Fenner (University of South Carolina – Columbia, US)

Joint work of Stephen A Fenner, Rabins Wosti
Main reference Stephen Fenner, Rabins Wosti: "Implementing the fanout operation with simple pairwise interactions", arXiv, 2022.
URL http://dx.doi.org/10.48550/ARXIV.2203.01141

We give a swap-invariant diagonal gate U_n equivalent in constant depth to the n-qubit fanout gate. For $t = \pi/4$ and real coupling constants $\{\alpha_{i,j} : 1 \leq i, j \leq n\}$ with $\alpha_{i,j} = \alpha_{j,i}$, $\alpha_{ii} = 0$, the Hamiltonian $H_{\vec{\alpha}} := \sum_{i<j} \alpha_{i,j} Z_i Z_j$ implements U_n (i.e., $U_n = \exp(-iH_{\vec{\alpha}}t)$ up to a global phase factor) if and only if: (1) all the $\alpha_{i,j}$ are odd integers; and (2) for all i,

$\prod_{j \neq i} \alpha_{i,j} \equiv 1 \pmod 4$. We give tight constraints on $\{\alpha_{i,j}\}$ as above for spatial arrangements of identical qubits satisfying an inverse square law. These constraints are obtained using modular arithmetic on rational numbers.

Joint work with Rabins Wosti.

3.8 New Correlation Bounds for Quadratic Polynomials

Frederic Green (Clark University – Worcester, US)

Let p be an odd prime, $\zeta = e^{2\pi i/p}$ a complex primitive p^{th} root of unity, and $\chi : \mathbb{Z}_5 \to \mathbb{C}$ the quadratic character over $\mathbb{Z}_p$. Let $t \in \mathbb{Z}_p[x_1, \ldots, x_n]$ be an n-variable quadratic polynomial $\sum_{i,j} c_{ij} x_i x_j + \sum_i \ell_i x_i$. Consider the exponential sum,

$$S = \frac{1}{(p-1)^n} \sum_{\mathbf{x} \in \mathbb{Z}_p^n} \chi(\prod_{i=1}^{n} x_i) \zeta^{t(\mathbf{x})},$$

which can be interpreted as the correlation between the parity of the number of x_i's which are quadratic residues and whether $t(\mathbf{x}) \equiv 0$ (p). In 2001, Green (JCSS **69**, 2004, pp. 28–44) showed that for $p = 3$, $|S| \leq (|\zeta - -\overline{\zeta}|/2)^{\lceil n/2 \rceil}$, and that this bound can be met by $x_1x_2 + x_3x_4 + \ldots$. In this talk, we prove a tight bound for $|S|$ when $p = 5$: $|S| \leq (|\zeta - -\overline{\zeta}|)/2)^n$, which can be met by the polynomial $x_1^2 + x_2^2 + \cdots + x_n^2$. The technique relies on some of the simpler methods of those recently developed by Ivanov, Pavlovic, and Viola (ECCC TR22-092, July 2022). The latter paper consider sums of the form,

$$\frac{1}{2^n} \sum_{\mathbf{x} \in \{0,1\}^n} \zeta^{\sum_{i=1}^{n} x_i} (-1)^{t(\mathbf{x})},$$

again with t quadratic, and, remarkably, prove tight upper bounds met by symmetric polynomials for *any* complex unit ζ. It is not yet clear how to extend the simpler method for $p = 5$ to other odd moduli.

3.9 Set of rank-1 determinant polynomials is closed under approximations

Rohit Gurjar (Indian Institute of Technology – Mumbai, IN)

Joint work of Rohit Gurjar, Abhranil Chatterjee, Sumanta Ghosh, Roshan Raj

Consider the class of polynomials computed by rank-one determinants – determinants of the matrices of the form $(\sum_i A_i x_i)$ where each A_i is rank one. These polynomials appear naturally in the study of bipartite matching and related combinatorial problems. We show that this class of polynomials is closed under approximation. Interestingly, the proof of closure uses ideas from combinatorial optimization, specifically Rado's theorem on matroid transversals.

3.10 Depth-d Threshold Circuits vs. Depth-(d + 1) AND-OR Trees

William Hoza (University of California – Berkeley, US)

Joint work of William Hoza, Avishay Tal, Pooya Hatami, Roei Tell
Main reference Pooya Hatami, William Hoza, Avishay Tal, Roei Tell: "Depth-d Threshold Circuits vs. Depth-(d + 1) AND-OR Trees", Electron. Colloquium Comput. Complex., Vol. TR22-087, 2022.
URL https://eccc.weizmann.ac.il/report/2022/087

For $n \in \mathbb{N}$ and $d = o(\log \log n)$, we prove that there is a Boolean function F on n bits and a value $\gamma = 2^{-\Theta(d)}$ such that F can be computed by a uniform depth-$(d+1)$ AC^0 circuit with $O(n)$ wires, but F cannot be computed by any depth-d TC^0 circuit with $n^{1+\gamma}$ wires. This bound matches the current state-of-the-art lower bounds for computing explicit functions by threshold circuits of depth $d > 2$, which were previously known only for functions outside AC^0 such as the parity function. Furthermore, in our result, the AC^0 circuit computing F is a monotone *read-once formula* (i.e., an AND-OR tree), and the lower bound holds even in the average-case setting with respect to advantage $n^{-\gamma}$.

Our proof builds on the *random projection* procedure of Håstad, Rossman, Servedio, and Tan, which they used to prove the celebrated average-case depth hierarchy theorem for AC^0 (J. ACM, 2017). We show that under a modified version of their projection procedure, any depth-d threshold circuit with $n^{1+\gamma}$ wires simplifies to a near-trivial function, whereas an appropriately parameterized AND-OR tree of depth $d+1$ maintains structure.

3.11 The algebraic geometry of the closure properties of #P

Christian Ikenmeyer (University of Liverpool, GB)

Joint work of Christian Ikenmeyer, Igor Pak
Main reference Christian Ikenmeyer, Igor Pak: "What is in #P and what is not?", in Proc. of the 63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 – November 3, 2022, pp. 860–871, IEEE, 2022.
URL http://dx.doi.org/10.1109/FOCS54457.2022.00087

Since 1995 the functional closure properties of #P are beautifully classified via the coefficients in the expansion over the binomial basis. In order to study the #P (non-)membership of concrete problems related to counting versions of TFNP problems, we start the development of a classification of the #P closure properties on affine varieties. We obtain oracle separations between counting classes, where the existence of the oracle is based on properties of the vanishing ideal of an affine variety, which then translates to a specific polyhedron having no integer point. This is a part of the recent FOCS 2022 paper "What is in #P and what is not", which is joint work with Igor Pak.

References

1 Christian Ikenmeyer and Igor Pak. *What is in #P and what is not.* Proceedings FOCS 2022, full version on arXiv:2204.13149

3.12 Black Box Absolute Reconstruction for Sums of Powers of Linear Forms

Pascal Koiran (ENS – Lyon, FR)

Joint work of Pascal Koiran and Subhayan Saha
Main reference Pascal Koiran, Subhayan Saha: "Black Box Absolute Reconstruction for Sums of Powers of Linear Forms", in Proc. of the 42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2022, December 18-20, 2022, IIT Madras, Chennai, India, LIPIcs, Vol. 250, pp. 24:1–24:17, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
URL http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2022.24

We study the decomposition of multivariate polynomials as sums of powers of linear forms. We give a randomized algorithm for the following problem: If a homogeneous polynomial $f \in K[x_1, ..., x_n]$ (where $K \subseteq \mathbb{C}$) of degree d is given as a blackbox, decide whether it can be written as a linear combination of d-th powers of linearly independent complex linear forms. The main novel features of the algorithm are:

- For $d = 3$, we improve by a factor of n on the running time from the algorithm in (Koiran and Skomra, 2020). The price to be paid for this improvement is that the algorithm now has two-sided error.
- For $d > 3$, we provide the first randomized blackbox algorithm for this problem that runs in time poly(n, d) (in an algebraic model where only arithmetic operations and equality tests are allowed). Previous algorithms for this problem (Kayal, 2011) as well as most of the existing reconstruction algorithms for other classes appeal to a polynomial factorization subroutine. This requires extraction of complex polynomial roots at unit cost and in standard models such as the unit-cost RAM or the Turing machine this approach does not yield polynomial time algorithms.
- For $d > 3$, when f has rational coefficients (i.e. $K = \mathbb{Q}$), the running time of the blackbox algorithm is polynomial in n, d and the maximal bit size of any coefficient of f. This yields the first algorithm for this problem over $\mathbb{C}$ with polynomial running time in the bit model of computation.

These results are true even when we replace $\mathbb{C}$ by $\mathbb{R}$. We view the problem as a tensor decomposition problem and use linear algebraic methods such as checking the simultaneous diagonalisability of the slices of a tensor. The number of such slices is exponential in d. But surprisingly, we show that after a random change of variables, computing just 3 special slices is enough. We also show that our approach can be extended to the computation of the actual decomposition. This step relies on matrix diagonalisation which is not an algebraic step over $\mathbb{C}$. In forthcoming work we plan to extend these results to overcomplete decompositions, i.e., decompositions in more than n powers of linear forms.

3.13 Turning Turing Machines into Boolean Circuits

Michal Koucký (Charles University – Prague, CZ)

We give a new simple proof for the classical result that Turing machines running in time $t(n)$ and space $s(n)$ can be simulated by boolean circuits of size $O(t(n) logs(n))$ and of depth $O(t(n))$. When we allow unbounded fan-in gates we can get circuits of the same size and depth $O(t(n)/loglogs(n))$.

3.14 The complexity of monomial symmetric polynomials

Nutan Limaye (IT University of Copenhagen, DK)

Joint work of Nutan Limaye, Radu Curticapean, Srikanth Srinivasan
Main reference Radu Curticapean, Nutan Limaye, Srikanth Srinivasan: "On the VNP-Hardness of Some Monomial Symmetric Polynomials", in Proc. of the 42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2022, December 18-20, 2022, IIT Madras, Chennai, India, LIPIcs, Vol. 250, pp. 16:1–16:14, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
URL http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2022.16

The determinant of the Vandermonde matrix has a very simple algebraic formula. However, the complexity of its permanent, denoted in this talk as Perm(V), is not known. The permanent of the Vandermonde matrix is a "monomial symmetric polynomial". In this talk we show that there exist monomial symmetric polynomials that are hard for VNP.

3.15 Decision tree rank for Boolean functions

Meena Mahajan (The Institute of Mathematical Sciences – Chennai, IN)

Joint work of Yogesh Dahiya, Meena Mahajan
Main reference Yogesh Dahiya, Meena Mahajan: "On (Simple) Decision Tree Rank", in Proc. of the 41st IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2021, December 15-17, 2021, Virtual Conference, LIPIcs, Vol. 213, pp. 15:1–15:16, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.
URL http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2021.15

In this talk, I describe some relations between the minimum rank of a decision tree computing a Boolean function and other complexity measures of the function. I also describe a composition theorem in terms of rank and decision tree depth, and show how it simplifies some known lower bounds on decision tree size and rank.

Joint work with Yogesh Dahiya.

3.16 Radical Sylvester-Gallai theorem for cubics – and beyond

Rafael Mendes de Oliveira (University of Waterloo, CA)

Joint work of Rafael Oliveira, Akash Kumar Sengupta
Main reference Rafael Mendes de Oliveira, Akash Sengupta: "Radical Sylvester-Gallai for Cubics", Electron. Colloquium Comput. Complex., Vol. TR22-131, 2022.
URL https://eccc.weizmann.ac.il/report/2022/131

In 1893, Sylvester asked a basic question in combinatorial geometry: given a finite set of distinct points $v_1, \ldots, v_m \in \mathbb{R}^N$ such that the line defined by any pair of distinct points v_i, v_j contains a third point v_k in the set, must all points in the set be collinear?

Generalizations of Sylvester's problem, which are known as Sylvester-Gallai type problems, have found applications in algebraic complexity theory (in Polynomial Identity Testing – PIT) and coding theory (Locally Correctable Codes). The underlying theme in all these types of questions is the following:

Are Sylvester-Gallai type configurations always low-dimensional?

In 2014, Gupta, motivated by such applications in algebraic complexity theory, proposed wide-ranging non-linear generalizations of Sylvester's question, with applications on the PIT problem.

In this talk, we will discuss these non-linear generalizations of Sylvester's conjecture, their intrinsic relation to algebraic computation, and a recent theorem proving that radical Sylvester-Gallai configurations for cubic polynomials must have small dimension.

Joint work with Akash Kumar Sengupta.

3.17 High-dimensional expanders from Chevalley groups

Ryan O'Donnell (Carnegie Mellon University – Pittsburgh, US)

Joint work of Ryan O'Donnell, Kevin Pratt
Main reference Ryan O'Donnell, Kevin Pratt: "High-Dimensional Expanders from Chevalley Groups", in Proc. of the 37th Computational Complexity Conference, CCC 2022, July 20-23, 2022, Philadelphia, PA, USA, LIPIcs, Vol. 234, pp. 18:1–18:26, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
URL http://dx.doi.org/10.4230/LIPIcs.CCC.2022.18

In this talk I discussed recent joint work with Kevin Pratt on constructing high-dimensional expanders.

Let Φ be an irreducible root system (other than G_2) of rank at least 2, let $\mathbb{F}$ be a finite field with $p = \operatorname{char} \mathbb{F} > 3$, and let $\mathrm{G}_\Phi\mathbb{F}$ be the corresponding Chevalley group. We describe a strongly explicit high-dimensional expander (HDX) family of dimension rank(Φ), where $\mathrm{G}_\Phi\mathbb{F}$ acts simply transitively on the top-dimensional faces; these are λ-spectral HDXs with $\lambda \to 0$ as $p \to \infty$. This generalizes a construction of Kaufman and Oppenheim (STOC 2018), which corresponds to the case $\Phi = A_d$. Our work gives three new families of spectral HDXs of any dimension ≥ 2, and four exceptional constructions of dimension 4, 6, 7, and 8.

3.18 Highly-efficient local proofs

Noga Ron-Zewi (University of Haifa, IL)

Joint work of Noga Ron-Zewi, Ron Rothblum

The celebrated PCP theorem from the 90's shows that any mathematical proof can be encoded in such a way that its correctness can be verified locally by reading only a tiny number of bits from the encoding. A fundamental question that has drawn a great amount of interest is what is the minimal overhead in encoding that is needed to allow for such highly efficient local verification. While the original PCP theorem only guarantees a polynomial overhead, a beautiful line of work has culminated in remarkably short encodings with only a poly-logarithmic overhead. Motivated by cryptographic applications, we study a relatively new interactive variant of PCPs, called Interactive Oracle Proofs, and show that for this model the overhead in the encoding can be made arbitrarily small (approaching 1), and moreover, the prover complexity overhead can be made constant.

The improved efficiency was obtained by replacing polynomial-based codes, commonly used in such proof systems, with more efficient (tensor-based) codes. In particular, these constructions bypassed a barrier imposed by the need to encode the computation using a multiplication code.

References

1 Noga Ron-Zewi and Ron D. Rothblum, Proving as fast as computing: succinct arguments with constant prover overhead, STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 – 24, 2022, 1353–1363, ACM, 2022,

2 Noga Ron-Zewi and Ron D. Rothblum, Local Proofs Approaching the Witness Length [Extended Abstract], 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020, 846–857, IEEE, 2020

3.19 An Algorithmic Approach to Uniform Lower Bounds

Rahul Santhanam (University of Oxford, GB)

We propose a new family of sampling tasks such that non-trivial algorithmic solutions to certain tasks from this family imply frontier uniform lower bounds such as "NP not in uniform ACC^0" and "NP does not have uniform depth-2 threshold circuits". Indeed, the most general versions of these sampling tasks have implications even for central open problems such as PSPACE vs P and NP vs P.

We observe that these sampling tasks do have non-trivial solutions under standard cryptographic assumptions. Moreover, we can use our framework to capture uniform versions of known non-uniform lower bounds, as well as classical results such as the space hierarchy theorem and Allender's uniform lower bound for the Permanent. Our framework can also be used to show that NP does not have uniform AC^0 circuits with a bottom layer of Mod 6 gates – the non-uniform version of this lower bound appears to be an open question.

3.20 Demystifying the border of depth-3 algebraic circuits

Nitin Saxena (Indian Institute of Technology Kanpur, IN)

Joint work of Pranjal Dutta, Prateek Dwivedi, Nitin Saxena
Main reference Pranjal Dutta, Prateek Dwivedi, Nitin Saxena: "Demystifying the border of depth-3 algebraic circuits", in Proc. of the 62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022, pp. 92–103, IEEE, 2021.
URL http://dx.doi.org/10.1109/FOCS52979.2021.00018
Main reference Pranjal Dutta, Prateek Dwivedi, Nitin Saxena: "Deterministic Identity Testing Paradigms for Bounded Top-Fanin Depth-4 Circuits", in Proc. of the 36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference), LIPIcs, Vol. 200, pp. 11:1–11:27, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.
URL http://dx.doi.org/10.4230/LIPIcs.CCC.2021.11

Border (or approximative) complexity of polynomials plays an integral role in GCT approach to P$\neq$NP. This raises an important open question: can a border circuit be *efficiently* debordered (i.e. convert from approximative to exact)? Or, could the approximation involve exponential-precision which may not be efficiently simulable? Circuits of depth 3 or 4, are a good testing ground for this question.

Recently, (Kumar ToCT'20) proved the universal power of the border of top-fanin-2 depth-3 circuits. We recently solved some of the related open questions. In this talk we outline our result: border of bounded-top-fanin depth-3 circuits is relatively easy– it can be computed by a polynomial-size algebraic branching program (ABP). Our de-bordering paradigm has many applications, especially in identity testing and lower bounds.

Based on the works with Prateek Dwivedi & Pranjal Dutta (CCC 2021) (FOCS 2021, invited to SICOMP).

3.21 Convex influences and a quantitative Gaussian correlation inequality

Rocco Servedio (Columbia University – New York, US)

Joint work of Anindya De, Shivam Nadimpalli, Rocco Servedio
Main reference Anindya De, Shivam Nadimpalli, Rocco A. Servedio: "Convex Influences", in Proc. of the 13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 – February 3, 2022, Berkeley, CA, USA, LIPIcs, Vol. 215, pp. 53:1–53:21, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
URL http://dx.doi.org/10.4230/LIPIcs.ITCS.2022.53
Main reference Anindya De, Shivam Nadimpalli, Rocco A. Servedio: "Quantitative Correlation Inequalities via Semigroup Interpolation", in Proc. of the 12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference, LIPIcs, Vol. 185, pp. 69:1–69:20, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.
URL http://dx.doi.org/10.4230/LIPIcs.ITCS.2021.69

The Gaussian correlation inequality (GCI), proved by Royen in 2014, states that any two centrally symmetric convex sets (say K and L) in Gaussian space are positively correlated. We establish a new quantitative version of the GCI which gives a lower bound on this correlation based on the "common influential directions" of K and L. This can be seen as a Gaussian space analogue of Talagrand's well known correlation inequality for monotone Boolean functions.

To obtain this inequality, we propose a new approach, based on analysis of Littlewood type polynomials, which gives a recipe for transferring qualitative correlation inequalities into quantitative correlation inequalities. En route, we also give a new notion of influences for symmetric convex symmetric sets over Gaussian space which has many of the properties of influences of Boolean functions over the discrete cube. Much remains to be explored about this new notion of influences for convex sets.

Based on joint work with Anindya De and Shivam Nadimpalli.

3.22 Lossless Condensers from Multiplicity Codes

Amnon Ta-Shma (Tel Aviv University, IL)

Joint work of Itay Kalev, Amon Ta-Shma
Main reference Itay Kalev, Amnon Ta-Shma: "Unbalanced Expanders from Multiplicity Codes", in Proc. of the Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022, September 19-21, 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference), LIPIcs, Vol. 245, pp. 12:1–12:14, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
URL http://dx.doi.org/10.4230/LIPIcs.APPROX/RANDOM.2022.12

In 2007 Guruswami, Umans and Vadhan gave an explicit construction of a lossless condenser based on Parvaresh-Vardy codes. This lossless condenser is a fundamental building block in many constructions, and, in particular, is behind state-of-the-art extractor constructions.

We give an alternative construction that is based on Multiplicity codes. While the bottom-line result is similar to the GUV result, the analysis is very different. In GUV (and Parvaresh-Vardy codes) the polynomial ring is closed to a finite field, and every polynomial is associated with related elements in the finite field. In our construction a polynomial from the polynomial ring is associated with its iterated derivatives. Our analysis boils down to solving a differential equation over a finite field, and uses previous techniques, introduced by Kopparty for the list-decoding setting. We also observe that these (and more general) questions were studied in differential algebra, and we use the terminology and result developed there.

We believe these techniques have the potential to get better constructions and solve the current bottlenecks in the area.

3.23 Matrix multiplication via matrix groups

Christopher Umans (California Institute of Technology – Pasadena, US)

Joint work of Christopher Umans, Henry Cohn, Jonah Blasiak, Josh Grochow, Kevin Pratt
Main reference Jonah Blasiak, Henry Cohn, Joshua A. Grochow, Kevin Pratt, Chris Umans: "Matrix multiplication via matrix groups", CoRR, Vol. abs/2204.03826, 2022.
URL http://dx.doi.org/10.48550/arXiv.2204.03826

Cohn and Umans proposed a group-theoretic approach to bounding the exponent of matrix multiplication. Previous work within this approach ruled out certain families of groups as a route to obtaining $\omega = 2$, while other families of groups remain potentially viable. In this work we turn our attention to matrix groups, whose usefulness within this framework was relatively unexplored.

We first show that finite groups of Lie type cannot prove $\omega = 2$ within the group-theoretic approach. This is based on a representation-theoretic argument that identifies the second-smallest dimension of an irreducible representation of a group as a key parameter that determines its viability in this framework. Our proof builds on Gowers' result concerning product-free sets in quasirandom groups. We then give another barrier that rules out certain natural matrix group constructions that make use of subgroups that are far from being self-normalizing.

Our barrier results leave open several natural paths to obtain exponent 2 via matrix groups. To explore these routes we propose working in the continuous setting of Lie groups, in which we develop an analogous theory. Obtaining the analogue of exponent 2 in this potentially easier setting is a key challenge that represents an intermediate goal short of actually proving $\omega = 2$. We give constructions in the continuous setting, which evade our two barriers, and indeed are "best-possible" in a precise sense. We then describe a new ingredient – "separating polynomials" – which allow us to recover a full-fledged framework yielding actual algorithms in the Lie setting (rather than constructions whose interest is only by analogy).

3.24 Almost Chor-Goldreich Sources and Adversarial Random Walks

David Zuckerman (University of Texas – Austin, US)

Joint work of Dean Doron, Dana Moshkovitz, Justin Oh, David Zuckerman
Main reference Dean Doron, Dana Moshkovitz, Justin Oh, David Zuckerman: "Almost Chor-Goldreich Sources and Adversarial Random Walks", Electron. Colloquium Comput. Complex., Vol. TR22-103, 2022.
URL https://eccc.weizmann.ac.il/report/2022/103

A Chor-Goldreich (CG) source is a sequence of random variables where each has min-entropy, even conditioned on the previous ones. We extend this notion in several ways, most notably allowing each random variable to have Shannon entropy conditioned on previous ones. We achieve pseudorandomness results for Shannon-CG sources that were not known to hold even for standard CG sources, and even for the weaker model of Santha-Vazirani sources.

Specifically, we construct a deterministic condenser that on input a Shannon-CG source, outputs a distribution that is close to having constant entropy gap, namely its min-entropy is only an additive constant less than its length. Therefore, we can simulate any randomized algorithm with small failure probability using almost CG sources with no multiplicative slowdown. This result extends to randomized protocols as well, and any setting in which we cannot simply cycle over all seeds, and a "one-shot" simulation is needed. Moreover, our construction works in an online manner, since it is based on random walks on expanders.

Our main technical contribution is a novel analysis of random walks, which should be of independent interest. We analyze walks with adversarially correlated steps, each step being entropy-deficient, on good enough lossless expanders. We prove that such walks (or certain interleaved walks on two expanders) accumulate entropy.

Participants

- Eric Allender
Rutgers University – Piscataway, US
- Markus Bläser
Universität des Saarlandes – Saarbrücken, DE
- Andrej Bogdanov
The Chinese University of Hong Kong, HK
- Peter Bürgisser
TU Berlin, DE
- Prerona Chatterjee
The Czech Academy of Sciences – Prague, CZ
- Arkadev Chattopadhyay
TIFR – Mumbai, IN
- Gil Cohen
Tel Aviv University, IL
- Julian Dörfler
Universität des Saarlandes – Saarbrücken, DE
- Stephen A. Fenner
University of South Carolina – Columbia, US
- Michael A. Forbes
University of Illinois – Urbana-Champaign, US
- Lance Fortnow
Illinois Institute of Technology – Chicago, US
- Anna Gál
University of Texas – Austin, US
- Frederic Green
Clark University – Worcester, US
- Rohit Gurjar
Indian Institute of Technology – Mumbai, IN
- William Hoza
University of California – Berkeley, US
- Christian Ikenmeyer
University of Liverpool, GB
- Valentine Kabanets
Simon Fraser University – Burnaby, CA
- Pascal Koiran
ENS – Lyon, FR
- Antonina Kolokolova
University of Newfoundland – St. John's, CA
- Michal Koucký
Charles University – Prague, CZ
- Sophie Laplante
University Paris Diderot, FR
- Nutan Limaye
IT University of Copenhagen, DK
- Meena Mahajan
The Institute of Mathematical Sciences – Chennai, IN
- Rafael Mendes de Oliveira
University of Waterloo, CA
- Ryan O'Donnell
Carnegie Mellon University – Pittsburgh, US
- Natacha Portier
ENS – Lyon, FR
- Noga Ron-Zewi
University of Haifa, IL
- Rahul Santhanam
University of Oxford, GB
- Nitin Saxena
Indian Institute of Technology Kanpur, IN
- Rocco Servedio
Columbia University – New York, US
- Ronen Shaltiel
University of Haifa, IL
- Amir Shpilka
Tel Aviv University, IL
- Srikanth Srinivasan
Aarhus University, DK
- Amnon Ta-Shma
Tel Aviv University, IL
- Jacobo Torán
Universität Ulm, DE
- Christopher Umans
California Institute of Technology – Pasadena, US
- Mary Wootters
Stanford University, US
- David Zuckerman
University of Texas – Austin, US
- Jeroen Zuiddam
University of Amsterdam, NL

Report from Dagstuhl Seminar 22372

Knowledge Graphs and their Role in the Knowledge Engineering of the 21st Century

Paul Groth*[1], Elena Simperl*[2], Marieke van Erp*[3], and Denny Vrandečić*[4]

1 **University of Amsterdam, NL.** p.t.groth@uva.nl
2 **King's College London, GB.** elena.simperl@kcl.ac.uk
3 **KNAW Humanities Cluster, NL.** marieke.van.erp@dh.huc.knaw.nl
4 **Wikimedia Foundation, San Francisco, US.** denny@wikimedia.org

Abstract

This report documents the programme and outcomes of Dagstuhl Seminar 22372 "Knowledge Graphs and their Role in the Knowledge Engineering of the 21st Century" held in September 2022.

The seminar aimed to gain a better understanding of the way knowledge graphs are created, maintained, and used today, and identify research challenges throughout the knowledge engineering life cycle, including tasks such as modelling, representation, reasoning, and evolution. The participants identified directions of research to answer these challenges, which will form the basis for new methodologies, methods, and tools, applicable to varied AI systems in which knowledge graphs are used, for instance, in natural language processing, or in information retrieval.

The seminar brought together a snapshot of the knowledge engineering and adjacent communities, including leading experts, academics, practitioners, and rising stars in those fields. It fulfilled its aims – the participants took inventory of existing and emerging solutions, discussed open problems and practical challenges, and identified ample opportunities for novel research, technology transfer, and inter-disciplinary collaborations. Among the topics of discussion were: designing engineering methodologies for knowledge graphs, integrating large language models and structured data into knowledge engineering pipelines, neural methods for knowledge engineering, responsible use of AI in knowledge graph construction, other forms of knowledge representations, and generating user and developer buy-in. Besides a range of joint publications, hackathons, and project proposals, the participants suggested joint activities with other scientific communities, in particular those working on large language models, generative AI, FAccT (fairness, accountability, transparency), and human-AI interaction.

The discussions were captured in visual summaries thanks to Catherine Allan – you can find more about her work at https://www.catherineallan.co.uk/. The summaries are arrayed throughout this report. Lastly, knowledge about the seminar is captured in Wikidata at https://www.wikidata.org/wiki/Q113961931

Seminar September 12–14, 2022 – http://www.dagstuhl.de/22372

2012 ACM Subject Classification Computing methodologies → Knowledge representation and reasoning; Computing methodologies → Natural language processing; Computing methodologies → Machine learning; Information systems → Information retrieval; Computing methodologies → Ontology engineering; Computing methodologies → Reasoning about belief and knowledge; Human-centered computing → Collaborative and social computing theory, concepts and paradigms

Keywords and phrases Dagstuhl Seminar

Digital Object Identifier 10.4230/DagRep.12.9.60

* Editor / Organizer

Dagstuhl Seminar 22372, *Dagstuhl Reports*, Vol. 12, Issue 9, pp. 60–120
Editors: Paul Groth, Elena Simperl, Marieke van Erp and Denny Vrandečić
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Marieke van Erp (KNAW Humanities Cluster – Amsterdam, NL)
Elena Simperl (King's College London – London, GB)
Denny Vrandečić (Wikimedia Foundation – San Francisco, US)
Paul Groth (University of Amsterdam, NL)

Knowledge engineering has changed dramatically in the last twenty years. When the organisers of this seminar were starting out, it used to be about gathering highly curated knowledge from experts and encoding it into computational representations in knowledge bases. It was primarily a manual process, focusing more on how knowledge was structured and organised, for instance, as schemas or ontologies, and less on tying in existing data into that process. The results were used in expert systems and required considerable up-front investment. Today, knowledge base construction is a largely automatic process with human-in-the-loop. Owing to greater availability of data in different modalities and to advances in data management, machine learning, and crowdsourcing, knowledge bases today incorporate large amounts of knowledge. Provided access to data and (off-the-shelf) AI capabilities, an organisation can create a large knowledge base at a fraction of the costs from decades ago. It's for these reasons that we see knowledge bases, in particular in the form of knowledge graphs, routinely applied in anything from search and intelligent assistants to digital twins, supply chain management, and legal compliance. Many socio-technical challenges remain, which the seminar aimed to address with a mix of invited talks, deep-dives, and small-group workshops as following:

Landscape review: as the field has changed so much, both in research and practices, it was important to take inventory of approaches, methods, techniques, and tools by analysing real-world case studies where knowledge bases and knowledge graphs are created and used. Participants reflected on core lessons learned, knowledge gaps, and opportunities to create and maintain knowledge graphs at scale in various domains.

The knowledge graph life cycle: participants discussed extant knowledge engineering pipelines and identified gaps and connections between knowledge sources and methods and tools used in the construction and maintenance of knowledge graphs, including large language models and generative AI systems. There was consensus that we need a sustained effort to update and upgrade classical ontology engineering methodologies and develop a prototype infrastructure to make the most of the latest neurosymbolic technologies and tools. One specific challenge identified during the seminar was around taking knowledge engineering and knowledge graphs beyond structured data e.g., tables and information extraction from text to other modalities.

Using AI responsibly: as knowledge graph construction is slowly but surely embracing more and more sophisticated AI capabilities to scale, it is critical that processes and outcomes are aligned with fairness, accountability, and transparency guidance and standards. Solutions need to consider a range of end-users and stakeholders, including those that are unique to knowledge engineering settings such as domain experts, information scientists and librarians, and knowledge graph developers. Participants discussed the need for setting up task-based studies and in-depth analyses of human-centric challenges, and for developing bespoke explainability solutions and bias and fairness assessments.

Knowledge and technology transfer: knowledge graphs and knowledge engineering do not exist in isolation. From a research point of view, participants suggested activities to build capabilities to use the latest neurosymbolic technologies and tools in knowledge graph

construction, including tutorials, workshops, and hackathons, and to jointly develop frameworks and methodologies. From an application point of view, it was recognised that there is a need to promote knowledge graphs to the wider developer community and communicate their benefits, for instance, alongside neural methods.

2 Table of Contents

3 Overview of Talks

3.1 Day 1: History, Practices, Lessons Learned

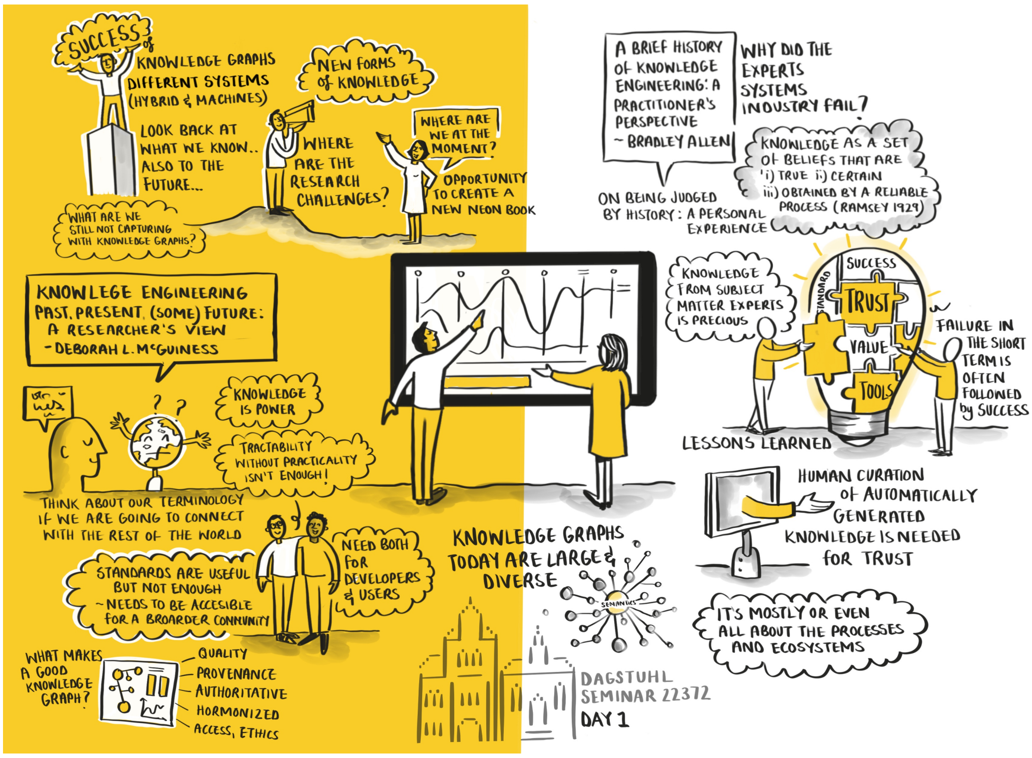


Figure 1 A History of Knowledge Engineering.

3.2 A Brief History of Knowledge Engineering: A Practitioner's Perspective

Bradley P. Allen (Merit International, Inc. – Millbrae, US, bradley.p.allen@gmail.com)

3.2.1 An Approach to the History of Knowledge Engineering

This talk is an attempt to outline the evolution of the discipline of knowledge engineering practice over time, draw some lessons from that evolution, and then raise a number of questions that this seminar is in a position to address, towards the end of defining paths forward for knowledge engineering with knowledge graphs in the 21st century.

Knowledge engineering as a discipline has changed considerably since its initial flowering during the period associated with expert systems development during the nineteen-eighties. If we define knowledge as a set of beliefs that are "(i) true, (ii) certain, [and] (iii) obtained by a reliable process" [2], we can further define knowledge engineering as the discipline of building and maintaining processes that produce knowledge. We argue that this gives us a framework to understand the history of knowledge engineering to date through the evolution of stated requirements for such knowledge production processes.

3.2.2 Seventy Years of Evolving Requirements

During the period from 1955 to today, we can identify four distinct periods, each of which began with the addition of a new set of requirements for knowledge production processes intended to address perceived shortcomings of systems developed during the preceding period (see fig. 2).

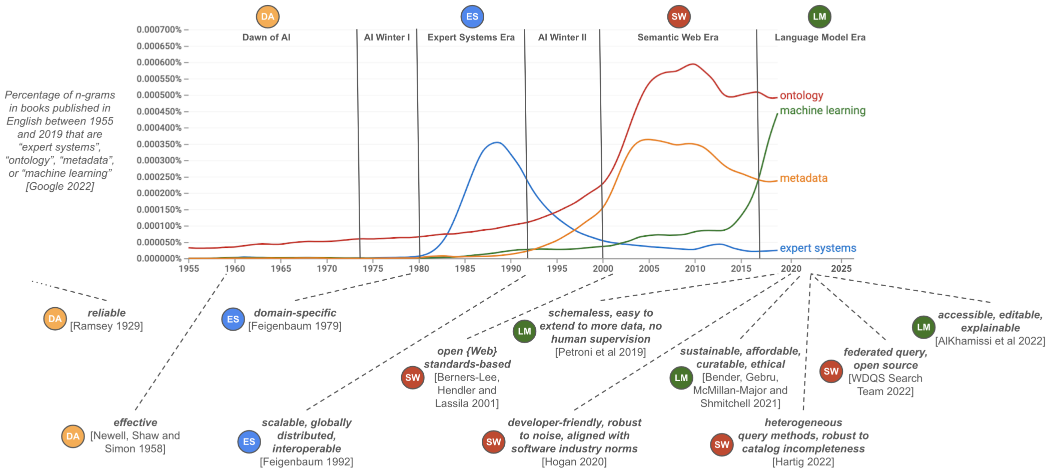


Figure 2 Seventy years of evolving requirements for knowledge production processes [1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13].

3.2.2.1 The Dawn of AI

Starting from Ramsey's simple requirement that such processes be reliable, some of the earliest work in AI identified the additional requirement that such processes also be effective, in the sense that they complete in a reasonable amount of time [3]. Newell and Simon were optimistic about the potential of goal-directed search using heuristics as a general approach to problem solving to be useful for practical applications, but by the beginning of the nineteen-seventies, it was clear that such systems were difficult to use in developing applications that were recognizably more than just toy tasks.

3.2.2.2 The Expert Systems Era

By the mid-seventies, having been deeply involved in attempting to apply Newell and Simon's model, Feigenbaum became convinced that automating knowledge production required a domain-specific focus to succeed [4]. His evangelism of knowledge engineering (a term he was instrumental in propagating the use of) engendered a period of intense activity in the construction of expert systems for the purposes of decision support in business enterprise settings. By the early nineteen-nineties, however, Feigenbaum and others acknowledged that the expert systems approach resulted in systems that were brittle and hard to maintain. Without abandoning his requirement that knowledge production be domain-specific in application focus and thus heavily dependent on subject matter expertise, he argued that future knowledge-based systems also be scalable, globally distributed, and interoperable to address these shortcomings [5]. At that point in time, however, there was no consensus about how such requirements could be addressed, but in retrospect, one can argue that in [5] Feigenbaum anticipated several aspects of what several years later would come to be known as the World Wide Web.

3.2.2.3 The Semantic Web Era

With the establishment of the Web and the emergence of Web architectural principles, Berners-Lee argued for a "Web of Data" based on linked data principles, standard ontologies, and data sharing protocols that not only provided an implementation of Feigenbaum's requirements, but with a single stroke established Web-centric open standards that anyone could adopt [6]. The subsequent twenty years witnessed the development of a globally federated open linked data "cloud", as well as the refinement of techniques for ontology engineering (i.e., the development and publishing of shared data schemas with semantics using linked data principles). Enterprises in particular found better value propositions for the use of such techniques toward the improvement of access and discovery of Web content and data, in contrast to the automation of decision making that was the primary value proposition for knowledge-based systems during the expert systems era [8]. However, while progress was made in building systems based on such principles, general adoption of specific principles advocated for by the semantic web community by the broader community of software developers and web application designers was slow, leading to semantic web researchers to identify additional requirements for broader adoption, for example that the core tools and standards used in semantic web application be more developer-friendly and more directly aligned with software industry norms, and that measures be taken to make federated open data more robust to noise [9]. Addition focus on improving the effectiveness of federated query, which proved hard to scale, and on handling the problem of data catalog incompleteness, all the while maintaining the practical benefits of open source and open standards led to new requirements towards those ends [11, 12].

3.2.2.4 The Language Model Era

The success of connectionist methods arising from the proliferation of graphical processing hardware for matrix arithmetic and concurrent innovations in neural network architectures [14] has led to a new set of possibilities for the production of knowledge graphs. This is an area that at the time of this writing is difficult to summarize due to the rapid rate of research publication, but two perspectives on the relation between language models and knowledge bases have emerged over the last several years. First, that the language model can serve directly as a knowledge base that is queryable using natural language prompts; secondly, that a language model can be a useful component in a knowledge production workflow that combines techniques based on the use of language models together with more traditional symbolic approaches [13]. Regardless of which of these perspectives is most valid, both are sure to result in work that will have an impact on the ability to produce and use knowledge graphs in knowledge engineering work moving forward.

3.2.3 Seventy Years of Lessons Learned

This decades-long evolution of knowledge engineering, bringing us to the current situation where the production of knowledge as knowledge graphs is gaining industrial acceptance at the same time as an entirely new paradigm of knowledge production through the use of large language models may be beginning to emerge, provides us with lessons learned along the way. In addition to these lessons from the history of knowledge engineering, it is also worth noting as well that this period also saw the evolution of software engineering best practices and patterns, as well as the emergence of both the software products and Internet services industries, and that many of the lessons learned in those contexts can be applied to improve the practice of knowledge engineering, particularly from a methodological perspective. Below we call out seven such lessons.

3.2.3.1 Manually Authored Knowledge from Subject Matter Experts is Precious

The digital library community has long argued that manually-created metadata is of vital importance in the creation of robust search resources, and much of the development of the World Wide Web (and continuing on to the Open Linked Data cloud) was informed by that assumption [15]. The effort of designing ontologies, taxonomies, and entity and relationship data has historically depended on expensive, labor-intensive manual effort. In many respects, the work generated by this labor is irreplaceable, and must be treated with respect. Acknowledging the essential nature of these foundational knowledge resources is not only important for an understanding of how knowledge is produced, but also to gain a clear understanding of the labor economics involved in these processes, from both a cost and an ethical perspective [16].

3.2.3.2 Automatic Generation of Knowledge is Needed for Scale

Automatic generation of knowledge graphs are needed to scale the extraction and production of knowledge. With the emergence of statistical natural language processing capable of dealing with training corpora on the order of half a trillion tokens, text available in massive curated corpora or the Web at large are now a effective source of manually authored knowledge. The sheer amount of human-authored content across the Web and in hand-crafted ontologies for linked open data require the automation of the knowledge graph creation process. Automation can also reduce time-to-market and enable larger and more up-to-date knowledge graphs to be generated, making knowledge graphs more accessible and useful.

3.2.3.3 Human Curation of Automatically Generated Knowledge is Needed for Trust

While automated systems can produce large knowledge graphs, they are limited in their ability to interpret and contextualize this output (though with the advent of language models this may be changing). Human curation is needed to verify that the knowledge graph production process is accurate. This process of verification is a necessary condition in many applications for users to be able to trust the knowledge and use it effectively. Additionally, human curation can provide insights into the data that automated systems may miss, such as potential ethical implications, biases, and areas for improvement.

3.2.3.4 User Buy-in to the Value Proposition is Essential

The failure of expert systems in delivering value to commercial enterprises can be viewed as an example of the failure of software product developers to understand users' needs and to effectively communicate value propositions to their users [17]. In striving to replace human decision makers, knowledge engineering in the expert systems era was attempting to solve a problem that ultimately turned out to be not of great importance to many enterprises. The Semantic Web era saw a realignment of knowledge engineering with user values by developing knowledge graphs that supported the needs of organizations to develop ways of guiding their users to the right sources of knowledge and information.

3.2.3.5 Developer Buy-in is Critical for Adoption of Standards and Tools

Software developer buy-in is critical for the successful adoption of standards and tools in any given field. Without their buy-in, the standards and tools will not be leveraged correctly by developers, or at all. We see this in the controversies around the adoption of Semantic Web standards and tools. In part, some developers are hesitant to use these standards and tools due to limited support by commercial vendors, and the lack of resources to help them understand the technology and how to incorporate it into their projects. Without the buy-in of software developers, knowledge graph standards and tools will continue to lack widespread adoption. In instances where commercially-useful enterprise knowledge graphs have been produced, such as Google Knowledge Graph [18], Amazon's Product Graph [19], and Wikidata [20], this has led to a reliance on custom architectures and approaches, which does not address the requirements of interoperability and federation of knowledge resources identified by Feigenbaum and Berners-Lee [5, 6].

3.2.3.6 Open Source/Access/Standards are a Huge Accelerant for Adoption

Open source/access/standards promote adoption because they make it easier to share, collaborate, and replicate research. For example, the pace of research and development in the area of large language models has been greatly accelerated by open source initiatives such as GPT-3 [21], TensorFlow [22], and PyTorch [23]. Initiatives such as these have provided researches in both academic and industrial contexts quick and easy access to cutting-edge tools and datasets, which in turn allows researchers to share, replicate, and collaborate on research quickly and easily through open access publication platforms such as Arxiv [24]. As a result, researchers are able to develop more sophisticated models and applications faster than ever before; this is in contrast with the experience of knowledge engineering in the expert systems era, which was heavily dependent on proprietary, closed source tools and technologies, and hence compromised with respect to the speed of innovation and technology transfer.

3.2.3.7 Failure in the Short Term is Often Followed by Success in the Long Term

It is easy to be disillusioned by the inability to deliver clear benefits out of the early adoption of technologies that initially seemed to carry significant promise. But often that perception of failure is due to insufficient time yet invested in working through the challenges of deployment and adoption. The history of speech recognition is a wonderful example of this. The initial approaches taken by participants in the ARPA Speech Understanding Research Project of the mid-nineteen-seventies laid the groundwork for much of what has come to be the statistical and neural language processing technologies approaching universal adoption today, at levels of accuracy barely dreamed of by the researchers of the time. At the conclusion of that effort, however, the evaluation of the project's result was decidedly mixed, with some expert evaluators arguing that the effort had in fact been a step backwards for the research area [25]. This example argues for patience in the effort to demonstrate the benefits of the use and application of knowledge graphs in knowledge engineering.

3.2.4 The Road Ahead: Questions for the Seminar

This seminar provides us with an opportunity to reflect on the past and come up with a set of goals for future progress towards the continuing evolution of knowledge engineering. Below are five questions that we believe need to be addressed to arrive at a robust set of goals.

In what ways does knowledge engineering deliver value today?
Knowledge graphs have demonstrated their ability to improve knowledge access, knowledge discovery, and heterogeneous data integration. But in many respects these are incremental improvements over what has been accomplished with software engineering in general. Can we identify economically and societally important problems either cannot be solved

without knowledge engineering, or are best solved with it? These problems will give us the basis for reinforcing the case for the benefits of knowledge engineering, that can be used to drive further adoption.

What should be the requirements for knowledge production processes?
Best practice in software product development requires us to clearly establish that our technology choices are properly motivated by our users' needs. What are we to carry forward from the cumulative sum of requirements articulated in the body of worked referenced in fig. 2 above? The requirements for knowledge graph production processes should include capabilities for data integration from multiple structured sources, data quality checks, entity resolution, merges and links, query optimization, and natural language processing. Moreover, the production processes should be automated to enable efficient updates and maintenance of the knowledge graph. Finally, the production process should incorporate mechanisms for security and privacy, as well as access control mechanisms to ensure that the data stays secure and only authorized users have access. It is worth observing that many of these issues have been explored to date in the more generic context of data engineering and data science architectures and platforms. To what extent does knowledge engineering add value to those architectures and platforms, and how current knowledge engineering and knowledge graphs tools and standards can be best integrated with them?

Why do we believe that knowledge graphs are a key enabling technology?
A fundamental premise of this seminar is there is a consensus that knowledge graphs are the preferred representation for knowledge for knowledge engineering. What evidence do we have for this assertion? Anecdotally, there is a better developer experience associated with the use of graphs as opposed to, e.g., rules, but what evidence has been gathered to support this view?

What other enabling technologies are there, and how do they interact with knowledge graphs?
Large language models show early promise as a enabling technology that can significant improve and complement knowledge graphs. Can they be harnessed to this end, or do they instead they present an alternative approach to knowledge engineering? In addition, graph databases are necessary for the storage and querying of knowledge graphs, but there is a bifurcation within the community between the use of RDF graphs and labeled property graphs to represent knowledge graphs. How can we reconcile these two approaches (for example, as described in [26])?

How can we convince people that knowledge graph engineering is mainstream software engineering?
Finally, and perhaps most importantly, the majority of software engineering efforts today do not involve the use of knowledge engineering techniques, even in use cases where knowledge engineers can see clear benefits to be gained in their use. Knowledge engineering is still a niche skill set that is unfamiliar to most practicing software engineers. However, the architectures and methodologies emerging from the commercial applications of machine learning, data science, and data engineering [27] in many ways borrow heavily from those developed to support knowledge engineering. How can we better relate knowledge engineering concepts, tools and methodologies to the industry consensus and ecosystem that has been established for data engineering and data science platforms, and drive mainstream adoption in the future?

References

1 Google Books Ngram Viewer. "percentage of n-grams in books published in english between 1955 and 2019 that are "expert systems", "ontology", "metadata", or "machine learning"". `https://books.google.com/ngrams/graph?content=expert+systems%2Contology%2Cmachine+learning%2Cmetadata&year_start=1955&year_end=2019&corpus=26&smoothing=1`, 2022. Accessed: 2022-08-10.

2 F.P. Ramsey. Knowledge. In *F.P. Ramsey: Philosophical Papers*, pages 110–111. Cambridge University Press, 1929.

3 Allen Newell, John Calman Shaw, and Herbert A Simon. Elements of a theory of human problem solving. *Psychological review*, 65(3):151, 1958.

4 Edward A Feigenbaum. The art of artificial intelligence: Themes and case studies of knowledge engineering. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, volume 2. Boston, 1977.

5 Edward A Feigenbaum. *A personal view of expert systems: Looking back and looking ahead.* Knowledge Systems Laboratory, Department of Computer Science, Stanford ..., 1992.

6 Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):34–43, 2001.

7 Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*, 2019.

8 James Hendler, Fabien Gandon, and Dean Allemang. *Semantic web for the working ontologist: Effective modeling for linked data, RDFS, and OWL.* Morgan & Claypool, 2020.

9 Aidan Hogan. The semantic web: Two decades on. *Semantic Web*, 11(1):169–185, 2020.

10 Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.

11 Olaf Hartig. "reflections on linked data querying and other related topics". `https://olafhartig.de/slides/Slides-DKG-SWSA-Talk.pdf`, 2022. Accessed: 2022-03-17.

12 WDQS Search Team. "wdqs backend alternatives: The process, details and results". `https://www.wikidata.org/wiki/File:WDQS_Backend_Alternatives_working_paper.pdf`, 2022. Accessed: 2022-08-15.

13 Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona Diab, and Marjan Ghazvininejad. A review on language models as knowledge bases. *arXiv preprint arXiv:2204.06031*, 2022.

14 Haohan Wang and Bhiksha Raj. On the origin of deep learning. *arXiv preprint arXiv:1702.07800*, 2017.

15 Jane Greenberg. Big metadata, smart metadata, and metadata capital: toward greater synergy between data science and metadata. *Journal of Data and Information Science*, 2(3):19–36, 2017.

16 Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11):100336, 2021.

17 Adrian Payne, Pennie Frow, and Andreas Eggert. The customer value proposition: evolution, development, and application in marketing. *Journal of the Academy of Marketing Science*, 45(4):467–489, 2017.

18 Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. Industry-scale knowledge graphs: Lessons and challenges: Five diverse technology companies show how it's done. *Queue*, 17(2):48–75, 2019.

19 Xin Luna Dong. Challenges and innovations in building a product knowledge graph. In *Proceedings of the 24th ACM SIGKDD International conference on knowledge discovery & data mining*, pages 2869–2869, 2018.

20 Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

21 Robert Dale. Gpt-3: What's it good for? *Natural Language Engineering*, 27(1):113–118, 2021.

22 Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, 2016.

23 Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

24 Paul Ginsparg. Lessons from arxiv's 30 years of information sharing. *Nature Reviews Physics*, 3(9):602–603, 2021.

25 Wayne A Lea and June E Shoup. Review of the arpa sur project and survey of current technology in speech understanding. Technical report, Office of Naval Research, January 1979. Final Report.

26 Renzo Angles, Harsh Thakkar, and Dominik Tomaszuk. Mapping rdf databases to property graph databases. *IEEE Access*, 8:86091–86110, 2020.

27 Matt Bornstein, Jennifer Li, and Casado. Martin. Emerging architectures for modern data infrastructure. `https://future.com/emerging-architectures-modern-data-infrastructure/`, 2020. Accessed: 2022-12-02.

3.3 Knowledge Engineering Past, Present, (some) Future: A Researcher's View

Deborah L. McGuinnness (Rensselaer Polytechnic Institute – Troy, US, dlm@cs.rpi.edu)

I provide a brief historical perspective on significant research contributions and highlight some key lessons, some of which may be particularly worthy for reflection as we move forward. I begin with expert systems as a foundational motivating area but I also highlight the evolution and contributions from the structured object and ontology communities. I also reflect on the early notion of knowledge engineering as the applied side of artificial intelligence (from Feigenbaum) and present that notion in the grounding of the 21st century environment. I also present a range of characteristics as considerations for evaluating if a knowledge engineered system is "good".

I then present some perspectives on the our current landscape that may be significantly different from the past. These include: much greater need for knowledge graph interoperability (as well, of course, as the needs for compatibility and interoperability with a wide range of ontologies); The very large linked open data world ; the significantly more diverse architectures for hybrid AI systems, with large language models as an increasing component; the increasingly diverse community of co-designers and co-authors of large "smart" data portals. I conclude with a set of driving research questions along with a take home message that process and methodology is becoming even more critical for our field to increase impact and buy-in.

3.4 Automated Knowledge Graph Construction

Lise Stork (Vrije Universiteit of Amsterda, NL, l.stork@vu.nl)

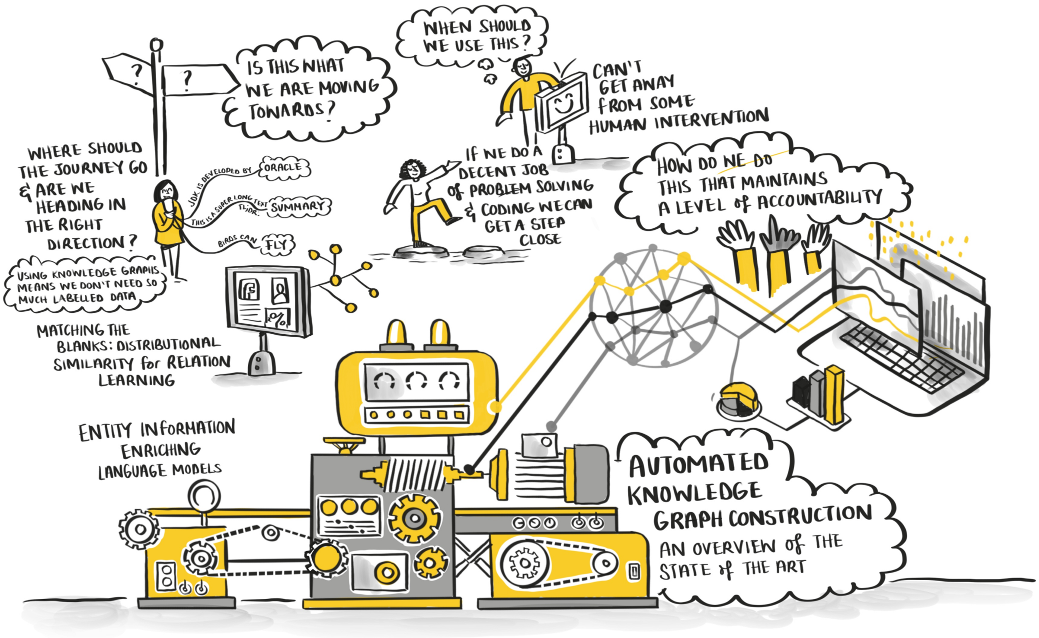


Figure 3 Automated Knowledge Graph Construction.

The talk gave an overview of Knowledge Graph Construction (KGC) methods, four big focus shifts in the development of these methods [1], and sketched some open challenges and future directions for KGC.

Over the past decade, many methods have been proposed for KGC: human-based collaborative or curated approaches in which experts work together to create and curate knowledge graphs, but also automated approaches, classified broadly into approaches that use a predefined schema for extraction, versus open information extraction (IE) [2, 3]. Tasks become increasingly harder (i) with less data available for training, (ii) when relationships are increasingly complicated to extract (binary vs n-ary relations) and (iii) the openness of the task: schema-driven vs open IE.

Methods proposed for KGC have shifted focus from the engineering of features, to the engineering of model architectures, the engineering of tasks or objectives, to the engineering of prompts [1]. Before 2013, domain experts used their expertise about a domain to define salient textual features to be used in an NLP task. After the rise of Neural Nets, the focus shifted towards Architecture Engineering: convolutional neural networks as well as recurrent neural networks such as LSTMs and BiLSTMs were applied in a fully supervised manner, where features were learned jointly with the supervised classification task.

Around 2017, the power of language models increased, mostly due to the discovery that proximity in the input is less important than expected, and context is much better represented when sentences are processed fully, using attention mechanisms [4], instead of sequentially. Such a method at the same time proved easier to train. Since these models, when trained on large corpora, appeared powerful enough to be used in a variety of down-stream tasks, the focus then began to shift towards the fine-tuning of pre-trained LLMs specific tasks [5, 6, 7, 8, 9].

Lastly, since 2019, it was found that these LLMs are interesting to probe, given that they have learned a lot of interesting facts. It was hypothesized that LLMs could serve as knowledge graphs themselves; new ways had to be discovered to query them. Therefore, the focus shifted towards creating, either manually or algorithmically, prompts in order to get out the interesting facts these LLMs models had learned, and 'prompt engineering' became an active field of research [1, 10, 11].

Open challenges that were proposed:

1. how to automatically construct "higher-order or higher-ary knowledge", such as scopes, context, degrees of belief, confidence, and how to evaluate these;
2. how to deal with n to M relations;
3. how do we integrate LMs in the knowledge engineering pipeline;
4. how to deal with bias, trust and control in LMs as KGs; how to add provenance to statements in LMs;
5. how to deal with explainability of answers from prompts;
6. how to update facts in LLMs as KGs;
7. what types of knowledge representations do we extract.

References

1 Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *CoRR*, abs/2107.13586, 2021.

2 Gerhard Weikum, Xin Luna Dong, Simon Razniewski, and Fabian Suchanek. Machine knowledge: Creation and curation of comprehensive knowledge bases. *Foundations and Trends in Databases*, 10(2-4):108–490, 2021.

3 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4):1–37, 2021.

4 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

5 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

6 Shanchan Wu and Yifan He. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 2361–2364, 2019.

7 Christoph Alt, Marc Hübner, and Leonhard Hennig. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction. *arXiv preprint arXiv:1906.08646*, 2019.

8 Xu Han, Tianyu Gao, Yankai Lin, Hao Peng, Yaoliang Yang, Chaojun Xiao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. More data, more relations, more context and more openness: A review and outlook for relation extraction. *arXiv preprint arXiv:2004.03186*, 2020.

9 Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158*, 2019.

10 Dimitrios Alivanistos, Selene Báez Santamaría, Michael Cochez, Jan-Christoph Kalo, Emile van Krieken, and Thiviyan Thanapalasingam. Prompting as probing: Using language models for knowledge base construction. *arXiv preprint arXiv:2208.11057*, 2022.

11 Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*, 2019.

3.5 Day 2 Challenges and Future Directions

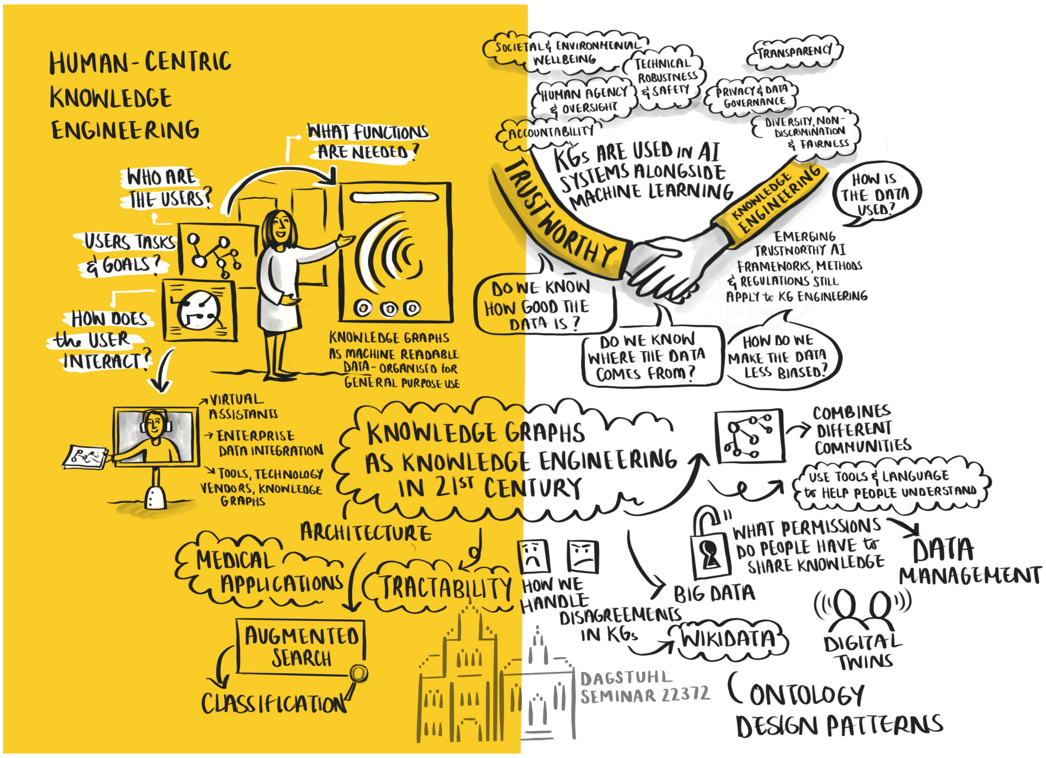


Figure 4 Future Directions: Human-Centric Knowledge Graph Construction.

3.6 Human-Centric Knowledge Engineering: Making Knowledge Engineering Trustworthy

Elena Simperl (King's College London, GB, elena.simperl@kcl.ac.uk)

Knowledge engineering has changed dramatically since I was doing my PhD. It was always meant to be and has remained a challenging, inter-disciplinary subject – the question of how to encode a domain in a computational representation will always be non-trivial and will require some form of human-in-the-loop. The field has advanced considerably: the knowledge bases we build today are much larger and more complex than twenty years ago, there are a range of technologies and end-user tools to support standard tasks, as well as several notable open-source projects delivering large knowledge graphs that can bootstrap applications without massive up-front investment. And yet, our understanding of user-centric aspects of knowledge engineering remains limited. The reasons for this are as often sociotechnical, but the result is clear: we are not (yet) in a position to fully answer questions like these:

- Who are the users?
- What are the users' tasks and goals?
- How does a user interact with the knowledge graph?
- What are the users' experience levels with it, or with similar environments?
- What functionalities does the user need?
- What additional information might the users need, and in what form do they need it?
- How does the user think knowledge engineering tools should work?
- Is the user multitasking? Are they working on a mobile phone, desktop computer etc?
- Does the interface utilise different input modes, such as touch, speech, gestures or orientation?
- How can we support multi-disciplinary teams?
- How can we support remote work, decision making, conflict resolution?

Answering such questions will require studies of specific knowledge engineering projects or tool environments, but would deliver invaluable insights to improve both user experience and knowledge graph outcomes. In time, it would lead to a culture of user-centric design and to research advances that are applicable beyond knowledge engineering contexts. with the recent changes, it is also worth revisiting the surveys and handbooks written a decade or so ago to deliver up-to-date comparative surveys and tool evaluations, relevant to how knowledge graphs are built today in terms of scale, complexity, and degree of automation.

Using automation, particularly the latest AI capabilities, raises interesting human-centric challenges, which other communities such as natural language processing and computer vision are starting to explore. These are grouped under the banner of trustworthy AI, which is concerned with questions of transparency, accountability, fairness, human agency and oversight, and sustainability when AI is used by (or impacts) different groups of people. There is a large body of work happening right now to define frameworks, guidance, regulation, and standards for trustworthy AI[1] – for instance, the European Commission has proposed seven dimensions for designing AI systems, shown in Figure 5 and there are many standardisation activities at national and international levels (e.g. ISO).[2]

---

[1] For an overview, see e.g., OECD AI Policy Observatory, `https://oecd.ai/`, visited in September 2022
[2] See a list of AI-related ISO standards at `https://www.iso.org/committee/6794475/x/catalogue/`, visited in December 2022

Figure 5 Dimensions of Trustworthy AI according to the European Commission.

Ongoing research in trustworthy AI proposed a core set of methods and best practices to meet the regulatory requirements of trustworthy AI systems [1]. These include factsheets [2], model cards [3], canvases,[3] explainability methods [4] and fairness and debiasing methods [5]. Knowledge graphs and knowledge graph construction systems need to build on these works to ensure the processes we follow and their outcomes can be trusted by end-users and stakeholders. This includes: the domain expert or business analyst involved in knowledge acquisition, the knowledge engineer building the knowledge graph construction pipeline, the crowd worker labelling training data, the developer of downstream applications using the knowledge graph, for instance in the form of embeddings [6] and the users of those applications.

In my team we undertook research into knowledge communities such as DBpedia and Wikidata to understand how different components of trust emerge and propose socio-technical methods to improve the quality of the knowledge graph and make it more complete, up-to-date and less biased. The research pursued questions such as:

Do we know how good the data in the knowledge graph is? In [7] we surveyed 28 quality approaches and methods for Wikidata and proposed a joint framework.

Do we know where the data comes from In [8] and [9] we proposed an AI architecture with human-in-the-loop to assess quality of triple provenance across five languages.

Do we know how to audit our data to make it less biased? In [10] we proposed a method to detect content gaps in open knowledge graphs and applied it to three main types of biases: gender, recency, and socioeconomic biases.

---

3 For example, the data ethics canvas of the Open Data Institute in the UK, `https://theodi.github.io/interactive-data-ethics-canvas/`, visited December 2022

Do we know how the data came about? Even when automation is at play, knowledge construction is a social process. For this reason, we analysed the link between quality of knowledge graph entities and the make-up of the editor teams that worked on those entities [11, 12]

Do we know how the data is used? One application of knowledge graphs is natural language processing. In [13] we evaluated a natural language generation system that takes Wikidata triples and creates Wikipedia articles in different languages. We ran user studies to understand if and when the presence of automation changes editor perceptions and practices.

Knowledge engineering remains as exciting of a field as ever, with a range of human-centric challenges that cannot and should not be overlooked given the advanced in the field and in related fields such as machine learning, natural language processing, and computer vision. Looking ahead, I would like to see more work into establishing user-centric design and empirical methods more firmly into the ways we build our tools and applications. In particular, we need to ensure the way we make knowledge graphs today is interpretable and trustworthy, and ongoing research in the area of responsible AI, including transparency, accountability, and fairness can deliver new impulses for interdisciplinary research.

References

1. Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 624–635, 2021.
2. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.
3. Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.
4. Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. *Explainable AI: interpreting, explaining and visualizing deep learning*, volume 11700. Springer Nature, 2019.
5. Alexandra Chouldechova and Aaron Roth. A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, 63(5):82–89, 2020.
6. Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.
7. Alessandro Piscopo and Elena Simperl. What we talk about when we talk about wikidata quality: a literature survey. In *Proceedings of the 15th International Symposium on Open Collaboration*, pages 1–11, 2019.
8. Alessandro Piscopo, Lucie-Aimée Kaffee, Chris Phethean, and Elena Simperl. Provenance information in a collaborative knowledge graph: an evaluation of wikidata external references. In *International semantic web conference*, pages 542–558. Springer, 2017.
9. Gabriel Amaral, Alessandro Piscopo, Lucie-Aimée Kaffee, Odinaldo Rodrigues, and Elena Simperl. Assessing the quality of sources in wikidata across languages: a hybrid approach. *Journal of Data and Information Quality (JDIQ)*, 13(4):1–35, 2021.
10. David Abián, Albert Meroño-Peñuela, and Elena Simperl. An analysis of content gaps versus user needs in the wikidata knowledge graph. In *International Semantic Web Conference*, pages 354–374. Springer, 2022.

11 Alessandro Piscopo, Chris Phethean, and Elena Simperl. What makes a good collaborative knowledge graph: group composition and quality in wikidata. In *International Conference on Social Informatics*, pages 305–322. Springer, 2017.
12 Alessandro Piscopo and Elena Simperl. Who models the world? collaborative ontology creation and user roles in wikidata. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–18, 2018.
13 Lucie-Aimée Kaffee, Pavlos Vougiouklis, and Elena Simperl. Using natural language generation to bootstrap missing wikipedia articles: A human-centric perspective. *Semantic Web*, 13(2):163–194, 2022.

3.7 Everything is Expensive

Denny Vrandečić (Wikimedia Foundation – San Francisco, US, denny@wikimedia.org)

The creation of ontologies is expensive. It is hard to achieve the initial buy-in from developers, and often developers are mandated into the use of a specific ontology. This often leads to less than enthusiastic support. The ontologist often has to "cat-herd" technical leads and product managers across several departmenets and organizations. Generating agreements often takes a long time and many meetings with discussions and not directly tangible outcomes. And meetings are expensive. Even in Wikidata, property creation is one of the major bottlenecks.

The trouble with triples. Single triples cannot express complex statements (known as n-ary statement, but also not frames or events). So patterns of triples are required to represent such complex statements. But for users of a triple store, these are atomic statements. Tools, user experience, metrics, processes all become much more complex and expensive due to this mismatch.

Will large language models lead to cheap ontologies? It is expected that a technology-driven companies there will be an initial surplus of trust in large language models, which may backfire when these models lead to expensive errors. On the other side, technology-skeptical organizations such as in journalism or in Wikidata, may start with a deficit of trust, which may hamper the usage of these technologies. The biggest problem is actually the same as with handmade ontologies: how to make people understand, commit to, and trust the created ontologies? The cost is not in creating the ontologies, but the agreement.

Knowledge acquisition is expensive. Once we have the ontology, how do we efficiently populate it? How do we let humans efficiently check a large amount of data before product launch? Important are the possibility to sample parts of the knowledge graph, which are either particularly impactful or particularly interesting. Rules have been very good at discovering inconsistencies and incompleteness. Machine learning has also been well used to suggest anomalies.

Knowledge maintenance is expensive. Now that we have large lists of inconsistencies and incomplete data, what do we do with that? We also need to keep and maintain metadata about exceptions (because the world is always more complex than your rules). If we allow for feedback from end users, how do we capture and classify that feedback? If we don't allow for feedback, what is the point of the knowledge graph? How do we channel feedback in order to maintain the knowledge?

3.8 Tools and User Experience for KG Engineering

Filip Ilievski (University of Southern California – Marina del Rey, US, ilievski@isi.edu)

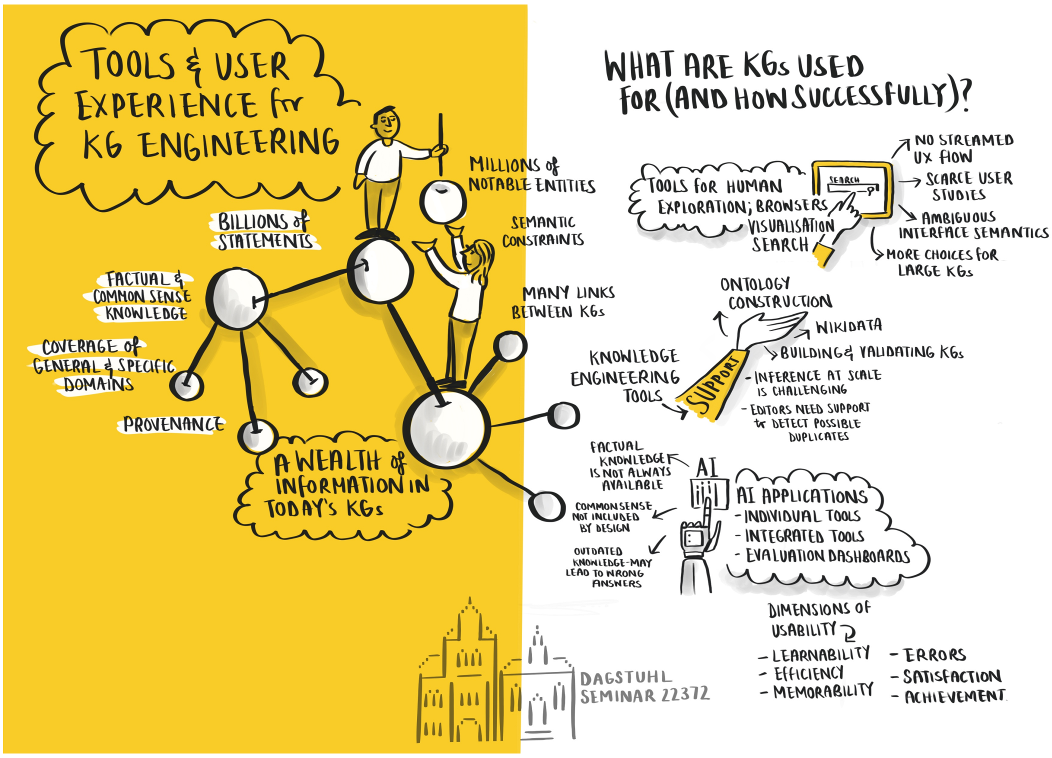


Figure 6 Tools and User Experience.

Today's knowledge graphs contain a wealth of information. For instance, Wikidata has billions of statements for millions of notable entities, with recorded provenance and semantic constraints. In this talk, I ask the question: What are KGs used for and how successfully? I consider three user profiles: an end user, an AI/DS engineer, and a knowledge engineer.
1) An end user might use knowledge graphs to explore knowledge, browse answers to their questions, or develop new ideas. These tasks can be supported by browsers, visualization tools, or tools for textual and faceted search. Key pain pointers from an end-user perspective are the lack of streamed workflow from high- to micro-level, the lack of user studies, the ambiguity of interface semantics, and issues with compositionality and data quality.
2) An AI/DS engineer might use knowledge technologies to perform automatic question answering, recommendation, search, or content enrichment. These tasks can be pursued with a pipeline of existing tools that perform operations like entity linking, semantic similarity, lexicalization, and embedding learning. Integrated tools, databases, or libraries allow developers to perform a set of these operations in the same framework, avoiding the need to compose different toolkits, formats, and standards themselves. Evaluation dashboards for tasks like knowledge graph completion and question answering enable fine-grained auditing of system performance. Pain points for AI engineers include: sparsity of factual and commonsense knowledge, consistency of ontological knowledge, the lack of decision support tooling, and potentially outdated knowledge.

3) A knowledge engineer might contribute or edit new knowledge or provenance, perform semantic modeling and validation, infer new knowledge, or engineer a new knowledge graph. Key tools for knowledge engineers include ontology editors, tooling for Wikidata contributors, and knowledge construction and validation tools. The key pain points for knowledge engineers are that inference at scale is challenging, identity is hard to establish, different is-a flavors are difficult to distinguish, the lack of tool integration, and the lack of user studies and logging practices.

3.9 Social and Technical Biases in Knowledge Graphs

Harald Sack (FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, DE & Karlsruhe Institute of Technology (KIT), DE, harald.sack@fiz-karlsruhe.de)

Figure 7 Biases in Knowledge Graphs.

Knowledge graphs as a key tool for organizing and presenting information in the modern world constitute networks of interlinked data that help us to make sense of the vast amounts of information available to us. Once constructed, knowledge graphs are often considered as "gold standard" data sources that safeguard the correctness of other systems. Thereby, objectivity and neutrality of the represented information have become an important issue. Biases inherent to knowledge graphs may become magnified and spread through knowledge graph based systems. Traditionally, bias can be defined as "*a disproportionate weight in favor* *of or against an idea or thing, usually in a way that is closed-minded, prejudicial, or unfair*"[4]. Taking into account the bias networking effect for knowledge graphs, it is crucial that we acknowledge and address various types of bias already in knowledge graph construction [1].

Biases in knowledge graphs as well as potential means to address them, are different from those in linguistic models or image classification. Knowledge graphs are sparse by nature, i.e. only a small number of triples are available per entity. In difference, linguistic models learn the meaning of a term from its context within large corpora and image classification learns classes from millions of labeled images. Biases in knowledge graphs may originate in the very design of the knowledge graph, in the source data from which it is created (semi-)automatically, and in the algorithms used to sample, aggregate, and process that data. These source biases typically appear in expressions, utterances, and text sources, and can carry over into downstream representations such as knowledge graphs and knowledge graph embeddings. Furthermore, we also have to consider a large variety of human biases, as e.g. reporting bias, selection bias, confirmation bias, overgeneralization, etc.

Biases in knowledge graphs can arise from multiple sources. Data bias occurs already in the data collection process for the knowledge graph or simply from the available source data. Schema bias depends on the chosen ontology for the knowledge graph or simply is already embedded within the used ontologies [1]. Inferential bias might result from drawing inferences on the represented knowledge. Ontologies are typically defined by a group of knowledge engineers in collaboration with domain experts and consequently (implicitly) reflect the worldviews and biases of the development team. Ontologies are also prone to encoding bias depending on the chosen representation language (fragment of description logics). Moreover, biases in knowledge graph embeddings may also arise from the embedding method. Inferential biases in knowledge graphs arise at inferencing level, such as reasoning, querying, or rule learning. A simple example might be the different SPARQL entailment regimes, which in consequence, might be responsible for different results that different SPARQL endpoints deliver despite containing the same knowledge graph.

Collaboratively built knowledge graphs, as e.g. DBpedia or GeoNames also exhibit social bias, often arising from the western centered world view of their main contributors [2]. In addition, some "truths" represented in those knowledge graphs might be considered as controversial or opinionated, which underlines the importance of provenance information.

For knowledge graph embeddings that represent a vector space based approximation of the structural and semantic information contained in a knowledge graph, one of the main sources of bias lies in the sparseness and incompleteness of most knowledge graphs. Thereby, knowledge graph embeddings trained on incomplete knowledge graphs might favour entities for which more information is available [3]. However, if the underlying knowledge graph is biased, then also knowledge graph embeddings trained on this base data. De-biasing of knowledge graph embeddings requires methods for detecting as well as removing bias in knowledge graph embeddings. Depending on the underlying embedding model, this task might become complex and requires finetuning of embeddings with respect to certain sensitive relations [4, 5, 6].

References

1 Krzysztof Janowicz, Bo Yan 0003, Blake Regalia, Rui Zhu, and Gengchen Mai. Debiasing knowledge graphs: Why female presidents are not like female popes. In Marieke van Erp, Medha Atre, Vanessa López, Kavitha Srinivas, and Carolina Fortuna, editors, *Proceedings*

[4] Wikipedia article on Bias, `https://en.wikipedia.org/wiki/Bias`

of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to – 12th, 2018, volume 2180 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

2 Gianluca Demartini. Implicit bias in crowdsourced knowledge graphs. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, page 624–630, New York, NY, USA, 2019. Association for Computing Machinery.

3 Wessel Radstok, Melisachew Wudage Chekol, and Mirko Tobias Schäfer. Are knowledge graph embedding models biased, or is it the data that they are trained on? In *Wikidata@ISWC*, 2021.

4 Joseph Fisher. Measuring social bias in knowledge graph embeddings. *ArXiv*, abs/1912.02761, 2019.

5 Joseph Fisher, Arpit Mittal, Dave Palfrey, and Christos Christodoulopoulos. Debiasing knowledge graph embeddings. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7332–7345, Online, November 2020. Association for Computational Linguistics.

6 Mario Arduini, Lorenzo Noci, Federico Pirovano, Ce Zhang, Yash Raj Shrestha, and Bibek Paudel. Adversarial learning for debiasing knowledge graph embeddings. *ArXiv*, abs/2006.16309, 2020.

4 Lightning Talks

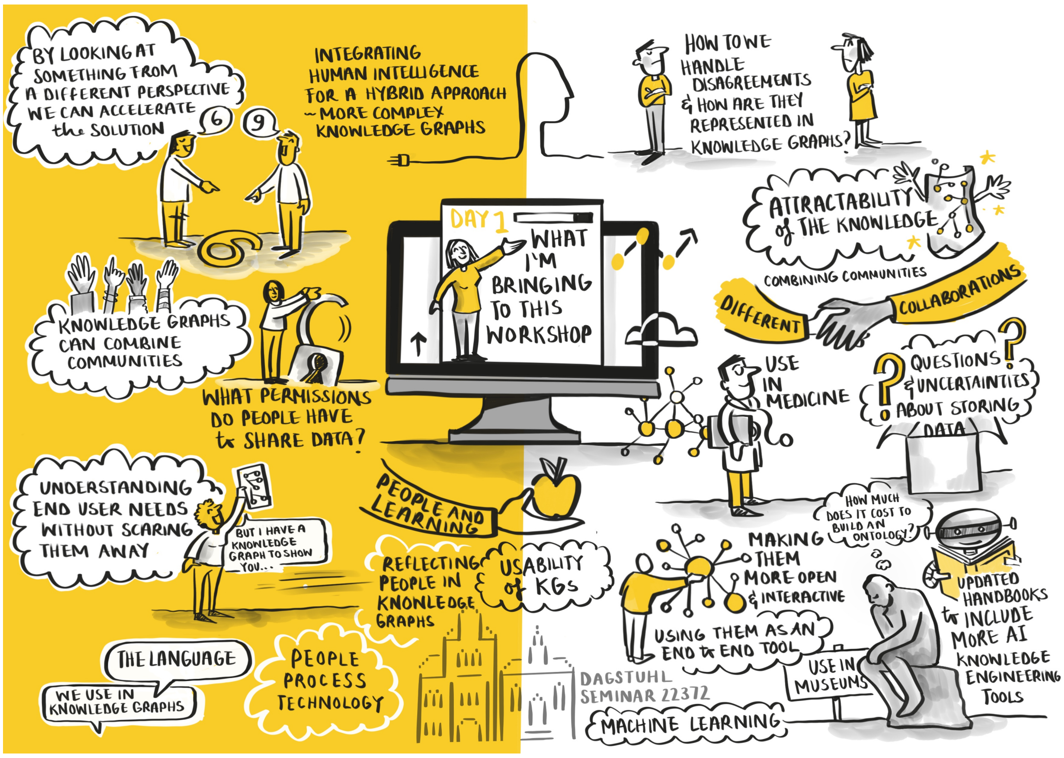


Figure 8 Participant Perspectives – Lightning Talks.

4.1 Organizing Scientific Contributions in the Open Research Knowledge Graph

Sören Auer (TIB - Hannover, DE)

The transfer of scholarly knowledge has not changed fundamentally for many hundreds of years: It is usually document-based-formerly printed on paper as a classic essay and nowadays as PDF. With around 2.5 million new research contributions every year, researchers drown in a flood of pseudo-digitized PDF publications. As a result, research is seriously weakened. We argue for representing scholarly contributions in a structured and semantic way as a knowledge graph. The advantage is that information represented in a knowledge graph is readable by machines and humans. As an example, we give an overview of the Open Research Knowledge Graph (ORKG[5]), a service implementing this approach. For creating the knowledge graph representation, we rely on a mixture of manual (crowd/expert sourcing) and (semi-)automated techniques. Only with such a combination of human and machine intelligence, we can achieve the required quality of the representation to allow for novel exploration and assistance services for researchers. As a result, a scholarly knowledge graph such as the ORKG can be used to give a condensed overview of the state-of-the-art addressing a particular research quest, for example as a tabular comparison of contributions according to various characteristics of the approaches. Further possible intuitive access interfaces to such scholarly knowledge graphs include domain-specific (chart) visualizations or answering natural language questions.

A detailed presentation including screenshots of the demo can be found here.

4.2 dblp as a Knowledge Graph

Marcel R. Ackermann (Schloss Dagstuhl LZI – Trier, DE)

For more than 20 years, a full XML dump of the *dblp computer science bibliography*[6] has been available as open data for download and reuse. Snapshots of the dblp XML dump have been converted to RDF before by members of the community. However, these snapshots are usually severely out of sync with the continuously curated dblp data, in some cases up to several years. To remedy this problem, the dblp team has now started to release its data also in RDF via APIs and as a full dump download. The goal is to provide a semantically rich knowledge graph of bibliographic information that is up to date and in sync with the curated and disambiguated dblp data. Just as with any other data provided by dblp, the RDF data is made available under CC0 1.0 Public Domain Dedication license.

In its initial release, the *dblp knowledge graph*[7] forms a simple person-publication graph, consisting (as of October 2022) of more than 3 million person entities, 6.3 million publication entities, and 340 million RDF triples in total. More than 15 million external resource URIs

---

[5] `https://orkg.org`
[6] `https://dblp.org`
[7] `https://dblp.org/rdf/`

are linked in the data set. Numerous metadata aspects, like journals/conference series or the affiliation of an author, are currently provided only as string literals. Future iterations of the schema will see these and further aspects being added as true entities, together with their own metadata, persistent IDs, and links to external resources. Hence, we don't see the current dblp knowledge graph as final, but rather as a first step in providing the semantics of the dblp dataset in a more structured way. We also aim to provide a proper SPARQL endpoint in the near future.

4.3 Triples are not Enough

Denny Vrandečić (Wikimedia Foundation – San Francisco, US)

Abstract Wikipedia aims to cover the whole breadth of knowledge that is in a usual Wikipedia article. Wikidata cannot comfortably represent the kind of knowledge necessary for the natural language text of such a Wikipedia article. We decided to work with two knowledge representations beyond triples: functions, in order to generate natural language text, and frames, in order to capture n-aries and other complex statements [1].

See [1] for more details.

4.4 Making Knowledge Graph Embeddings a First Class Citizen

Heiko Paulheim (Universität Mannheim, DE)

Knowledge graph embeddings have become a major area in the knowledge graph research landscape. There are quite a few downstream applications which do not consume the knowledge graph per se, but only the embeddings.

At the same time, embeddings are not very well integrated in current tool stacks. In many cases, developers download a dump of a knowledge graph, compute embeddings, and then feed them into the application at hand. Such a model can neither incorporate any knowledge graph dynamics, nor is it suitable if only a small excerpt of a large knowledge graph is of interest for an application at hand. [2].

Services which serve knowledge graph embeddings like KGvec2go [3] are still rare. Moreover, embeddings are rarely integrated with other KG toolstack services, such as query interfaces. For those reasons, if we want to unleash the full potential of knowledge graph embeddings, we have to integrate them more tightly into our current knowledge graph tool stacks.

4.5 Knowledge Graph Completion using Embeddings

Mehwish Alam (FIZ-Karlsruhe, Leibniz Institute for Information Infrastructure, DE & Karlsruhe Institute of Technology, DE)

Knowledge Graphs (KGs) constitute a large network of real-world entities and relationships between these entities. KGs have recently gained attention in many tasks such as recommender systems, question answering, etc. Due to automated generation and open-world assumption, these KGs are never complete. Recent years have witnessed many studies on link prediction using KG embeddings which is one of the mainstream tasks in KG completion. These KG completion methods also include methods for entity type prediction [4], i.e., given the structural, textual, or another kind of information about an entity the task is to predict the type of an entity. Over the past few years, many methods have been proposed that also utilize language models, as well as a few benchmark datasets, have also been proposed [5]. A challenge remains as to how these methods can further be applied to real-world problems such as the biomedical domain, materials sciences, cultural heritage, scholarly data [6], etc. How do these existing methods scale to a particular domain? Moreover, multilingualism is also an important aspect that is under explored, i.e., how different language chapters of a KG such as Wikidata or DBpedia can help complete a KG in one language?

4.6 Knowledge Engineering for Semantic Web Machine Learning Systems

Marta Sabou (Vienna University of Economics and Business, AT)

In line with the general trend in artificial intelligence research to create intelligent systems that combine learning and symbolic components, a new sub-area has emerged that focuses on combining machine learning (ML) components with techniques developed by the Semantic Web (SW) community – Semantic Web Machine Learning (SWeML for short). Of particular interest are the emerging variations of processing patterns used in these SWeML systems in terms of their inputs/outputs and the order of the processing units. While several such neuro-symbolic system patterns were identified previously, we performed a systematic study and analyzed nearly 500 papers published in the last decade in this area. Overall we discovered 41 different system patterns, which we categorized into six pattern types. We observed that simple patterns that only incorporate one ML module are the most frequent, however the number of modules used in SWeML Systems is growing over time leading to increasingly complex and sophisticated system architectures for these novel systems. This development raises interesting questions for our community: What does the emergence of these complex systems mean for knowledge engineering? Do we need to rethink how we create, evaluate and evolve knowledge resources to better fit the requirements of such systems? What are typical SWeML systems patterns that can be used for various knowledge engineering tasks? Can we make use of these system patterns to guide the development of knowledge-based intelligent systems?

4.7 Shifting from a Triple-centric View to a Knowledge Components View in KGs

Eva Blomqvist (Linköping University, SE)

Tool support and partial automation is essential in today's Knowledge Engineering (KE) practices. This is true both for creating schemas, e.g. ontologies, and corresponding knowledge graphs. It is rarely the case that a single triple in a KG answers a user's query, rather, users of knowledge intensive systems most often need much more complex knowledge structures. An example from our previous experience is the notion of a crime, in the policing domain. A naive look at the concept of may lead to modelling a direct relation between a crime concept and a person that committed that crime. While this may be to some extent valid in a historic record, for an ongoing police investigation however, there are only suspects that to a certain degree can be connected to the crime, based on specific evidence. Even the crime in itself may need to be represented not as a single event, but as a series of actions, that could lead to certain charges being applicable in court. On the other hand, end users, in this case the police investigator also need ways to abstract from highly complex relations, to get an overview of the main connections between events and people involved in the investigation. Hence, it becomes essential that the knowledge engineering process captures all these end-user relevant levels of granularity, i.e. not only the triple-level but as more complex knowledge components. Some previous work on ontology design patterns, and recently conceptual components, point in this direction. However, this has not yet been fully brought into KE methodologies, tools, visualisations, and reasoning methods. Even further, when automating parts of the KE methodologies, such as the population of KGs, there is a need for knowledge extraction not only at the triple level, but at the level of detecting and extracting such complex components, e.g. from natural language text, where many open challenges exist.

4.8 A Normative Knowledge Graph for Verified Identity Applications

Bradley Allen (Merit International, Inc. – Millbrae, US)

The Merit Graph is a commercial example of a knowledge graph. However, in contrast to other commercial knowledge graphs, and because of the sensitivity of its areas of application in licensing, certification, and emergency services, the Merit Graph takes special care to address the problem of ensuring that the data it contains is managed to the highest standards of truth and trust. The Merit Graph maintains metadata about the provenance of statements about relations and entities, and uses that information to establish access control over data in the graph. This metadata supports verifiable and fine-grained policies that are meant to ensure the trustworthiness of the data, as well as to prevent improper sharing of personal data with third parties. The normative specification of these policies uses principles derived from action and deontic modal logic, allowing the control of not only who can access certain data, but also who is permitted to share data they have access to with whom, a capability necessary to provide organizations the tools needed to ensure that data that they are responsible for is not compromised or abused by others with whom they share that data. The Merit Graph is formally defined in a way that it can be transformed into a set of logical statements which, combined at processing time with rules, can be used to perform automated reasoning about the data in the graph. Rules are managed as part of the schema associated with the graph, through user interfaces used by system administrators to establish policies and provide domain expertise for specific use cases. This capability is used to automatically perform syntactic and semantic validation, transformation, and enhancement of data during the ingestion and issuance of merits, personas, and folios. It can also be used to perform advanced analytics, for example, link prediction in support of the recommendation of career or educational opportunities for licensed individuals, or normative reasoning to establish additional permissions and obligations of entities represented in the graph.

4.9 Semantic Interoperability at Conceptual Level: Not Easy but Necessary

Valentina Presutti (University of Bologna, IT)

Knowledge graphs (KG) have the potential of enabling meaning-aware artificial intelligence (AI) as opposed to statistically-aware AI. Let us consider recommending systems as an example. Most of them rely on features such as popularity: if I like a song, I will be suggested to listen to other music that is very "popular" among consumers who listened to the same song before me. To act differently and being able to personalise recommendations and motivate them, AI systems need to be aware of the meaning associated with the music (or any other item) they recommend and of the preference that emerge from a consumer's previous behaviour. Encoding the meaning of music or of other subjects is a hard problem but knowledge graphs and their ability to capture and formalise domain knowledge can push AI systems toward this achievement. One main issue is that specialised, domain knowledge is often overlooked. We are literally sitting on an unprecedented global, distributed source of knowledge addressing all sorts of specialised domains (Linked Open Data – LOD) but most KG-related research is limited to analyse and reuse encyclopedic knowledge. From a study that analyses the alignment between LOD ontologies [7] it emerges that LOD is poorly linked at the conceptual level (and I speculate that these alignments are mostly based on labels and common sense). There is an opportunity and a challenge to analyse LOD's knowledge from specialised domains, to enrich it and properly link it at the conceptual level. We shall resume the Semantic Web agenda about alignment and reuse of distributed ontologies, which opens to numerous research paths: to define more expressive and flexible knowledge representation languages, informed by empirical semantics; to standardise ontology design patterns (ODP); to provide tool support that makes it easy to reuse ODP, to perform ODP-based ontology alignment, to document (automatically) ontologies and KGs, to perform ontology testing, to lexicalise ontologies, etc. Only with semantic interopertability at the conceptual level and by properly addressing specialised domains shall we make a step towards meaning-aware AI systems.

4.10 Modelling Complex Concepts

Marieke van Erp (KNAW Humanities Cluster - Amsterdam, NL)

The success of AI technologies on standardised benchmark datasets, invites us to move towards more difficult and more complex concepts and tasks. The digital humanities domain presents many opportunities for investigating the recognition and modelling of complex concepts thanks to massive digitisation efforts that have made available large and varied datasets, in multiple modalities. My work now specifically highlights the complexities in modelling a concept such as smell, dealing with its representations in various media, and how the temporal dimension of historical and linguistic research forces us to deal with issues such as changing social norms and our colonial history.

4.11 KG Magic Requires KE Magic

Stefan Schlobach (VU University Amsterdam, NL)

Expert Systems have been among the initial success stories of Knowledge Representation, showing the potential of (mostly rule-based) formalised knowledge in a variety of tasks in various domains. The enormous costs of producing such high-quality knowledge led to the development of a variety of Knowledge Engineering (KE) methodologies in the Nineties and the decades after, which focused on the challenge of creating systematic processes to formalise tacit and tribal knowledge that, while being essential for the success of a system, is very often neither explicit, nor formalised. Nowadays, Knowledge Graphs (KG) are often considered to be some kind of magic wands of modern AI with the promise to extend purely statistical, learning-based, approaches by more generalisability and explainability. This has lead to increased interest in the development of Knowledge Graphs by commercial partners. The engineering challenges for constructing such high-quality knowledge remain the same as 10, 20 or 30 years ago; tribal and tacit knowledge is still as non-explicit and non-formalised as it used to be then. My research ambition is to extend the proven socio-technical KE methodologies with recent technological advances, e.g. based on Language Models or other statistical learning-based methods, to scale-up to the required complexity of modern AI-based systems.

References

1 Denny Vrandečić. Building a multilingual Wikipedia. *Commun. ACM*, 64(4):38–41, 2021.
2 Jan Portisch, Michael Hladik, and Heiko Paulheim. Rdf2vec light–a lightweight approach for knowledge graph embeddings. *arXiv preprint arXiv:2009.07659*, 2020.
3 Jan Portisch, Michael Hladik, and Heiko Paulheim. Kgvec2go–knowledge graph embeddings as a service. *arXiv preprint arXiv:2003.05809*, 2020.
4 Russa Biswas, Radina Sofronova, Harald Sack, and Mehwish Alam. Cat2type: Wikipedia category embeddings for entity typing in knowledge graphs. In Anna Lisa Gentile and Rafael Gonçalves, editors, *K-CAP '21: Knowledge Capture Conference, Virtual Event, USA, December 2-3, 2021*, pages 81–88. ACM, 2021.
5 Genet Asefa Gesese, Mehwish Alam, and Harald Sack. Literallywikidata – A benchmark for knowledge graph completion using literals. In Andreas Hotho, Eva Blomqvist, Stefan

Dietze, Achille Fokoue, Ying Ding, Payam M. Barnaghi, Armin Haller, Mauro Dragoni, and Harith Alani, editors, *The Semantic Web – ISWC 2021 – 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24-28, 2021, Proceedings*, volume 12922 of *Lecture Notes in Computer Science*, pages 511–527. Springer, 2021.
6 Cristian Santini, Genet Asefa Gesese, Silvio Peroni, Aldo Gangemi, Harald Sack, and Mehwish Alam. A knowledge graph embeddings based approach for author name disambiguation using literals. *Scientometrics*, 127(8):4887–4912, 2022.
7 Luigi Asprino, Wouter Beek, Paolo Ciancarini, Frank van Harmelen, and Valentina Presutti. Observing LOD using equivalent set graphs: It is mostly flat and sparsely linked. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtech Svátek, Isabel F. Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web – ISWC 2019 – 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part I*, volume 11778 of *Lecture Notes in Computer Science*, pages 57–74. Springer, 2019.

5 Breakout Groups

5.1 Integration of Language Models and Structured Data

Juan Sequeda (data.world – Austin, US)
Mehwish Alam (FIZ-Karlsruhe, Leibniz Institute for Information Infrastructure, DE & Karlsruhe Institute of Technology, DE)
Soren Auer (TIB - Hannover, DE)
George Fletcher(Eindhoven University of Technology, NL)
Harald Sack (FIZ-Karlsruhe, Leibniz Institute for Information Infrastructure, DE & Karlsruhe Institute of Technology, DE)

This group focused on how Large Language Models (LLMs) can be integrated or used for structure data.

5.1.1 Discussed Problems

Large-scale Language Models (LLMs) have shown impressive results in terms of language generation, question answering, but also software source code generation or translation. Following the presentation of the overview of the state of the art in Automated Knowledge Graph Construction, it was observed that the surveyed methods focused on automated approaches to construct knowledge graphs from unstructured sources. The question is whether these results can be applied for automatically constructing knowledge graphs from structured data (e.g. tabular, relational), and mapping structured data to ontologies and knowledge graphs.

The following initial observations were made:

- There are two streams to consider: 1) Automatic Knowledge Graph Construction from structured data, namely given structured data as input, the output is a knowledge graph, and 2) Automatically Mapping structured data to Knowledge Graph, namely, given structured data and an existing knowledge graph as input, the output is an augmented knowledge graph.

- The second stream considers the traditional data integration challenges of schema matching and entity linking.
- Language models have common sense knowledge. However, to do the mapping, it would also need to have specific business/domain knowledge, which may not exist in language models today.

5.1.2 Possible Approaches

There have been few recent studies focusing on tabular data to KG matching [1] along with many recent efforts by the semantic web community which designs a group benchmark dataset for this problem, i.e., the SemTab [2, 3, 4] challenge where the community joins forces and presents their systems targeting the problem of tabular data to KG. None of the existing studies so far utilize LLMs for performing this matching.

On the other hand, there have been several efforts where the latent representations are learned directly from the tabular data such as Tab2Vec [5], TaBert [6], etc. Tab2Vec is then evaluated on row completion, table completion, and table retrieval tasks. TaBERT is a pre-trained model that learns representations for natural language sentences and tabular data. These efforts should further be explored and exploited for mapping structured data to ontologies and KGs. A brief collection of methods following this line of research has been discussed in [7]. A deep dive into Machine or Deep Learning methods for tabular data is required.

5.1.3 Open Research Questions

The high-level questions to consider are:

- How do we automatically construct a knowledge graph from structured data?
- How do we automatically construct mappings from structured data to a knowledge graph?

Diving deeper into these questions, we discussed the following questions:

- Do we even need LLMs for this problem? It seems that we are turning this problem into a nail for the Language Model hammer, thus we should try to use this tool. As observed, language models consist of common-sense/domain-agnostic knowledge and may lack specific domain/business knowledge, thus the limits of existing language models need to be investigated. On the other hand, if we look at schema.org, we have evidence of a manual, low-effort, distributed, community-driven, and scalable approach to adding semantics to web pages.
- What is the cost/benefit tradeoff to using LLMs? What do we do with the results of a language model? A user will most likely need to review the result. For this approach to be cost beneficial, it would need to be drastically reduced to the cost of creating the knowledge graph manually/non-language model approaches.
- How can mappings be learned with additional context provided as input, for example, mapping patterns?
- What happens when the input is just tabular data vs relational data (SQL DDL, constraints)?
- Would there be a need to denormalize the data into a single flat table?
- What are the frameworks for evaluation?

5.1.4 Next Steps

Some anecdotal evidence arising from some experiments done during the breakout session[8] showed that while LLMs are able to perform some form of data mapping on typical textbook examples, they quickly fail when data structures are more original (and thus less likely to be included in the LLM training data). Also, due to the lacking explainability of LLM results, it is extremely cumbersome and thus not feasible to manually verify the results, since the required effort for this task might easily exceed a manual mapping. However, LLMs could possibly be used for generating smaller (e.g. property) mapping or documentation suggestions.

An interesting question is whether the experience of LLMs can be applied to generate novel large-scale structured data models, which are trained with millions of data schemata, ontologies, and mappings and will thus be better suited for mapping generation tasks. However, this might not be practically feasible since many of the required artifacts are private. Possibly some federated learning of such large-scale structured data models could alleviate this problem.

When considering applying Language Models for the problems of schema mapping, it's key to understand the state of the art in order to create bridges. For example, there is formal work on learning schema mappings [8] and queries from examples of structured data.

One of the possible solutions could be to align two embeddings generated from different structures such as tabular data and the knowledge graphs and perform alignments or make use of both kinds of embeddings and use it in the downstream task of matching.

What would prompt engineering for data integration look like? This may be an extension of the existing SemTab challenge.

Finally, we should be careful and not just jump on using the Language Model hammer and start pounding on that hammer to see what works. It is paramount to have a systematic approach to understanding and evaluating how language models and structured data can be combined to automatically construct knowledge graphs.

References

1 Jixiong Liu, Yoan Chabot, Raphaël Troncy, Viet-Phi Huynh, Thomas Labbé, and Pierre Monnin. From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods. *Journal of Web Semantics*, page 100761, 2022.

2 Ernesto Jiménez-Ruiz, Oktie Hassanzadeh, Vasilis Efthymiou, Jiaoyan Chen, Kavitha Srinivas, and Vincenzo Cutrona, editors. *Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching (SemTab 2020) co-located with the 19th International Semantic Web Conference (ISWC 2020), Virtual conference (originally planned to be in Athens, Greece), November 5, 2020*, volume 2775 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

3 Ernesto Jiménez-Ruiz, Oktie Hassanzadeh, Vasilis Efthymiou, Jiaoyan Chen, Kavitha Srinivas, and Vincenzo Cutrona. Results of semtab 2020. In Ernesto Jiménez-Ruiz, Oktie Hassanzadeh, Vasilis Efthymiou, Jiaoyan Chen, Kavitha Srinivas, and Vincenzo Cutrona, editors, *Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching (SemTab 2020) co-located with the 19th International Semantic Web Conference (ISWC 2020), Virtual conference (originally planned to be in Athens, Greece), November 5, 2020*, volume 2775 of *CEUR Workshop Proceedings*, pages 1–8. CEUR-WS.org, 2020.

[8] `https://beta.openai.com/playground`

4 Vincenzo Cutrona, Jiaoyan Chen, Vasilis Efthymiou, Oktie Hassanzadeh, Ernesto Jiménez-Ruiz, Juan Sequeda, Kavitha Srinivas, Nora Abdelmageed, Madelon Hulsebos, Daniela Oliveira, and Catia Pesquita. Results of semtab 2021. In Ernesto Jiménez-Ruiz, Vasilis Efthymiou, Jiaoyan Chen, Vincenzo Cutrona, Oktie Hassanzadeh, Juan Sequeda, Kavitha Srinivas, Nora Abdelmageed, Madelon Hulsebos, Daniela Oliveira, and Catia Pesquita, editors, *Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching co-located with the 20th International Semantic Web Conference (ISWC 2021), Virtual conference, October 27, 2021*, volume 3103 of *CEUR Workshop Proceedings*, pages 1–12. CEUR-WS.org, 2021.

5 Li Zhang, Shuo Zhang, and Krisztian Balog. Table2vec: Neural word and entity embeddings for table population and retrieval. In Benjamin Piwowarski, Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer, editors, *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, pages 1029–1032. ACM, 2019.

6 Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. Tabert: Pretraining for joint understanding of textual and tabular data. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 8413–8426. Association for Computational Linguistics, 2020.

7 Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, and Gjergji Kasneci. Deep neural networks and tabular data: A survey. *CoRR*, abs/2110.01889, 2021.

8 Balder ten Cate, Víctor Dalmau, and Phokion G. Kolaitis. Learning schema mappings. *ACM Trans. Database Syst.*, 38(4):28, 2013.

5.2 Knowledge Engineering with Language Models and Neural Methods

Elena Simperl (King's College London, GB, elena.simperl@kcl.ac.uk)
Paul Groth (University of Amsterdam, NL, p.t.groth@uva.nl)
Steffen Staab (Universität Stuttgart, DE, steffen.staab@ipvs.uni-stuttgart.de)
Marta Sabou (Vienna University of Economics and Business, AT, marta.sabou@wu.ac.at)
Eva Blomqvist (Linköping University, SE)
Bradley Allen (Merit International, Inc. – Millbrae, US, bradley.p.allen@gmail.com)

5.2.1 Discussed Problems

Knowledge engineering remains expensive. The emergence of new powerful automation tools (e.g. large language models) opens new avenues for exploration to bring down the cost of knowledge engineering. While there is much work using machine learning for knowledge engineering (e.g. ontology learning, curation), we know much less about the overall picture of the incorporation of these new machine learning (ML) techniques.

5.2.2 Open Research Questions

- How does state of the art machine learning, including large language models augment knowledge engineering processes and projects?
- What is the existing user / developer experience of machine learning tools?

- What are the roles of people and machines in current knowledge engineering process?
- How do you evaluate the added value of automation to knowledge engineering processes?
- How do you control the outputs of ML-based systems are what you need for knowledge engineering?
- What is the interplay knowledge graph engineering (with or without AI) and system engineering?

5.2.3 Possible Approaches

Classic knowledge engineering approaches (e.g. NEON [1]) tend to distinguish between management, support, development activities while knowledge graph engineering approaches are still in their infancy and tend to focus largely on development activities (e.g. relation extraction, learning class representations, refinement). Therefore, the potential for automation is much bigger when considering the management and support activities, for instance, re-purposing existing knowledge graphs in new contexts, search, automatic generation of documentation (e.g. labels in multiple languages, entity and relation descriptions), or process optimization. These are just some examples, hence, what is needed is a systematic study of current practices, roles, and tools that support them. This can be achieved in several points:

- analyze emerging knowledge engineering processes to assess their automation potential building on [2, 3, 4] (e.g. through literature reviews or empirical analysis of existing code);
- run task-based studies in which the tasks would be to build knowledge graphs following established knowledge engineering methodologies using existing out-of-the-box automation tools (e.g. HuggingFace);
- case study analysis of existing knowledge engineering projects that include an AI element.

Within this analysis, a key emerging technology, is prompt engineering [5], whose outputs, based on large language models, could inform knowledge engineering activities in several ways. Here, a mapping to between the state-of-the-art in prompt engineering and knowledge engineering would be beneficial. In particular, there is a question as to how these technologies can be suitably controlled for knowledge engineering processes.

Evaluating and understanding the impact of technology is an established field with its own methodologies and approaches. In particular, there has been considerable work by researchers, practitioners and regulators around the use of machine learning in a range of applications, which resulted in frameworks for responsible/trustworthy AI [6, 7]. Studies with technical users of AI seem to suggest data scientists and other technical roles tend to over-trust the outcomes of machine learning systems and do not always fully grasp how they work, or, where applicable, their explanations [8]. End-users of downstream applications need means to provide feedback and adjust the outputs of the application, which often involves improving the underlying data – often, knowledge graph embeddings are a source of such data, hence it is important when evaluating the added value of automation in building a graph to consider questions of end-user agency and control from the start. Here, extending approaches [9] that look at machine teaching[9] to examples from knowledge graphs appears promising.

Any way to understand to knowledge engineering with AI systems should be based on the existing extensive work with designing and planning for AI systems. This includes a series of practices including, following human-centered design, identifying multiply evaluation

[9] https://github.com/cleanlab/cleanlab

metrics, extensive testing and continued monitoring while in deployment.[10] This also includes reflecting on fairness and interpretability, which come with their own set of best practices and techniques.

Furthermore, often knowledge graph engineering is described as one-off activity, where the project is finished when a knowledge graph is complete. In practice, this is not the case as seen by the examples discussed at the seminar. Therefore, there is to study the ongoing maintenance of knowledge graphs the roles and automation involved. This should be done with a grounding in the current thinking around data-centric AI and MLOps[10].

5.2.4 Next Steps

- Perform the user studies, case studies and reviews mentioned above;
- Organizie a workshop bringing together prompt engineering and knowledge engineering experts.

References

1 María del Carmen Suárez de Figueroa Baonza. *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse.* PhD thesis, Universidad Politécnica de Madrid, Madrid, Spain, June 2010.
2 Anna Breit, Laura Waltersdorfer, Fajar J. Ekaputra, Tomasz Miksa, and Marta Sabou. A lifecycle framework for semantic web machine learning systems. In *Database and Expert Systems Applications – DEXA 2022 Workshops*, pages 359–368, Cham, 2022. Springer International Publishing.
3 Gytė Tamašauskaitė and Paul Groth. Defining a knowledge graph development process through a systematic review. *ACM Transactions on Software Engineering and Methodology*, 2022.
4 Lucie-Aimée Kaffee, Kemele M. Endris, and Elena Simperl. When humans and machines collaborate: Cross-lingual label editing in wikidata. In *Proceedings of the 15th International Symposium on Open Collaboration*, OpenSym '19, New York, NY, USA, 2019. Association for Computing Machinery.
5 Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *CoRR*, abs/2107.13586, 2021.
6 Lu Cheng, Kush R Varshney, and Huan Liu. Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, 71:1137–1181, 2021.
7 Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.
8 Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.
9 Curtis G Northcutt, Anish Athalye, and Jonas Mueller. Pervasive label errors in test sets destabilize machine learning benchmarks. *arXiv preprint arXiv:2103.14749*, 2021.

[10] `https://ai.google/responsibilities/responsible-ai-practices/`, `https://www.microsoft.com/en-us/ai/responsible-ai-resources`

10 Hima Patel, Shanmukha Guttula, Ruhi Sharma Mittal, Naresh Manwani, Laure Berti-Equille, and Abhijit Manatkar. Advances in exploratory data analysis, visualisation and quality for data centric ai systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, page 4814–4815, New York, NY, USA, 2022. Association for Computing Machinery.

5.3 Explainability of Knowledge Graph Engineering Pipelines

Axel-Cyrille Ngonga Ngonga (Universität Paderborn, DE, axel.ngonga@upb.de)
Diana Maynard (University of Sheffield, GB, d.maynard@sheffield.ac.uk)
Marcel R. Ackermann (Schloss Dagstuhl LZI – Trier, DE, marcel.r.ackermann@dagstuhl.de)

5.3.1 Discussed Problems

There is currently no standard definition for explanation and explainability. We consider the following explanation scenario [1]: An explainer is to provide an explanation for an explanandum to an explainee via explanans. In knowledge engineering for knowledge graphs, the explainer is commonly a system driven by some background knowledge. The explanandum could consist of the graph as a whole or a single statement, but also the process. The explainee can be a human or another system. Finally, the explanans can range from natural language to a set of assertions in a formal language. In this setting, explanation is clearly an iterative process within which the explainee can request supplementary information (e.g., pertaining to previous explanans) to reach the explanation goal. When modelled as such, the function of an explanation is to empower the explainee to understand enough about the explanandum to take action. It is rather unclear how explanations and the accompanying processes are to be tailored and evaluated.

Explainability is central for several aspects of the knowledge engineering process including building trust, quantifying uncertainty, hypothesis exploration, due diligence support, compliance and liability, data and process audits, and data usage agreements. *Trust* is a key element of explainability, enabling the explainee to evaluate the correctness / usefulness and/or actionability of the output. *Quantifying uncertainty* ensures that the explainee has a measurement for the reliability of the explanation process and hence of the explanandum. Devising pareto-optimal explanation processes that can cater for several of these aspects is a challenge which is currently not widely addressed.

5.3.2 Possible Approaches

The body of works on interpretability and explainability covers various disciplines ranging from psychology [1] to machine learning theory [2]. The ML community has developed post-hoc methods for explainability, including approaches such as LIME [3], SHAP [4], and MVU [5]. Ante-hoc solutions such as verbalization techniques for class expressions [6, 7] serve a similar purpose in inductive logic programming based on description logics. Still, these are one-shot explanations, which do not fully implement the iterative explanation process described in Section 5.3.1. Currently, there seems to be no detailed study encompassing the state of the art in theory and practice, but rather a number of piecemeal attempts to solve various issues in tackling explainability.

5.3.3 Open Research Questions

Explaining is an intrinsically challenging task, as a good explanation for a given explanandum must fit potentially very different user and application requirements. The computation of explanations must hence be carried out in collaboration with the explainee – a process which is widely unexplored in knowledge engineering. Still, it seems obvious that one-shot explanations will rarely be enough to satisfy user needs. A clear specification of the relation between explanations and interpretations (e.g., as described in [2]) must be at the core of future research as the distinction between them is unclear and often misrepresented. Further key challenges include the need to provide measures for explanation, methods to evaluate them and to quantify their trustworthiness (both intrinsically and extrinsically) and to allow for measures of uncertainty. On the other hand, since the field of explainability in this context is both fast-evolving and application-dependent, it is therefore difficult and perhaps undesirable – especially in the near future – to develop rigid standards.

5.3.4 Next Steps

We plan to write a survey of the state of the art in explainability, with the aim of understanding better the current limitations and future directions. Several workshops on explainability are currently organized at major AI conferences including XAI4CV at CVPR, and SemEx at ISWC. We plan to support these efforts and contribute requirements and solutions from knowledge engineering. Our ultimate goal is to write a reference book on methods and applications of explainable AI for knowledge engineering.

References

1. Katharina J Rohlfing, Philipp Cimiano, Ingrid Scharlau, Tobias Matzner, Heike M Buhl, Hendrik Buschmeier, Elena Esposito, Angela Grimminger, Barbara Hammer, Reinhold Häb-Umbach, et al. Explanation as a social practice: Toward a conceptual framework for the social design of ai systems. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3):717–728, 2020.
2. Christoph Molnar. *Interpretable machine learning*. Lulu. com, 2020.
3. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144, 2016.
4. Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
5. Kilian Q Weinberger and Lawrence K Saul. Unsupervised learning of image manifolds by semidefinite programming. *International journal of computer vision*, 70(1):77–90, 2006.
6. Norbert E Fuchs, Kaarel Kaljurand, and Tobias Kuhn. Attempto controlled english for knowledge representation. In *Reasoning web*, pages 104–124. Springer, 2008.
7. Axel-Cyrille Ngonga Ngomo, Diego Moussallem, and Lorenz Bühmann. A holistic natural language generation framework for the semantic web. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 819–828, Varna, Bulgaria, September 2019. INCOMA Ltd.

5.4 Construction with Modalities and Types

Antoine Isaac (Europeana Foundation - Den Haag, NL, antoine.isaac@europeana.eu)
Marieke van Erp (KNAW Humanities Cluster - Amsterdam, NL, marieke.van.erp@dh.huc.knaw.nl)

This working group focused on investigating the gap between what is currently captured in Knowledge Graphs and what information is contained in other sources and modalities. The impetus for this break-out group came from the observation that most current KGs focus on factoid-information that is easily captured in triples such as information about entities and properties.

More freely structured data such as text and images contain information that may be difficult to capture in triple format. Procedural knowledge or other knowledge that has a clear sequence (e.g. word order in text) does not naturally fit into KGs. Solutions such as the NLP Interchange Format (NIF) have been proposed but lead to bulky modelling.

Information concerning more abstract concepts such as opinions or perspectives are often implicit and have a social and contextual dimension – what is acceptable in one context may not be acceptable in another. This type of information intersects with commonsense knowledge as well as social norms. Something that, to the best of our knowledge, is currently not captured in KGs.

This break-out group therefore poses the following questions:

- What can (or should be) be included knowledge graph?
- Which source can it come from?
- What is the purpose of (elements of) the knowledge graph?

5.4.1 Discussed Problems – Towards a Typology of Knowledge for KGs

The main problems that we discussed are the identification and characterization of the types above, as well as the representation techniques that can be used to handle them.

Table 1 Types of knowledge typically found in knowledge graphs.

Type of knowledge	Examples	(Typical) representation level
Conceptual knowledge	Entities: classes, properties. Typically the knowledge written in OWL	T-Box, R-Box
Factual knowledge	Entities and statements: objects, relationships	A-Box
Procedural knowledge		
Rules		
Commonsense and encyclopedic	Naive physics, "knives are used for cutting", "dinner is at 6pm"	
Causality		
Sentiment		
Arguments, claims	Political, scientific arguments	Qualifiers, named graphs, nanopublications
Beliefs	"No man landed on moon"	
Provenance, references		Qualifiers, named graphs
Perspectives, narratives, frames, interpretations	Colonial perspectives on objects being "given", Tonality of a music piece (which changes as the applied theory changes), Wikidata has different entries for Jesus (in which he may be the last or second-to-last prophet)	Representations of situations, (trans-)actions and (sociological) roles
Moral and ethical judgements	"Abortion is a crime" according to certain groups of people in the US	

Definitions (and to some extent, terminology) need to be confirmed and refined for these types. For example our group discussed commonsense knowledge, only to conclude that while we have a general idea of the notion – e.g., it includes what is needed to understand the newspapers – it remains extremely vague and the term is quite overloaded.

5.4.2 Possible Approaches

A first way to refine and better structure the notions (roughly) laid down above would be to identify suitable dimensions of analysis and position the various types of knowledge along these dimensions. This idea follows upon the example of the "expressiveness spectrum" produced at AAAI99, which was presented by McGuinness at the Seminar (see fig. 9).

A complementary, more bottom-up approach, would be to inventorise the elements of knowledge used in actual KGs. As a first attempt, and recognizing that the types of knowledge present in KGs heavily depends on the domain or the application considered, the group embarked on identifying knowledge elements that are typically (or less typically) found in a few selected domains.

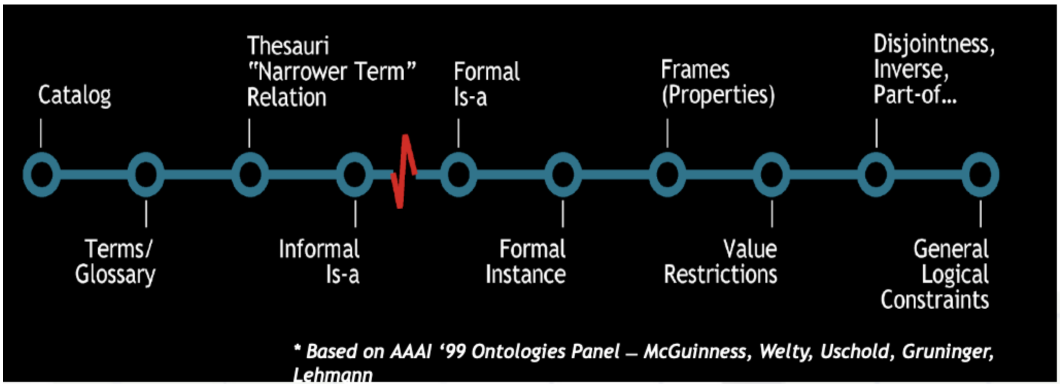


Figure 9 Expressiveness spectrum as presented by McGuinness during the seminar (see Section 3.3 for an overview of the talk).

Table 2 Matrix in which the break-out group brainstormed types of knowledge that may be included in a knowledge graph for a particular use case.

	Conceptual	Facts	Rules	Procedural	Common sense	Sentiment, moral/ ethical judgment	Arguments	Beliefs	Provenance	Narratives/Perspectives	Causality
Contested heritage	x	x	x	(x)	x?	x	x	x	x	x	
Musicology	x	x	x	?	(x)	x	x	?	x	x	
Engineering	x	x	(x)	x	(x?)		x		x		
Agriculture	x	x		x	x		x		x	x	
Medical											
Wikidata											
Astronomy				(x)							
Bibliography											

5.4.3 Next Steps

The discussion in our group was only a first attempt at charting the landscape of the various types of knowledge that can appear in KGs. This effort needs to be continued, especially on:

- Surveying of cases and the types of knowledge that can or should be relevant for them.
- Working on one or several "spectra" of expressiveness and other dimensions, for the knowledge that can be represented in KGs.

This work, which could be progressed in a workshop-like setting (especially in order to agree on types and dimensions) and long-term community outreach effort (especially for the surveying), should be eventually presented in a written form that can benefit researchers and practitioners on the longer term – either as a separate paper or part of a wider book on knowledge engineering methodology.

5.5 Knowledge Graphs vs. Other Forms of Knowledge Representation

Paul Groth (University of Amsterdam, NL, p.t.groth@uva.nl)
Aidan Hogan (University of Chile – Santiago de Chile, CL, ahogan@dcc.uchile.cl)
Lise Stork (Vrije Universiteit Amsterdam, NL, l.stork@vu.nl)
Katherine Thornton (Yale University Library – New Haven, US, katherine.thornton@yale.edu)
Denny Vrandečić (Wikimedia Foundation – San Francisco, US, denny@wikimedia.org)

This working group focused on how knowledge graphs relate to, complement, and could be combined with, or even replaced by, other forms of knowledge representation, including traditional forms of knowledge representation like text and tables, as well as novel forms of knowledge representation such as (large) language models.

5.5.1 Discussed Problems

- How do knowledge graphs relate to other types of knowledge representation?
- What kinds of knowledge are knowledge graphs good at representing?
- Will knowledge graphs still be needed given the advancements in large language models?
- Are knowledge graphs the best target for knowledge extraction from large language models?

5.5.2 Discussion

Despite the growing popularity of knowledge graphs, it is not always clear for what sorts of knowledge (or knowledge-centric applications) such graphs are appropriate representations. Our discussion thus covered different forms of knowledge representation and knowledge, how knowledge graphs relate to modern forms of knowledge graphs like large language models, and what forms of representation are useful in what settings.

5.5.2.1 Different Representations

Knowledge graphs are a particular class of knowledge representation; some members of this class include RDF, RDF*, property graphs, Wikidata, etc. Knowledge graphs have garnered a lot of attention for their ability to integrate knowledge from diverse sources at large scale. However, they are only one instance of a particular "modality" of knowledge representation used by humans. One may thus ask: How do knowledge graphs relate to other modalities of knowledge representation? When are they more or less useful than the other alternatives? Can we model all knowledge within a knowledge graph, or do we need different representations for different types of knowledge? How could knowledge graphs be combined with, or interact with, these other representations?

With respect to modalities of knowledge representation, we can identify, for example, the following:

- *Textual*: books, literature, rich text, emails
- *Lexicographical*: thesauri, lexemes, vocabulary, dictionaries
- *Tabular*: CSV, spreadsheets, relational tables
- *Temporal*: edit histories, chronologies, stock tickers, temporal databases
- *Graph*: (social/transport/biological) networks, knowledge graphs
- *Hierarchical*: taxonomies, classifications, XML, JSON

- *Logical*: rules, ontologies, first-order logic, frames, scripts, schemas
- *Procedural*: code, instructions, workflows, tutorials
- *Multimedia*: video, audio, images
- *Diagrammatic*: UML, ER, pie charts, Sankey diagrams
- *Numeric*: embeddings, language models, matrices
- *Mental*: human memory, epigenetic memory
- *Social*: word of mouth, gossip, stories, songs, institutional memory

It is clear that knowledge graphs are not intended to replace all of these modalities, nor is it clear that they even can. To understand this in more detail, we identify some different types of knowledge:

- Factual: expressing declarative statements representing claims of truth (e.g., *the capital of Nigeria is Abuja*).
- Quantified: expressing statements for existential or universally quantified elements (e.g., *all countries have a capital*).
- Contextual: expressing statements that are claimed to be true within a certain context, such as a probability or fuzzy quantification of truth (e.g., *a country probably only has one capital*); a temporal context (*the capital of Nigeria has been Abuja since 1991*), etc.
- Procedural: expressing ways of doing things, often involving a sequence of actions and their effects (e.g., *how to prepare the Nigerian dish Tuwo shinkafa*).
- Narrative: expressing a series of statements building a model and working within that model to communicate knowledge
- Tacit: implicit knowledge often gained through lived experience; may involve qualia, such as taste, smell, touch, sight (e.g., *what Tuwo shinkafa tastes like*); socially-acquired knowledge relating to customs, values, etc. (e.g., *that it would be strange to eat Tuwo shinkafa with marmalade*), and so forth.
- Counterfactual: expressing statements of possible world states, representing what would be true under varying circumstances and often including modal terms such as "possibly" (*if I would take the bike, I would possibly not be on time).*

Knowledge graphs are perhaps strongest when representing factual knowledge, particularly when such knowledge is expressed as binary relations. When combined with rules or ontologies, they can further represent quantified knowledge. When combined with techniques such as annotated logic, reification, named graphs, RDF* or property graphs, etc., they can also be used to capture contextual knowledge. For reasoning over such complex objects, however, new formalisms would be required [1]. Though knowledge graphs are only one possible representation, they have shown certain advantages and disadvantages when compared with tables (SQL, CSV, etc.), trees (JSON, XML, etc.), images, and so forth.

How knowledge graphs can be used to capture procedural knowledge is less clear. If, for example, we wanted to represent the sequence of steps in a recipe, while we could potentially structure the recipe as a graph of dependent steps or causal relations, the steps themselves will likely be described in natural language, such as "*mash the rice with a wooden spoon*". While such unstructured steps could potentially be decomposed (potentially recursively) into a structured sequence of sub-steps, and the instruments they involve, etc., and while a more fine-grained structure might help to later find recipes satisfying certain criteria, the resulting representation will not convey very well how to actually make the recipe (a video would be better).

Moreover, it is nontrivial how to represent hypothetical knowledge – such as counterfactual knowledge – for which statements can be equally likely depending on different world states. RDF* does allow for contextualised statements without any truth value assigned to them.

These are *quoted triples*,[11] which are statements not asserted and thus not evaluated in the knowledge graph. However, keeping track of the epistemic status of a contextualised statement, given situational facts, is not yet supported.

Likewise tacit knowledge is often inherently difficult to express, particularly as structured knowledge (including knowledge graphs).

5.5.2.2 Use for Knowledge Graphs in the Age of Large Language Models

Knowledge graphs will inevitably begin to compete with – and potentially complement – language models [2, 3], which have recently captured not only interest within academia, but also the more general public, in terms of their ability to seemingly understand and communicate knowledge through natural language. Such language models thus increase machine interpretability of human language. However, many of the modalities of knowledge representation introduced previously – including knowledge graphs – were primarily introduced as a way to make knowledge available in "machine-readable" structured formats. If language models are paving the way for human natural language to become "machine-readable", and if natural language is a more intuitive way for humans to capture, express, communicate and conceptualise knowledge (including procedural and tacit knowledge), then a key question arises: assuming that language models continue to improve over time, will we even need structured representations like knowledge graphs in the future?

Knowledge graphs require knowledge to be structured, while language models are designed to capture unstructured knowledge. Thus the question of how knowledge graphs relate to language models is contained within the broader scope of how structured knowledge relates to unstructured knowledge. Viewed in this light, knowledge graphs are more efficient for query answering, are more reliably modifiable, are more transparent, cover the long tail better, and lend themselves better to explainability than large language models (which we expect to hold, also, for the medium-term future). In more detail:

- For **query answering**, looking up a triple in a triple store is more efficient than retrieving text from a large generative model with billions of parameters; e.g. it is less expensive to look up the capital of France in Wikidata than to generate that answer from GPT-3.
- If the world changes, or if an error in the knowledge is discovered, we can easily **fix, edit and update** the knowledge graph, but it is currently an open research question how a language model would need to update its weights to reflect such a change in the world or correct an existing error; e.g. if the capital of Kazakhstan gets a new name, or if the British monarch dies, how do we update a language model to incorporate that change?
- Knowledge graphs can be more **transparent** and can have clearer **provenance** as they can contain references, sources, or other ways to establish trust in the knowledge in the graph. Conversely, language models do not currently capture the connection between the weights and the textual sources used to learn these weights in a fine-grained way.
- Relatedly, knowledge graphs allow for more **explainability** than large language models. With a symbolic system we can display the involved ground statements, and the inferences that took place, whereas with language models, generating explanations is a very popular and challenging topic of active research [4].
- Knowledge graphs also **cover the long tail** better, and can be more easily extended to cover the long tail. A naive approach to increasing coverage for a language model

[11] https://w3c.github.io/rdf-star/cg-spec/editors_draft.html#dfn-quoted

is to retrain or refine it with more text about the topics to be covered; in a structured knowledge base you can just explicitly add the required structure. Anecdotal experience indicates that if we want to increase coverage of, e.g. different file types, we can either write or search for documents about these file types – and writing a new document may take dozens of minutes if not hours – or we can create a new item in a knowledge base, which may take half a minute.

The aforementioned advantages of knowledge graphs versus language models have clear parallels elsewhere in terms of the advantages and disadvantages of structured/deductive/symbolic methods vs. unstructured/inductive/numeric methods. In the context of Natural Language Processing and Information Extraction, for example, while machine learning methods have led to major advances in the state-of-the-art, more traditional rule- or pattern-based approaches are still often preferred for certain applications (particularly in domain-specific scenarios) as they provide more control over the process, provide more transparent and explainable results, and can work better for the long tail or for emerging knowledge (where training data is sparse).

In conclusion, we think that knowledge graphs will not become redundant due to language models, but rather both can clearly complement each other.

5.5.2.3 Representations in Practice

Large language models are good at understanding (the distribution of) language and can therefore be used in a variety of downstream tasks such as named entity recognition. However, they also come with some important draw-backs and challenges, such as the lack of provenance and explainability (as discussed in the previous section). Humans express and record a lot of knowledge in unstructured form, but even before the advent of digital computers, humans were applying structure to knowledge for the purpose of understanding as well as communication (e.g., the Periodic Table).

Existing diverse representational structures (as enumerated in Subsection 5.5.2.1) each have their specific merits. In some cases, working with just a bunch of screenshots decreases cognitive load over working with free text. In others, we need formal rules or ontologies where transparency and clarity are key, whereas in other cases a table will suffice.

The downside of using a variety of data structures alongside one another is the lack of integration and harmonisation. The question then arises, how do we enable data federation in the case of heterogeneous data structures?

One solution would be a single data structure for heterogeneous data, such as the multimodal knowledge graph described in [5], as the go-to data structure. Such a data structure integrates multi-modal data such as lists, images, etc., queryable through a single query language. Potential pitfalls of such a heterogeneous data structure could be the added modeling complexity, resulting in data silos that would be hard to query/use, or structures that are difficult to query or understand by users. Another way to go forward would be a single knowledge based system integrating multiple types of knowledge, with a single unified query interface.

5.5.3 Open Research Questions

- How to allow for a knowledge based system that integrates different kinds of knowledge representation, but yet allow for a unified query interface?
- What types of tasks require which kind of knowledge representations?

- Language models are good in smoothly dealing with the brittleness problem of symbolic knowledge representations. How can we combine language models with knowledge graphs to gain the advantage of language models?
- Would increased use of datatypes be advantageous for knowledge graph engineering? Is that a potential approach for combining knowledge graphs with more knowledge representations?
- Can we separate individual facts or knowledge out of a language model, and store it in a more efficient representation, and thus save on parameters that would encode that knowledge, making them smaller and more efficient, while allowing them to access a knowledge graph?
- How could we track provenance for language models? How could we represent and explain where this response came from? How would we trace the lineage of statements in large language models?
- How can language models be updated?
- How can language models be adapted to better cover the long tail, emerging knowledge, etc.?

5.5.4 Next Steps

- Invite collaboration on a prototypical infrastructure that demonstrates the usefulness of combining a knowledge graph with other modalities, e.g., images and a language model.
- Setting up tasks or challenges that are expected to be very difficult for certain types of knowledge representations, and easy for others. Often benchmarks are biased towards tasks that are solvable with a given approach; for example, benchmarks for question answering over knowledge graphs will include questions that are answerable over the target knowledge graph, and might tend to exclude questions like *how can I mash rice?* or *what is EUR12.53 in USD?* that knowledge graphs are not well-suited for, even if users may often like to answer such questions. These challenges should include tasks that are easy for, say, knowledge graphs, but very challenging for language models; and vice versa. The challenges should also consider "meta-tasks", like updating the knowledge, curating answers, explaining them, etc. The challenges should be promoted within the wider machine learning communities. The best-performing approaches will likely require combining different forms of representation; thus the challenges might stimulate research on hybrid approaches.

References

1 Fabian Suchanek. The need to move beyond triples. *CEUR Workshop Proceedings*, 2593:95–104, 2020.

2 Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*, 2019.

3 Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona Diab, and Marjan Ghazvininejad. A review on language models as knowledge bases. *arXiv preprint arXiv:2204.06031*, 2022.

4 Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144, 2016.

5 Xander Wilcke, Peter Bloem, and Victor De Boer. The knowledge graph as the default data model for learning on heterogeneous knowledge. *Data Science*, 1(1-2):39–57, 2017.

5.6 Bias in Knowledge Graph Systems

Mehwish Alam (FIZ Karlsruhe, DE)
George Fletcher (Eindhoven University of Technology, NL)
Antoine Isaac (European Foundation – Den Haag, NL)
Aidan Hogan (University of Chile – Santiago de Chile, CL)
Diana Maynard (University of Sheffield, GB)
Heiko Paulheim (Universität Mannheim, DE)
Harald Sack (FIZ Karlsruhe, DE)
Elena Simperl (King's College London, GB)
Lise Stork (VU University Amsterdam, NL)
Marieke van Erp (KNAW Humanities Cluster – Amsterdam, NL)
Hideaki Takeda (National Institute of Informatics – Tokyo, JP)

The starting point of this discussion was the overview lecture on social and technical bias in knowledge graphs presented by Harald Sack. Bias often is characterised as a disproportionate weight in favour of or against a person, group, an idea or thing, usually in a way that is considered closed-minded, prejudicial, or unfair, especially one that is preconceived or unreasoned. Biases in Knowledge Graphs (KGs) as well as potential means to address them are different from those in other AI systems, as e.g. in large language models or in image classification. KGs store human knowledge about the world in structured format, e.g., triples of facts or graphs of entities and relations, to be processed by AI systems. In the past decade, extensive research efforts have gone into constructing and utilising KGs for tasks in natural language processing, information retrieval, recommender systems, and many more. In difference to language models and image classification systems, KGs are sparse, i.e. typically only a small number of triples exist per entity. Once constructed, KGs are often considered as objective and neutral reference data sources that safeguard the correctness of other systems. In reality this is often not the case, since KGs are created with specific application context in mind. This has the undesirable effect that biases inherent to KGs may become magnified and spread through KG based systems (Bias Network Effect).

Basically, biases in KGs may arise from the following sources [1]:

Data Bias: Bias may be already inherent in the source data from which the KG is created in an automated or semi-automated way. For KGS that are collaboratively created or based on collaboratively collected information, all forms of human biases might be already incorporated. Furthermore, bias can also be introduced by the algorithms used to sample, aggregate, and process that data.

Schema Bias: Bias may be introduced via the chosen ontology as the basis for a KG, or simply be embedded within ontologies. Most times, ontologies are developed in a top-down manner with application needs or certain philosophical paradigms in mind Typically defined by a group of knowledge engineers in collaboration with domain experts, ontologies consequently (though often implicitly) reflect the worldviews and biases of the development team (human bias and anthropocentric thinking). In addition, the ontology and its modelling often depends on the chosen representation language, i.e. typically a fragment of DL, and not the other way around.

Inferential Bias: Inferential biases in KGs arise at inferencing level, such as reasoning, querying, or rule learning.

Bias inherent in KGs will directly carry over into downstream representations such as KG embeddings (KGE). Also errors and incompleteness in KGs might cause bias in KGEs due to an unbalanced distribution of facts or attributes that does not reflect an objective worldview. Furthermore the chosen KGE model might be the source of additional bias induced by application-specific loss functions.

5.6.1 Discussed Problems

Bias as a signal problem

One way in which representation bias might surface in knowledge graphs is that information which can be inferred is not explicitly represented in a knowledge graph. For example, the relation *is married to* is symmetric, and from *A is married to B*, one can infer that *B is married to A* also holds. From a logical standpoint, it is therefore sufficient to encode one of the two statements in the knowledge graph.

In [2], it was reported that a vast majority of *is married to* relations in DBpedia are only present in one direction, and there are far more statements where the subject is female and the object is male than vice versa [3]. This can be considered a gender-related representation bias in the knowledge graph, since the editors (of Wikipedia infoboxes, which DBpedia is created from) find this information more noteworthy for females than for males.

The same paper [2] also discussed logical inference as a means to cancel the representation bias. In this example, it would mean filling the slots for the symmetric relation in both directions, i.e., adding *B is married to A* for every occurrence of *A is married to B*. However, in an experimental setup, they showed that the performance of using the debiased knowledge graphs in a few downstream tasks actually leads to worse performance.

One interpretation of this outcome is that bias can actually be a signal, which can help downstream applications. The fact that a human editor considered the fact *A is married to B* noteworthy, but not *B is married to A*, actually conveys some information about A and B – mainly that B is better known for other things. Removing the bias here also implies removing the corresponding signal.

Bias as a legal and professional problem

The EU commission distinguishes between fair and unfair bias.[12] In general, national and international law, as well as the standards of professional bodies [4], provide norms regarding the development and use of knowledge and data-based systems and applications. What are the relationships between legal and professional norms such as (un)fairness, responsibility, accountability and bias in knowledge graph construction, maintenance, and use? How can we build and use knowledge graphs which reflect legal or professional guidelines regarding bias? To what extent can auditing and compliance checking of knowledge graphs be automated?

Bias as a context problem

Bias as an ethical and societal problem is another important aspect, rooted in the context of the knowledge graph, since the knowledge graph cannot be generated without context, which is usually implicit. Typical examples include political and cultural statements. The serious issue of such ethical/societal bias can be exacerbated by the naive use of a knowledge graph,

[12] https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

and may even cause the denial of a whole knowledge graph (for example, some nations forbid the use of Wikipedia). Data and knowledge graph quality methodologies and methods are also typically context-driven. Questions to explored include understanding the relationships between bias and quality of knowledge graphs.

5.6.2 Possible Approaches

Bias detection

Detecting biases is not a straightforward task, requiring knowledge of the world. For example, observing in the data that there are more females married to males than vice versa is an observed representation bias. On the other hand, observing in the data that there are far more male than female Nobel prize winners is not a representation bias in the data, but an accurate rendition of the state of the world.

Moreover, identifying bias requires some intuition of what to look for. While some sensitive attributes (e.g., gender or nationality) are quite straightforward, others are not, and may require several iterations of observing downstream behaviour in a system using a knowledge graph. For example, in [5], it was found that different language editions of DBpedia have a different information density of movies with respect to their genre – a representation bias that would be hard to anticipate without observing it in a downstream task.

Methods for detecting bias suggested in the literature so far often anticipate that the user knows which bias to look for, and then query the knowledge graph to get some statistics out (e.g., the proportion of male and female subjects in statements with a given property). A more open approach to this would be to learn patterns from the graphs, and then let a user decide whether those patterns represent biases in the data or distributions in the real world.

Representing and documenting bias

In contrast to language models or image classification systems, where bias can only be detected implicitly, and explicit bias descriptions have to be added separately, KGs offer the means for an explicit internal representation of bias, legal norms, and further guidelines by definition. Once bias is detected, it would be helpful to document it. If the bias occurs in the form of some pattern, this could be done using a pattern description language, such as SHACL.[13] Moreover, some statistical information would be required, as, e.g., defined by the VoiD vocabulary.[14]

Handling bias

Documenting bias is a first step to handling bias, but it is not the end of the line. Depending on the requirements and task at hand, different ways of further handling bias are possible. Applying negotiation protocols is an option for dealing with conflicting information, but may not be possible for truly controversial information. In such cases, the authors of [6] suggest allowing controversial information with additional metadata. Depending on the task at hand, bias may also be removed or handled by means of resampling methods. However, as the experiment reported above shows, this might not always be an efficient method.

---

[13] https://www.w3.org/TR/shacl/
[14] https://www.w3.org/TR/void/#statistics

5.6.3 Open Research Questions

Handling Bias

The *bias as a signal* observation gives rise to the conclusion that bias – although a mostly negatively connoted term – can also be helpful, and that blindly removing bias is therefore not the only or necessarily best option. Instead, more sophisticated ways of dealing with bias are required.

Detecting bias

As discussed above, bias detection requires world knowledge. For a fully automated detection of bias, one would therefore need a fully objective and bias-free knowledge base of world knowledge. This presumption – which, at least today, is impossible to meet – shows that fully automatic bias detection is currently impossible. Therefore, manual intervention will be required to detect bias in knowledge graphs, and the processes and models to do so in the best and most efficient way are still to be explored. This is not a solely computational issue. Rather, people with various disciplines should commit to the whole life cycle form generation to use of knowledge graph (diversity and inclusion issue).

Representing bias

As discussed above, a bias that is detected in a knowledge graph should at least be documented. However, to the best of our knowledge, no standards for documenting biases exist so far. Therefore, the representation of bias is still an open research issue.

Bias needs not only to be documented for humans, but also machines. Once a standard for bias representation has been defined, it would be another open question of how subsequent steps, e.g., machine learning operations, may be informed about that bias, and then how to carry out appropriate remedies (e.g., by internally re-sampling the data).

Compliance and auditing

How could we support KG engineers in building legally compliant KGs, and how could we support government bodies in (semi-)automated auditing of KGs for legal compliance (e.g., EU regulations on responsible data and AI)? How would these goals be balanced with methods and tools for bias negotiation (e.g., legal compliance vs. protecting personal safety)? More generally, further work at the intersection KG engineering and Computational Law is called for.

5.6.4 Next Steps

Targeted directions for continuing this discussion include:
- A vision paper, fleshing out the roadmap sketched above.
- Activities to bridge the knowledge graph engineering community and scholars working in the legal, professional, and cultural, societal aspects of data and knowledge. Avenues here include multidisciplinary workshops, panel discussions, Dagstuhl Seminars, research consortia (e.g., EU COST action), and working groups.
- Standardization of vocabularies and standards for bias representation.

References

1 Krzysztof Janowicz, Bo Yan, Blake Regalia, Rui Zhu, and Gengchen Mai. Debiasing knowledge graphs: Why female presidents are not like female popes. In Marieke van Erp, Medha Atre, Vanessa López, Kavitha Srinivas, and Carolina Fortuna, editors, *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th – to – 12th, 2018*, volume 2180 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

2 Andreea Iana and Heiko Paulheim. More is not always better: The negative impact of a-box materialization on rdf2vec knowledge graph embeddings. *arXiv preprint arXiv:2009.00318*, 2020.

3 Eduardo Graells-Garrido, Mounia Lalmas, and Filippo Menczer. First women, second sex: Gender bias in wikipedia. In *Proceedings of the 26th ACM conference on hypertext & social media*, pages 165–174, 2015.

4 Urs Gasser and Carolyn Schmitt. The Role of Professional Norms in the Governance of Artificial Intelligence. In *The Oxford Handbook of Ethics of AI*. Oxford University Press, 2020.

5 Michael Matthias Voit and Heiko Paulheim. Bias in knowledge graphs–an empirical study with movie recommendation and different language editions of dbpedia. *arXiv preprint arXiv:2105.00674*, 2021.

6 Gianluca Demartini. Implicit bias in crowdsourced knowledge graphs. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 624–630, 2019.

5.7 Generating User and Developer Buy-in

Filip Ilievski (University of Southern California – Marina del Rey, US, ilievski@isi.edu)
Lydia Pintcher (Wikimedia Deutschland – Berlin, DE, lydia.pintscher@wikimedia.de)
Florian Reitz (Schloss Dagstuhl LZI - Trier, dblp group, DE)
Bradley P. Allen (Merit International, Inc. – Millbrae, US, bradley.p.allen@gmail.com)
Axel-Cyrille Ngonga Ngomo (Universität Paderborn, DE, axel.ngonga@upb.de)
Katherine Thornton (Yale University Library – New Haven, US, katherine.thornton@yale.edu)
Paul Groth (University of Amsterdam, NL, p.t.groth@uva.nl)
Denny Vrandečić (Wikimedia Foundation – San Francisco, US, denny@wikimedia.org)

This working group focused on how knowledge graphs can be made more attractive for regular developers of applications and services. We wanted to figure out how to make the vast amount of value created in resources like Wikidata accessible to a broad developer audience.

5.7.1 Discussed Problems

We discussed how to make accessing the data available in knowledge graphs quicker, cheaper and more efficient for developers of applications and services, especially those who have not been in contact with knowledge graphs before. This is becoming especially relevant as artificial intelligence and machine learning systems are becoming more prevalent and knowledge graphs can be a powerful tool to improve them.

Developers trying to work with knowledge graphs are facing a number of pain points. We discussed the various pain points we encountered in our own work with developers building applications and services on top of different knowledge graphs. Some of these pain points relate to the data in the knowledge graph and some to the tooling around that data. The following pain points were identified:

- Getting incorrect answers to queries: Developers are getting incorrect answers to their queries, which directly harms adoption and trust. This may be caused by issues in the data, the modeling of the data or the query itself.
- Dislike of identifiers, especially opaque identifiers: Many developers seem to dislike the prevalent use of identifiers in knowledge graphs, especially opaque ones like they are used for example in Wikidata. Developers want to use human-readable labels in their code instead.
- Schema discovery: The same data can often be modeled in different ways in a knowledge graph. To write queries that give them the answers they need, developers need to first get an understanding of how the data they are interested in is modeled. This can be challenging, especially if exploratory tools are not at hand.
- Adapting to new interfaces: There are various user interfaces developers are expected to work with when developing with data from a knowledge graph such as a query UI. These have a learning curve.
- Unclear and unhelpful error messages: When writing queries developers make mistakes and sometimes produce syntactically invalid queries. The errors they get back from the query systems are often not helpful for them to identify the problem and improve their query.

During the discussion, it became clear that more work is needed to define the exact target group of this developer outreach to make it successful. We need to better understand their needs, motivations, additional pain points and the environment they are working in. We also need to articulate more clearly what problem areas knowledge graphs are particularly well equipped to solve. A list of prototypical example use cases was considered particularly helpful in addition.

It might help to analyze positive existing use cases for KGs in a commercial setting, including data unit testing, content enrichment, geographical visualization, easy access to multilingual labels, and infobox extraction with a single query. It also seems helpful to understand the experience and the motivation of the library community, which has bought into knowledge graphs, perhaps after being shown how Wikidata can answer questions that could not be answered before. However, applying the same approach to other communities and use cases might bring novel challenges.

5.7.2 Possible Approaches

To facilitate value creation for software developers brought by knowledge technologies, we propose a combination of the following eight approaches:

1. **Conduct user studies** to better understand the target group. Software developers should be asked to perform representative tasks, and monitoring tools should be included to understand their mental model. Users should be asked to provide feedback on what was easy, what was difficult, what went wrong, and what could be improved.
2. **Log user actions** to help us understand typical user needs that are expressed through their queries.
3. **Provide useful knowledge subsets** from Wikidata that users can easily download and plug in their tools. These subsets should be provided in a developer-friendly format, like TSV or JSON.
4. **Provide users with atomic functions for common operations,** based on the persona needs derived from user studies and logging (points 1 and 2). Some operations would

include getting labels and aliases for an item, describing an item, query for similar items, text search for entities or events, fact extraction, and extracting a subset for reuse.

5. **Provide example use cases** as Jupyter and Colab notebooks, to illustrate the simplicity, effectiveness, and efficiency of including knowledge technologies in developer tasks. This should include a discussion on why this technology is the one to use.
6. **Enable users to develop a proof of concept (POC)** quickly, based on the atomic functions and the example use cases. This POC is primarily meant to convince developers, but it is also essential to secure management buy-in, as people are best convinced by showing, not telling.
7. **Make knowledge technologies relevant to the developer world,** by designing them to follow developer best practices as closely as possible, including graphical interfaces, APIs, data unit testing, and GitHub actions and releases. We should not expect developers to change their habits and make sacrifices to adopt knowledge technologies.
8. **Solicit developer feedback** to understand remaining pain points and listen to their suggestions for further improvement.

5.7.3 Open Research Questions

There seems to be a limited understanding of knowledge graph adaptations by developers. As a first step, we need a better understanding of why developers are adopting knowledge graphs and – most importantly – why not. For this research, we first need a clear understanding of the developer role and the different applications for which knowledge graphs can be adopted. The developer role needs to be clearly distinguished from other roles, such as data engineer and end user. In a second step, we need to determine what categories of information are needed and how the pain points outlined above hinder the adoption of knowledge graphs. Amongst others, we have to look into aspects such as:

Data access and presentation: What types of interfaces do developers require and how are their requirements fulfilled by current tools? E.g., how are SPARQL endpoints perceived as points of access? What are advantages and what are problems that developers face when using them? In what ways should the data be represented, i.e., triple-based formats vs. other data formats?. A possible result would be a list of atomic interaction patterns (such as API calls) that are considered to be beneficial or to be avoided when providing a knowledge graph data interface.

Tool requirements: What tools are needed to access a knowledge graph – again, specifically from the perspective of a developer? E.g., what data inspection/visualization tools are needed and how do requirements for these tools differ from requirements of other roles?

Data ownership: What role does data ownership play in knowledge graph adoption? What are the hurdles/concerns in using a shared knowledge graph such as Wikidata, particularly in a commercial setting? What roles do data quality and trust issues play in adopting shared knowledge graphs?

In a second step we need to develop new or improve existing tools to better align with the needs of developers and to provide an overall better experience. These tools need to be evaluated based on the understanding of user requirements we developed. We also need to evaluate the usefulness of the possible approaches mentioned above to increase the buy-in of developers.

5.7.4 Next Steps

1. **Research on persona descriptions and needs** – The most urgent challenge with developer buy-in is social rather than technical. We need to understand what we mean exactly by developers, what are their knowledge needs, and what is their current workflow. A possible entry point is an existing user group in the library community, which has seen the value of knowledge and has gradually embraced knowledge technologies in its practices.
2. **Map of different knowledge types in relation to representation formats** – Knowledge graphs are likely not the optimal format to store every kind of data. Geo-coordinates, for instance, might be stored more efficiently in a database that supports numeric querying with high precision, and text-heavy knowledge might be better captured by language models or ElasticSearch indices. It is essential to provide a rule of thumb for which kind of representation and resource should be the primary source for which kind of knowledge. A comprehensive figure or webpage would be a great initial format for such a map.
3. **Cookbook style documentation for developers** – Performing a knowledge technology task is overwhelming without understanding the landscape of available knowledge graphs and tooling. This could be a challenge at the beginning of using this technology, but also later in the process. To improve the developer experience, we aim to develop cookbook-style documentation that will enable developers to find relevant knowledge sources and tools as efficient as possible. The cookbook should ideally also include example use cases with code as supplementary material.
4. **Release data subsets with high reuse potential** – Well-understood datasets like MNIST have been key drivers of user-friendly tools in data science, like scikit-learn. Similarly, developing high-quality Wikidata subsets with high reuse potential, like a list of all countries or English labels, will provide an attractive playground for developers and inspire them to include ready knowledge in their frameworks. The downloads of the published subsets should be tracked to understand which, if any, are adopted by developers.

5.8 A Core Knowledge Engineering Methodology for Knowledge Graphs

Eva Blomqvist (Linköping University, SE)
Deborah McGuinness (Rensselaer Polytechnic Institute – Troy, US)
Valentina Presutti (University of Bologna, IT)
Marta Sabou (Vienna University of Economics and Business, AT)
Juan Sequeda (data.world – Austin, US)
Steffen Staab (Universität Stuttgart, DE)

This working group focused on the state of Knowledge Engineering (K) methodologies today, their relation to ontology engineering methodologies and other types of methodologies, as well as identifying open issues and needs in this domain.

5.8.1 Discussed Problems

The discussion started with an inventory of methodologies used by the participants and experiences of using such methodologies for knowledge engineering in the past. The inventory of methodologies included mainly ontology engineering methodologies [1], [2], [3], [4], [5] and others. Some of the issues discussed were:

- Lack of transparency, e.g. provenance data, in many projects. Auditing the project can be an important tool to understand what went well and what did not.
- Methodologies need to be agile, to some extent. Waterfall-style processes do not work. Also the setup of the projects are different, i.e. centralized of decentralized.
- Methodologies need to start from use cases, and most of them do, but they differ in how much guidelines are given on how to actually capture and describe use cases, and to elicit requirements from them. A good template is essential, and it should include Competency Questions. It is also important to be able to pick or tailor the methodology base on the type of use case. An enterprise data integration project has different needs than a collaborative open data project. In addition, most projects also need to cater for the unknown use cases of tomorrow – how can a methodology incorporate that?
- Costs of the methodology should be considered – KE is expensive. There needs to be guidelines on how to reduce the costs, and adapt methodology to the available resources. Human-machine collaboration, and crowdsourcing can be such means to reduce costs.
- Types of stakeholders and users, and different roles of users, is another important aspect that a methodology should cover.
- Reuse, e.g. both of existing data artefacts, code lists, and existing ontologies, are not in focus of most methodologies, but often an ad-hoc add-on activity. Also design patterns is an important kind of reuse of best practices and proven solutions.
- Evaluation and testing is often overlooked, or restricted only to assessing basic sanity criteria. Very few test-driven methodologies, and to some extent activities such as ontology testing are still under explored, compared to in for instance software engineering. Evaluation needs to be more structured and with better tool support. However, human-centric evaluation methodologies are also crucial.
- Focusing only on ontology engineering is too restrictive. A KE methodology needs to include also data curation, data integration/mapping, population of the knowledge (graph), and should put the project into its context, e.g. software engineering.
- Current tool support is far from perfect, and new tools are emerging to automate additional steps in KE methodologies, e.g. through ML approaches and language models. Most tools operate on the triple level, but Knowledge Engineers, and in particular domain experts, think in terms of other conceptual units, i.e. more complex structures.
- Although methodologies should not be too prescriptive, knowledge engineers that are not experts need a good cookbook, with rules of thumb etc.

An observation of the group was that at a high level existing methodologies are quite similar, and can be updated and consolidated to give a more coherent view of the KE processes. However, they are also to some extent lacking in that they do not cover the whole process, and do not take into account the relation to, for instance, whether the project takes place in a software engineering context, is conducted more independently, or in an open collaborative setting, such as crowd-sourcing.

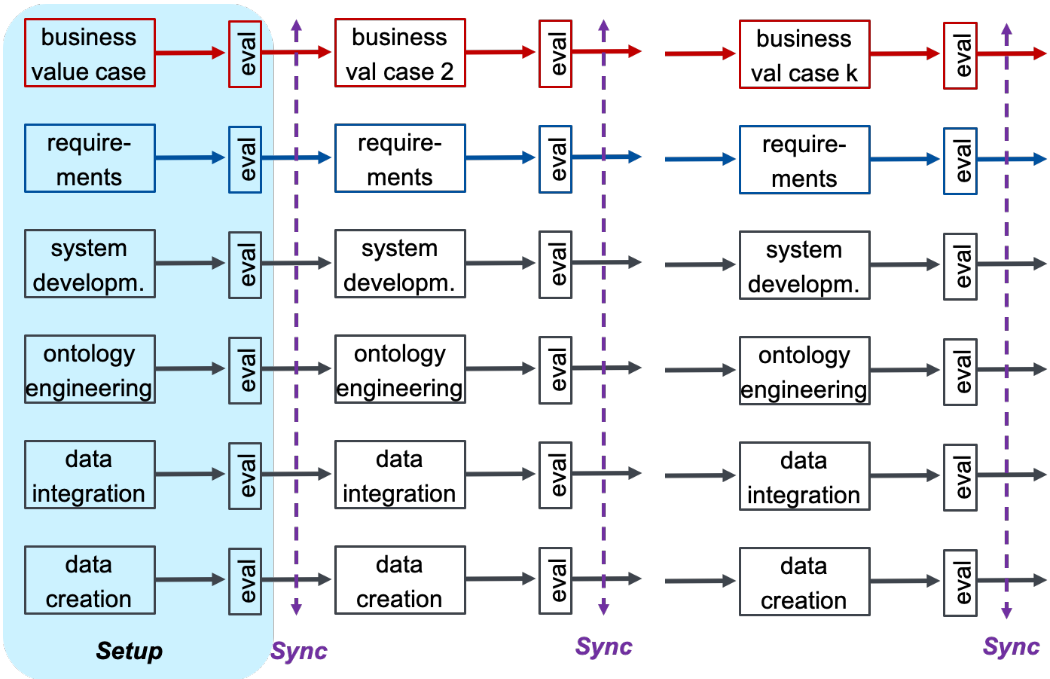


Figure 10 A blueprint knowledge engineering process that connects multiple methodologies for developing knowledge graphs.

5.8.2 Possible Approaches

There are many existing methodologies, both earlier KE methodologies and current ontology engineering methodologies. At an abstract level these are often quite similar, e.g. iterative processes to incrementally build up the knowledge model, but they are also often too narrow in scope, since they do not take into account the interaction with the context in which the KE process happens. Such context can for instance be a software project, intended to provide some business value to a company. In addition, many such methodologies also do not take into account the population of the knowledge model being built, i.e. the data integration and curation efforts needed to put the knowledge into use. Therefore some work is needed that considers the overall picture of KE in context, as illustrated in Figure 10.

Each step in such a process, i.e. the boxes in the figure, can then be more or less automated, and supported by various tools and detailed methodologies. However, the overall core KE methodology will still remain largely the same. A similar effort was also presented in [6].

5.8.3 Open Research Questions

- How does the sync between the methodologies in Figure 10 actually happen?
- What kind of evaluations are to be performed in each step, and overall?
- How can certain steps be automated or crowdsourced, and to what extent? What are the quality implications?
- How do current Knowledge Engineers actually perform these steps in practice? What are the bottlenecks and challenges?

5.8.4 Next Steps

Several possibilities for follow-up publications are discussed and will be pursued by the working group.

References

1. S. Staab, R. Studer, H.-P. Schnurr, and Y. Sure. Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1):26–34, 2001.
2. Juan F. Sequeda and Daniel P. Miranker. A pay-as-you-go methodology for ontology-based data access. *IEEE Internet Computing*, 21(2):92–96, 2017.
3. Eva Blomqvist, Karl Hammar, and Valentina Presutti. Engineering ontologies with patterns-the extreme design methodology. *Ontology Engineering with Ontology Design Patterns – Foundation and Applications*, (25):23–50, 2016.
4. James L Benedict, Deborah L McGuinness, and Peter Fox. A semantic web-based methodology for building conceptual models of scientific information. In *AGU Fall Meeting Abstracts*, volume 2007, pages IN53A–0950, 2007.
5. María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. Lot: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence*, 111:104755, 2022.
6. Gytė Tamašauskaitė and Paul Groth. Defining a knowledge graph development process through a systematic review. *ACM Transactions on Software Engineering and Methodology*, 2022.

6 Conclusions and Open Questions

Advances in neural and symbolic AI approaches [1], including knowledge graphs, prompted us to organise a Dagstuhl Seminar to chart the next frontiers of knowledge engineering in this brave new world. Participants reviewed the past, current, and emerging landscape of approaches, practices, and tools in knowledge base and knowledge graph construction, identified open research questions, and proposed next steps to address them.

The knowledge graph life cycle was a focal point of discussion. There was consensus that we need a sustained effort to update and upgrade classical ontology engineering methodologies [2] and develop end-to-end open-source infrastructure to make the most of the latest neurosymbolic technologies and tools, hence taking knowledge engineering and knowledge graphs beyond structured and semi-structured data to other modalities.

There are several canonical examples of knowledge graph architectures in use today. Within organisations, knowledge graphs are instrumental to data, content, and knowledge management [3]. KG projects essentially follow variants of classical ontology engineering methodologies, supported by a range of platforms and specialist tools e.g., taxonomy/ontology editors, graph databases, semantic mappers etc. In conjunction with machine learning, knowledge graphs are also used in semantic search, zero-shot learning, dialogue systems and recommender systems as a source of knowledge and explanations. Some of the best known knowledge graphs today, for instance in web search (Google, Microsoft), social networks (LinkedIn), and intelligent assistants (Siri, Alexa) achieve scales that were inconceivable decades ago – this is possible only with the help of automation, in particular using the latest developments in machine learning including generative models pre-trained on huge amounts of online data. It was recognised at the seminar that this AI-centric architecture with human-in-the-loop is not well supported in terms of methodologies and end-to-end tools. Finally, a third category of knowledge graphs are open-source and built by lively online communities [4]. While they have found considerable adoption in research and practice, their success is difficult to replicate in closed, proprietary settings, though they do provide invaluable insight into the sociotechnical ecosystem in which knowledge is created and shared.

As knowledge graph construction is making use of increasingly sophisticated, yet opaque AI capabilities, knowledge engineering must, like any other community using AI face its fairness, accountability, and transparency challenges. Several break-out workshops during the seminar considered common trustworthy AI concerns such as interpretability, biases, as well as human-AI interaction more generally, arguing for the need for bespoke solutions that target a range of end-users and stakeholders unique to knowledge engineering settings.

Finally, participants shared best practices and ideas to continue the knowledge and technology transfer efforts of the last two decades that have made knowledge graphs the backbone of systems as diverse as search engines, recommenders, chatbots, and enterprise data management platforms. They suggested activities to build capabilities and skillsets to use the latest neurosymbolic technologies and tools in knowledge graph construction, including tutorials, workshops, and hackathons, and agreed to work on joint frameworks and knowledge engineering methodologies. They also recognised the sustained need to promote knowledge graphs to the wider developer community and communicate their benefits, for instance, alongside neural methods.

As a community invested in knowledge representation and engineering, the participants embrace neural solutions such as language models for the step change they brought about in automating knowledge graph construction. At the same time, and looking back at decades of projects and experience with capturing knowledge in computational representations within organisations and on the open web, they are convinced that the use of such solutions will require human-in-the-loop approaches that are trusted and trustworthy. One of the reasons why enterprise knowledge graphs have become so successful is their ability to combine efficient, flexible storage of data with tractable representations of domain knowledge, while guaranteeing data integrity. If enterprise knowledge graph platforms are to adopt the latest advances in machine learning these guarantees will be as critical as ever. [15]

6.1 Continuing the Conversation

To continue the conversation, we provided organizers of EKAW 2022 the 23rd International Conference on Knowledge Engineering and Knowledge Management input for a *walkshop*. We prompted them with the following questions coming from the seminar:

- What ways does knowledge engineering deliver value today? What should be the requirements for knowledge production processes?
- What does user centric knowledge engineering look like including does it integrate into standard software engineering processes?
- How can new technologies help automate manual tasks such as knowledge elicitation, documentation, etc?
- How and to what extent do we integrate language models and knowledge engineering

[15] For an individual perspective of the seminar, we refer the reader to the trip report by Juan Sequeda: `http://www.juansequeda.com/blog/2022/09/20/knowledge-graphs-and-their-role-in-the-knowledge-engineering-of-the-21st-century-dagstuhl-trip-report/`

6.2 Open Questions

Finally, throughout this report we have identified many open questions for futher study. We itemize them here:

Knowledge Engineering Practice

- How does state of the art machine learning, including large language models augment knowledge engineering processes and projects?
- What is the existing user / developer experience of machine learning tools?
- What are the roles of people and machines in current knowledge engineering process?
- How do you evaluate the added value of automation to knowledge engineering processes?
- How do you control the outputs of ML-based systems are what you need for knowledge engineering?
- What is the interplay knowledge graph engineering (with or without AI) and system engineering?
- How doe we synchronize knowledge engineering methodologies?
- What kind of evaluations are needed to be performed in each step, and overall for knowledge engineering methodologies?
- How can certain steps in knowledge engineering be automated or crowdsourced, and to what extent? What are the quality implications?
- How do current Knowledge Engineers actually perform methodological steps in practice? What are the bottlenecks and challenges?

Types of Knowledge

- What are cases and the types of knowledge that can or should be relevant for people.
- Provide one or several "spectra" of expressiveness and other dimensions, for the knowledge that can be represented in KGs.
- How to allow for a knowledge based system that integrates different kinds of knowledge representation, but yet allow for a unified query interface?
- What types of tasks require which kind of knowledge representations?
- Language models are good in smoothly dealing with the brittleness problem of symbolic knowledge representations. How can we combine language models with knowledge graphs to gain the advantage of language models?
- Would increased use of data types be advantageous for knowledge graph engineering? Is that a potential approach for combining knowledge graphs with more knowledge representations?
- Can we separate individual facts or knowledge out of a language model, and store it in a more efficient representation, and thus save on parameters that would encode that knowledge, making them smaller and more efficient, while allowing them to access a knowledge graph?
- How could we track provenance for language models? How could we represent and explain where this response came from? How would we trace the lineage of statements in large language models?
- How can language models be updated?
- How can language models be adapted to better cover the long tail, emerging knowledge, etc.?

Explanations and Bias

- What is a clear specification of the relation between explanations and interpretations?
- What are measures for explanation and methods to evaluate them and to quantify their trustworthiness (both intrinsically and extrinsically) and to allow for measures of uncertainty?

- How doe we detecting bias using knowledge?
- How to representing bias?
- How can steps, e.g., machine learning operations, be informed about bias, and then how to carry out appropriate remedies?
- How could we support KG engineers in building legally compliant KGs, and how could we support government bodies in (semi-)automated auditing of KGs for legal compliance (e.g., EU regulations on responsible data and AI)?
- How would these goals be balanced with methods and tools for bias negotiation (e.g., legal compliance vs. protecting personal safety)?

Developer Experience

- What types of interfaces do developers require and how are their requirements fulfilled by current tools?
- What are advantages and what are problems that developers face when using tools?
- In what ways should the data be represented, i.e., triple-based formats vs. other data formats?
- What tools are needed to access a knowledge graph – again, specifically from the perspective of a developer? E.g., what data inspection/visualization tools are needed and how do requirements for these tools differ from requirements of other roles?
- What role does data ownership play in knowledge graph adoption? What are the hurdles/concerns in using a shared knowledge graph such as Wikidata, particularly in a commercial setting?

References

1 Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*, 2019.

2 Elisa F Kendall and Deborah L McGuinness. Ontology engineering. *Synthesis Lectures on The Semantic Web: Theory and Technology*, 9(1):i–102, 2019.

3 Juan Sequeda and Ora Lassila. Designing and building enterprise knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, 11(1):1–165, 2021.

4 Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

Participants

- Marcel R. Ackermann
Schloss Dagstuhl – Trier, DE
- Mehwish Alam
FIZ Karlsruhe, DE
- Bradley Allen
Merit – Millbrae, US
- Sören Auer
TIB – Hannover, DE
- Eva Blomqvist
Linköping University, SE
- George Fletcher
Eindhoven University of Technology, NL
- Paul Groth
University of Amsterdam, NL
- Aidan Hogan
University of Chile – Santiago de Chile, CL
- Filip Ilievski
USC – Marina del Rey, US
- Antoine Isaac
Europeana Foundation – Den Haag, NL
- Diana Maynard
University of Sheffield, GB
- Deborah L. McGuinness
Rensselaer Polytechnic Institute – Troy, US
- Axel-Cyrille Ngonga Ngomo
Universität Paderborn, DE
- Heiko Paulheim
Universität Mannheim, DE
- Lydia Pintscher
Wikimedia – Germany, DE
- Valentina Presutti
University of Bologna, IT
- Florian Reitz
Schloss Dagstuhl – Trier, DE
- Marta Sabou
Wirtschaftsuniversität Wien, AT
- Harald Sack
FZ Karlsruhe, DE
- Stefan Schlobach
VU University Amsterdam, NL
- Juan F. Sequeda
data.world – Austin, US
- Elena Simperl
King's College London, GB
- Steffen Staab
Universität Stuttgart, DE
- Lise Stork
VU University Amsterdam, NL
- Hideaki Takeda
National Institute of Informatics – Tokyo, JP
- Katherine Thornton
Yale University Library – New Haven, US
- Marieke van Erp
KNAW Humanities Cluster – Amsterdam, NL
- Denny Vrandečić
Wikimedia – San Francisco, US

Report from Dagstuhl Seminar 22381

Rational Design of RiboNucleic Acids

Sven Findeiß*[1], Christoph Flamm*[2], and Yann Ponty*[3]

1 **Leipzig University , DE.** `sven@bioinf.uni-leipzig.de`
2 **University of Vienna, AT.** `xtof@tbi.univie.ac.at`
3 **Ecole Polytechnique - Palaiseau, FR.** `yann.ponty@lix.polytechnique.fr`

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 22381 "Rational Design of RiboNucleic Acids" (RNAs). The seminar covered a wide array of models, algorithmic strategies, molecular scales and modalities, all targeting *in silico design* of RNAs performing predefined biological functions. It consisted in a series of talks, each being allocated a generous time budget enabling frequent (welcomed!) interruptions and fruitful discussions. Applications of rational RNA design include mRNA vaccines; RNAs acting as sensors; self-replicating RNAs, relevant to RNA world/origin of life studies; populations of RNAs performing computations, e.g. through strand-displacement systems; RNA origamis forming nano-architectures through self-assembly; weakly interacting RNAs inducing the formation of droplets within cells through liquid-liquid phase separation. Those diverse applications are typically tackled by Bioinformatics-inclined scientists, contributing to distinct areas of life science and, as a result, somewhat isolated and sometimes unaware of similar pursuits in neighboring fields. The overarching goals of this meeting were to gather computational scientists from multiple fields, increase awareness of relevant efforts in distant communities, and ultimately contribute to a transversal perspective where RNA design becomes an object of study in itself.

Seminar September 18–23, 2022 – http://www.dagstuhl.de/22381

2012 ACM Subject Classification Applied computing → Bioinformatics; Applied computing → Molecular evolution; Applied computing → Molecular structural biology; Theory of computation → Discrete optimization; Theory of computation → Dynamic programming; Theory of computation → Parameterized complexity and exact algorithms; Mathematics of computing → Optimization with randomized search heuristics; Computing methodologies → Discrete space search; Computing methodologies → Reinforcement learning; Computing methodologies → Neural networks

Keywords and phrases RNA, RNA design, Inverse folding, RNA structure, mRNA design, RNA sensors, Co-transcriptional folding, Molecular evolution, Distant homology, Drug design

Digital Object Identifier 10.4230/DagRep.12.9.121

1 Executive Summary

Sven Findeiß
Christoph Flamm
Yann Ponty

Context and selected takeaways

RiboNucleic Acids (RNAs) are ubiquitous macromolecules within biological systems, capable of performing a wide range of regulatory and catalytic functions. This versatility can

* Editor / Organizer

Rational Design of RiboNucleic Acids, *Dagstuhl Reports*, Vol. 12, Issue 9, pp. 121–149
Editors: Sven Findeiß, Christoph Flamm, and Yann Ponty
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

be harnessed, and RNAs are increasingly utilized to accurately monitor and control biological processes [19], leading to RNA being found at the core of modern therapeutics [18]. It is therefore not surprising that the RNA-guided CRISPR-Cas9 editing [10], rewarded by the 2020 Nobel Prize in Chemistry, and mRNA-based vaccines [12], are at the forefront of modern biotechnology. For many functional RNA families [11], decades of research have produced a deep understanding of the sequence and structural basis underlying their biological function(s). Such studies, coupled with mature computational methods for structure prediction [23], have paved the way for a rational design of RNAs targeting a wide diversity of biological function [8, 2, 13].

Accordingly, RNA design has emerged as an exciting open computational problems in molecular biology. Owing to the discrete nature of RNA sequence and popular structural representations (e.g. secondary structure), RNA design has inspired the contribution of a large number of diverse algorithms [9, 20, 14, 4] for the inverse folding problem, i.e. the design of an RNA sequence which preferentially and effectively folds into a predefined (secondary) structure. Given the, recently established, NP-Hardness of the problem, even for minimal energy models [1], many of those algorithmic predictions are either heuristics, exponential-time or based on a variety of machine learning techniques.

More generally, RNA design addresses the generation of sequences of nucleotides targeting a given biological function. A non-exhaustive list of classic design objectives includes:
- Preferential adoption of one or multiple given structures (inverse folding);
- Sequence specific constraints such as an overall (di-)nucleotide composition [21], encoding of an amino-acid sequence (mRNA design), presence/absence of motifs [22];
- Adoption of different conformations upon presence of ligand (RNA switches and sensors) [3];
- Effective and specific interactions with targeted partners (RNAs, proteins) cascading into system-level regulatory effects [15, 16];
- Self-assembly into large scale architectures, ultimately adopting a predefined 3D shape (RNA origami) [6];
- Exploit co-transcriptional folding, and more general out-of-equilibrium regimes to perform computations (strand displacement systems, oritatami) [5]

Typical applications of design include novel therapeutic strategies, control principles for existing biological systems, or sensors for the presence of small molecules [3], but designed sequences can also provide an objective experimental assessment of functional hypotheses, where designs are synthesized and their effect on the cellular context can be tested *in vitro* and, in turn, *in vivo*.

Over the course of the seminar, we witnessed a substantial recent expansion of the scope of applications. Beyond classic but still challenging objectives of design, including riboswitches addressed by Talk 5.8, 5.9 and 5.21, messenger RNAs towards vaccine objectives mentioned by Talk 5.27, and CRISPR gRNAs mentioned by Talk 5.11, novel applications of RNA design emerged during the seminar. Talk 5.11 introduced SARS CoV 2 sensors based on strand displacement, Talk 5.15 addressed self-replicating ribozymes connected with origin-or-life questions, and Talk 5.6 explored rational design principles for repetitive RNAs inducing the formation of cellular droplets through liquid-liquid phase separation.

RNA Design as a discrete (inverse) optimization problem

The inverse folding problem, one of the central elements of RNA design, is a hard computational problem [1]. Although attracting a wide interest from the community, it is also one of the very few problems in computational biology whose complexity status has remained open for a long time (about three decades). This difficulty can be attributed to a lack of a suitable conceptual framework for inverse combinatorial problems. Indeed, inverse folding can be viewed as the search of a pre-image, in a function that maps each RNA sequence to its most stable conformation, the latter being computed using a polynomial, yet non-trivial, dynamic programming algorithm [23]. Natural generalizations virtually include any instance of inverse optimization problems, and could be of general interest to the Computer Science research field. Prior works in this direction have led to characterization of designable structures based on formal languages and graph theory [7], revealing strong connections to many subfields of computer science (for instance, between positive design and graph coloring).

In Talk 5.18 it was discussed that a flexible inverse folding approach, e.g. by allowing the extension of helices by at most one base pair, seems to be easier than keeping the problem strict. Such a flexibility in the structural objective of design was also emphasized as desirable by Talk 5.5. The problem of classical inverse folding can be extended from one to multiple target structures, and Talk 5.27 showed that this can be solved by an elegant dynamic programming approach that is fixed parameter tractable. The resulting framework was further generalized, and is not only applicable to RNA design, but also to apparently more distant problems such as the alignment of RNAs with pseudoknots. *In silico* designs and analysis depend on the accuracy of the applied energy model. In Talk 5.12 it has been underpinned that a systematic perturbation of parameters can be used to define a notion of robustness of individual parameters of an energy model, and help to improve prediction accuracy. Talk 5.28 revisited the inverse folding problem as an inverse optimization problem, and showed that many local structural motifs do not admit a design, with consequences to the space of designable structures, but raised fundamental questions on a relatively new flavor of optimization.

RNA Design in Structural Bioinformatics

Inverse folding also represents the ultimate test of our understanding of the mechanisms governing the folding of macromolecules. Given a set of folding rules (typically, an energy model), a synthesis of *in silico* designed sequences combined with high-throughput experiments (e.g., structure probing) enables an assessment of the compatibility of the determined structure with the initial target. Observed discrepancies can then be used to assess the quality of predictive models, especially those based on statistical potentials which may be prone to overfitting. Systematic local imprecisions can also be used to refine energy models, enabling the generation of better designs, whose iteration represents a virtuous circle, ultimately contributing to a better understanding of folding principles.

A nascent RNA molecule typically folds during its transcription. Frameworks to simulate this kinetically driven process can help to interpret experimental results (Talk 5.4) but as neither the simulation nor the experiment is perfect, quality assurance (Talk 5.14) of the *in silico* investigations is essential and results have to be interpret with caution, as for instance the mapping of time scales is a non trivial task. Finally, complex RNA hybridization networks are designed *in silico* to perform regulatory functions with complex temporal dynamics. A simplified kinetic model, introduced in Talk 5.26, for RNA/RNA hybridization represents an attractive evaluation model for the design of interactions.

At a much more detailed 3D level, Talk 5.17 showed that high-resolution experimental techniques can be used to observe dynamic behaviors, sometimes triggered by the binding of a ligand, and could inform future objective functions. Talk 5.24 and 5.16 described coarse-grained models amenable to molecular dynamics. Interestingly, the latter can be leveraged in order to study kinetics behaviors at the 3D level.

RNA Design in Synthetic Biology and Natural Computing

This line of research applies various engineering principles to the design, and construction of artificial biological devices. While initially focused on hijacking naturally-occurring regulatory functions through a copy/paste of evolved genetic material [17], the need for a precise control and for a modularity/orthogonality of constructs, has increasingly led to a *de novo* design based on nucleic acids. Recently, RNA has been successfully used as a material for the design of whole regulatory circuits, or for the construction of complex programmable shapes (RNA origamis [6]), with promising applications as biomaterial.

Software frameworks, like the ones presented in Talk 5.10 and Talk 5.22, make the construction of large DNA and RNA nano-structures possible. Those designs are not only adopting the right structure *in silico* according to combinatorial folding algorithms, but can also be validated by simulations (Talk 5.24) and microscopy (Talk 5.1). This observation suggests that the difficulty of design could stem from the compactness of targeted ncRNAs, while larger (but more regular) RNAs may be easier to design, an element that could inspire future theoretical studies.

In the course of the seminar it became evident that information from the 2D and 3D level need to be mapped onto each other. Design could therefore benefit from multiscale approaches: selecting candidates with 2D objectives, use coarse grained 3D analysis (Talk 5.16) and go to a full atom final validation for critical sub-regions. The curation of refined and non-redundant 3D RNA structures (Talk 5.2) and the systematic extraction of information from such a data set can help to investigate for instance structural features of modified bases or to propose isosteric structural mutations (Talk 5.19) in order to generalize the design from 2D to compact 3D architectures.

Programmable RNA folding can also be used as a computational model, allowing for the computation of complex programs based on cotranscriptional folding phenomena. RNA regulatory circuits can be used to emulate Boolean functions, allowing a precise and expressive control of regulatory networks at an early stage of the gene expression process. Talk 5.23 introduced RNA oritatami, a Turing-complete model of computation based on cotranscriptional folding inspired by cellular automata. Talk 5.7 described exciting applications of design to generate easily-checkable QR codes that reveal contamination in a closed environment. However an application-agnostic implementation of the strand displacement systems underlying some of those applications still represent major challenges in RNA, as shown and discussed during Talk 5.3. Those include intra-molecular base pairs and an overall wasteful behavior that motivates efficient recycling strategies.

RNA evolution and Machine Learning

The analysis of new RNA families, such as the pervasive and poorly understood lncRNAs or the numerous viral/bacterial non-coding RNAs observed in metagenomics experiments, relies critically on the identification of an evolutionary pressure, allowing to hypothesize new functions. Given a family of homologous RNAs sharing established functional traits, it is classic to asks whether an observed property, such as the occurrence of a common motif or a given covariation pattern, is likely to reveal a yet-unknown selective pressure or, conversely, is merely the consequence of established functional traits. Classic bioinformatics methods rephrase the problem in a hypothesis-testing framework, and compute the probability that a sequence, generated at random in a model that captures existing constraints, features the observed property. Ideally, such sequences should represent solutions to an instance of the design problem, target established functions, while respecting a distribution that can either be derived from the targeted function, or learned from data.

Talk 5.5 presented a context where rational design methodologies were utilized to capture remote homologs of a suspected, but scarcely-populated, functional family of ncRNAs. Generative models can also be used for design, in cases where the underlying model of function is partially understood, and should be learned from the data. Talk 5.8 used Restricted Boltzmann Machines (RBM), an unsupervised learning approach, to pickup the intricate probability distribution of the statistical features of a naturally-occurring riboswitch. The RBM was then used to generate novel instances with the same distribution of features, resulting in an enrichment of functional designs as revealed by experimental validation. Direct Coupling Analysis was also used in Talk 5.15 to generate self-replicating ribozymes, using a complex definition of function that may require some element of learning. Interestingly, the efficacy of designs was ultimately shown to benefit from further refinements using classic combinatorial methods for inverse folding, suggesting future hybrid ML/combinatorial methodologies.

While RNA design is an increasingly important computational task in molecular biology, nanotechnology and medicine, methods for computational rational design are still lacking for many applications. Moreover, many design tasks are currently addressed using algorithmic techniques (e.g. Markov chain Monte Carlo) that are clearly superseded by the state-of-the-art in algorithmic research. Conversely, computer scientists considering design tasks usually limit themselves to inverse folding, overlooking a rich bestiary of computational problems whose consideration would, in turn, undoubtedly lead to the emergence of new algorithmic paradigms.

Talk 5.20 showed that basic ML architectures can be learned in the context of reinforcement learning and can be successful for basic inverse folding of RNA. Talk 5.25 presented a complete design story, describing a methodology to advance our understanding of tRNAs. In particular, a mechanistic understanding of the target function can be gained by masking constraints during redesign, and the differentiability of the design problem can lead to great speedups of the computation. However, ML approaches may not always represent a silver bullet in RNA bioinformatics, and Talk 5.13 dramatically illustrated this in the context of RNA folding, a context where the quality and biases in the data strongly impacts, and probably hinders for intrinsic reasons, the predictive capabilities of deep learning-based methods. As a consequence, synthetic data can and should be used to test the capacity of learning architectures on simplified problems before embarking into "real life" learning.

References

1 Édouard Bonnet, Paweł Rzążewski, and Florian Sikora. Designing RNA secondary structures is hard. In Benjamin J. Raphael, editor, *Research in Computational Molecular Biology - 22nd Annual International Conference, RECOMB 2018*, volume 10812 of *Lecture Notes in Computer Science*, pages 248–250, Paris, 2018. Springer.

2 Matthew G. Costales, Jessica L. Childs-Disney, Hafeez S. Haniff, and Matthew D. Disney. How we think about targeting RNA with small molecules. *Journal of Medicinal Chemistry*, 63(17):8880–8900, 2020. PMID: 32212706.

3 Sven Findeiß, Maja Etzel, Sebastian Will, Mario Mörl, and Peter F Stadler. Design of artificial riboswitches as biosensors. *Sensors (Basel, Switzerland)*, 17(9):E1990, August 2017.

4 Juan Antonio Garcia-Martin, Peter Clote, and Ivan Dotu. RNAiFOLD: a constraint programming algorithm for RNA inverse folding and molecular design. *Journal of Bioinformatics and Computational Biology*, 11(2):1350001, 2013.

5 Cody Geary, Pierre-Étienne Meunier, Nicolas Schabanel, and Shinnosuke Seki. Oritatami: A Computational Model for Molecular Co-Transcriptional Folding. *International Journal of Molecular Sciences*, 20(9):2259, May 2019.

6 Cody Geary, Paul W. K. Rothemund, and Ebbe S. Andersen. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science*, 345(6198):799–804, 2014.

7 Jozef Haleš, Alice Héliou, Ján Maňuch, Yann Ponty, and Ladislav Stacho. Combinatorial RNA Design: Designability and Structure-Approximating Algorithm in Watson-Crick and Nussinov-Jacobson Energy Models. *Algorithmica*, 79(3):835–856, Nov 2017.

8 Stefan Hammer, Christian Günzel, Mario Mörl, and Sven Findeiß. Evolving methods for rational de novo design of functional RNA molecules. *Methods*, 161:54–63, may 2019.

9 I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsch. Chem.*, 125:167–188, 1994.

10 Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096):816–821, 2012.

11 Ioanna Kalvari, Joanna Argasinska, Natalia Quinones-Olvera, Eric P Nawrocki, Elena Rivas, Sean R Eddy, Alex Bateman, Robert D Finn, and Anton I Petrov. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids research*, 46:D335–D342, January 2018.

12 David M. Mauger, B. Joseph Cabral, Vladimir Presnyak, Stephen V. Su, David W. Reid, Brooke Goodman, Kristian Link, Nikhil Khatwani, John Reynders, Melissa J. Moore, and Iain J. McFadyen. mRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences*, 116(48):24075–24083, 2019.

13 Na Qu, Yachen Ying, Jinshan Qin, and Antony K Chen. Rational design of self-assembled RNA nanostructures for HIV-1 virus assembly blockade. *Nucleic Acids Research*, 50(8):e44–e44, 12 2021.

14 Vladimir Reinharz, Yann Ponty, and Jérôme Waldispühl. A weighted sampling algorithm for the design of RNA sequences with targeted secondary structure and nucleotide distribution. *Bioinformatics*, 29(13):i308–i315, 2013.

15 Guillermo Rodrigo, Thomas E Landrain, and Alfonso Jaramillo. De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proceedings of the National Academy of Sciences U S A*, 109(38):15271–6, 2012.

16 Guillermo Rodrigo, Thomas E. Landrain, Eszter Majer, José-Antonio Daròs, and Alfonso Jaramillo. Full design automation of multi-state RNA devices to program gene expression using energy-based optimization. *PLoS Computational Biology*, 9(8):e1003172, 2013.

17 E Westhof, B Masquida, and L Jaeger. RNA tectonics: towards RNA design. *Folding & design*, 1(4):R78—88, 1996.

18 Melanie Winkle, Sherien M. El-Daly, Muller Fabbri, and George A. Calin. Noncoding RNA therapeutics – challenges and potential solutions. *Nat Rev Drug Discov*, 20:629–651, 2021.

19 Sherry Y. Wu, Gabriel Lopez-Berestein, George A. Calin, and Anil K. Sood. RNAi therapies: Drugging the undruggable. *Science Translational Medicine*, 6(240):240ps7, 2014.

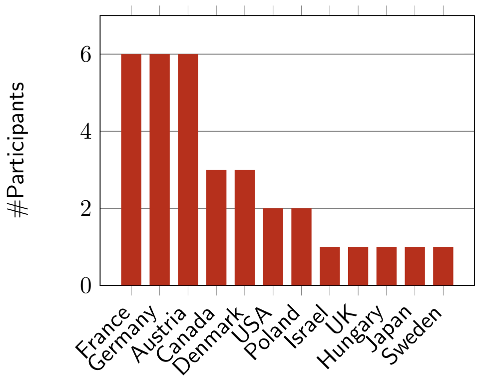
20 Joseph N Zadeh, Brian R Wolfe, and Niles A Pierce. Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry*, 32(3):439–52, 2011.

21 Yang Zhang, Yann Ponty, Mathieu Blanchette, Eric Lecuyer, and Jérôme Waldispühl. SPARCS: a web server to analyze (un)structured regions in coding RNA sequences. *Nucleic Acids Research*, 41(Web Server issue):W480–5, July 2013.

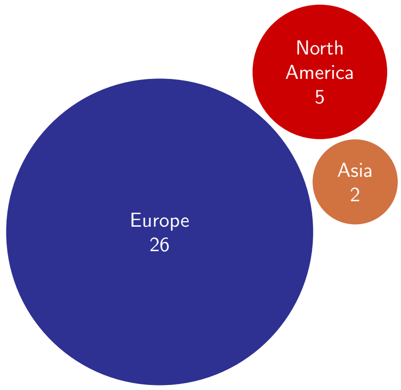
22 Yu Zhou, Yann Ponty, Stéphane Vialette, Jérôme Waldispühl, Yi Zhang, and Alain Denise. Flexible RNA design under structure and sequence constraints using formal languages. In Jing Gao, editor, *ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM-BCB 2013, Washington, DC, USA, September 22-25, 2013*, page 229. ACM, 2013.

23 M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.

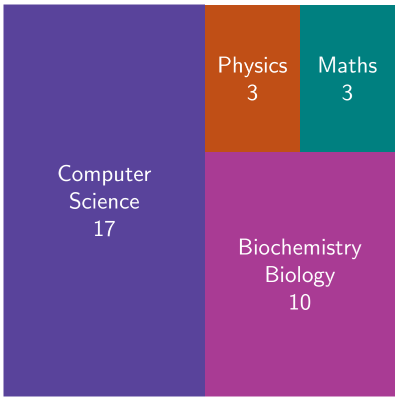
2 Table of Contents



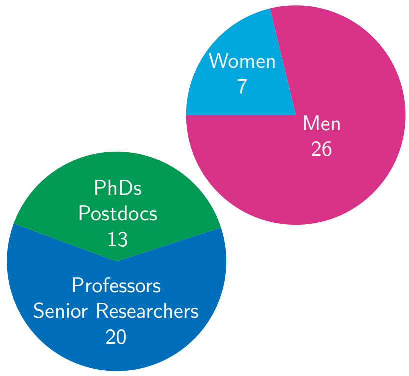
(a) Country of origin.



(b) Continent of institutions.



(c) Primary disciplinary background.



(d) Gender balance and academic seniority.

Figure 1 Participants statistics. Breakdown of workshop participants by country/continent of employing institution, primary disciplinary background, gender and academic seniority.

3 Participants and Group Composition

The list of participants included 33 researchers, selected for their outstanding contributions to RNA design, and/or their outstanding potential to impact future developments of the field. It should be noted that, in addition to those ultimate participants, a dozen confirmed researchers had to cancel, partly due to the ongoing pandemics.

As shown in Figures 1a and 1b, participants primarily originated from European institutions (26/33), but also from North America (5/33) and Asia (2/33), with five countries (Austria, France, Germany, Canada and Denmark) representing almost three quarters (24/33) of participants. While this concentration largely reflects the main centers of research on RNA bioinformatics, combined with the European location of the seminar center, the organizers regretted the absence of key players from North America and Asia, and will take this fact into consideration upon (possibly) organizing future editions of the seminar.

A key aspect of RNA design is that it requires a constant interdisciplinary dialogue, involving partners originating from diverse fields of research. Those include computational scientists to design algorithms and methods, modelers and experts in biochemistry to formulate models that are both accurate and computationally tractable, and end users/stakeholders from the fields of biology, biotechnology and medicine to assess the suitability of current design models and inform future developments. The organizers are proud to report that the seminar was able to strike such a critical interdisciplinary mix, as show in Figure 1c. While roughly half (17/33) of our participants were originally trained in Computer Science, one third (10/33) had initial background in Biology or Biochemistry, with the remaining participants being equally split between Physics and Mathematics (3/33 each). Interdisciplinary research also notoriously benefits from an early exposure of junior scientists to other fields of research. As shown in Figure 1d, the organizers were pleased to witness the presence of two fifths (13/33) of junior scientists (PhD candidates or Postdocs), among a majority (20/33) of more established scientists (Asst/Full prof. or permanent researchers).

Finally, partly due to late cancellation and despite a conscious mitigation effort by the organizers, the gender representation among participants certainly showed imbalance, with only a fifth of female researchers (7/33), Figure 1d. One possible reason, mentioned by some prospective participants, is the lack of support for daycare options (esp. towards small infants), which we understand is being considered by the center. While the organizers believe that this aspect is only partly within their control, it will nevertheless be the object of an increased focus while organizing future seminars.

4 Overall Organization and Schedule

Firstly, we wish to stress the impact of the ongoing COVID 19 pandemics on the organization of this seminar. Beyond the above-mentioned cancellations, the seminar was originally envisioned in February 2019, submitted in April 2019 and accepted in July of 2019, to be held in October 2020 and finally canceled due to a deep resurgence of COVID in Europe. A proposal was resubmitted in November 2020, and accepted in February 2021, to be finally held in September 2022. Overall, this workshop has been 3 1/2 years in the making, and the organizers were particularly excited to see it finally happen after such uncertainty.

The seminar itself consisted of talks, mostly scheduled before the seminar, while leaving ample time to impromptu discussions and spontaneous talk propositions. Wednesday afternoon was intentionally left open, to allow participants to interact in a less formal/public environment. We also finished the seminar after lunch on Friday, to allow most participants to reach home before the weekend. This left us with sufficient time to feature 28 talks, each of a duration of 30 or 45 minutes, structured in 7 sessions.

The seminar started with a joint talk by the organizers on Monday morning, aiming at providing sufficient context for all participants to follow and maximally benefit from the remaining talks. The afternoon session, named *Design Stories*, consisted of success stories in RNA design, a topic which we reasoned would expose most participants to the diversity of objectives required by applications, and base our future discussions on realistic use-cases. Tuesday morning's *Molecular Biology* session mentioned topics at the interplay of molecular modeling and evolution, while the afternoon's *Molecular Computing* session showcased design challenges and solutions arising in a context where RNA is used as a programmable material, capable of self-assembly and computation. The sole Wednesday morning session was dedicated to *Machine Learning* in the context of design, with a strong emphasis on generative models being used as a substitute for the classic specification/implementation philosophy. On Thursday, the morning *Combinatorial Design* session was dedicated to algorithmic and enumerative considerations of the yet-unsolved inverse folding problem, while the afternoon session focused on *3D and Design*, a very challenging context whose objective functions are still to be defined. Further increasing in difficulty, the Friday morning *Design for Dynamic Landscapes* session closed the seminar with contributions towards the design of RNA folding or interacting kinetically, out of the equilibrium regime.

Due to the interdisciplinary nature of the audience and topic, we ended up not eliciting to formally include an *open problem* session, although some formalized problems were mentioned during talks, and more thoroughly explored in smaller groups, past dinnertime.

5 Overview of Talks

5.1 Structural basis of RNA origami design, folding and flexibility

Ebbe Sloth Andersen (Aarhus University, DK)

Joint work of Ebbe Sloth Andersen, Helena Rasmussen, Ewan K. S. McRae, Jianfang Liu; Andreas Bøggild, Michael Truong-Giang Nguyen, Néstor Sampedro Vallina, Thomas Boesen, Jan Skov Pedersen, Gnag Ren, Cody Geary

The research field of RNA nanotechnology develops methods for the rational design of self-assembling RNA nanostructures with applications in nanomedicine and synthetic biology. Inspired by the cotranscriptional folding of biological RNA molecules, we developed the RNA origami method to design RNA nanostructures compatible with cotranscriptional folding [1, 2], advantageous for large-scale production *in vitro* and expression *in vivo*. However, advancing this technology further will require a better understanding of RNA structural properties and the non-equilibrium dynamics of the cotranscriptional folding process. Here, we use cryogenic electron microscopy to study a panel of RNA origami structures at sub-nanometer resolution revealing structural parameters of kissing loop and crossover motifs, that are further used to optimize designs by reduction of internal strain and global twist. In three-dimensional bundle designs, we discover a novel kinetic folding trap that forms during cotranscriptional folding and is only released 10-12 hours after transcription start. We characterize the conformational landscape of RNA origamis to reveal the RNA flexibility of helices and structural motifs. Finally, we demonstrate that large distinctive RNA origami shapes are visible by cryo-electron tomography pointing to potential use as markers in cellular environments. Our results improve understanding of RNA structure, folding, and dynamics, providing a basis for rational design of genetically encoded RNA nanodevices.

References

1 Geary, C., Rothemund, P. W. and Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. Science 345, 799-804, doi:10.1126/science.1253920 (2014).
2 Geary, C., Grossi, G., McRae, E. K. S., Rothemund, P. W. K. and Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. Nat Chem, doi:10.1038/s41557-021-00679-1 (2021).

5.2 Datasets for benchmarking RNA design algorithms

Maciej Antczak (Poznan University of Technology, PL)

Joint work of Maciej Antczak, Marta Szachniuk
Main reference Bartosz Adamczyk, Maciej Antczak, Marta Szachniuk: "RNAsolo: a repository of cleaned PDB-derived RNA 3D structures", Bioinform., Vol. 38(14), pp. 3668–3670, 2022.
URL https://doi.org/10.1093/bioinformatics/btac386
Main reference Jakub Wiedemann, Jacek Kaczor, Maciej Milostan, Tomasz Zok, Jacek Blazewicz, Marta Szachniuk, Maciej Antczak: "RNAloops: a database of RNA multiloops", Bioinform., Vol. 38(17), pp. 4200–4205, 2022.
URL https://doi.org/10.1093/bioinformatics/btac484

In this talk, we will present the databases developed to support benchmarking of bioinformatics algorithms targeting RNA, including the ones for RNA design. RNAsolo[1] collects experimentally determined 3D RNA structures from RNAs alone, protein-RNA complexes, and DNA-RNA hybrids and organizes them into classes of equivalent structures. Their sequences and tertiary structures are grouped in 192 benchmark sets ready for download and automated processing. RNAloops[2] aims to facilitate the study of multiloops in RNA molecules. It collects n-way junctions found in experimental RNA structures and allows to search them by sequence, secondary structure topology, or structure parameters. Both data sources address RNA-related studies by providing reliable sequence and structure data and efficient search facilities.

5.3 On the compilation of multi-stranded nucleic acids circuits

Stefan Badelt (Universität Wien, AT)

Main reference Stefan Badelt, Seung Woo Shin, Robert F. Johnson, Qing Dong, Chris Thachuk, Erik Winfree: "A General-Purpose CRN-to-DSD Compiler with Formal Verification, Optimization, and Simulation Capabilities", in Proc. of the DNA Computing and Molecular Programming - 23rd International Conference, DNA 23, Austin, TX, USA, September 24-28, 2017, Proceedings, Lecture Notes in Computer Science, Vol. 10467, pp. 232–248, Springer, 2017.
URL https://doi.org/10.1007/978-3-319-66799-7_15

Compilation from high level languages to low level languages is a fundamental concept in computer science, and it enables researchers to program silicon-based machines even though they have no understanding of assembler code or transistors. In previous work, we have shown that a description of nucleic acid circuits at the domain level (equipped with an approximate biophysical model for DNA), can be formally derived from a high level language, e.g. compilation from a boolean circuit to formal chemical reaction network to a domain-level strand displacement system.

The next challenge is to ensure a correct compilation from the domain-level system specification to the nucleotide level, which must involve both nucleic acid sequence design and a verification based on folding kinetics at the secondary structure level. Currently, we are exploring new sequence design techniques for large and complex nucleic acid reaction networks, that also systematically incorporate feedback from experimental work.

[1] `https://rnasolo.cs.put.poznan.pl/`
[2] `https://rnaloops.cs.put.poznan.pl/`

5.4 Simulations of cotranscriptional folding explain the impact of sequence mutations

Stefan Badelt (Universität Wien, AT)

Joint work of Stefan Badelt, Ronny Lorenz, Ivo L. Hofacker
Main reference Stefan Badelt, Ronny Lorenz, Ivo L. Hofacker: "DrTransformer: Heuristic cotranscriptional RNA folding using the nearest neighbor energy model", bioRxiv, Cold Spring Harbor Laboratory, 2022.
URL https://doi.org/10.1101/2022.09.08.507181

Cotranscriptional folding has the exciting potential to encode multiple functional important structures into a single molecule and visit them in a controlled manner. Unfortunately, the interpretation of experimental data on cotranscriptional folding is still heavily dependent on computational structure prediction, and it is easy to misinterpret data when using thermodynamic models. We use the stochastic simulator Kinfold, as well as a newly developed deterministic heuristic DrTransformer to show how cotranscriptional simulations can help with understanding experimental results and point out common mistakes in existing interpretations of data.

5.5 Eukaryotic riboswitch detection using inverse RNA folding

Danny Barash (Ben Gurion University - Beer Sheva, IL)

Joint work of Sumit Mukherjee, Matan Drory, Michelle Meyer, Danny Barash

The inverse RNA folding problem for designing sequences that fold into a given RNA secondary structure was introduced in the early 1990's in Vienna. By an extension of this problem we use a coarse-grained approach to possibly detect novel eukaryotic riboswitches. The approach can tentatively be used for other domains and applications.

5.6 Design of RNA tandem repeats creating RNA droplets forming liquid-liquid phase separation

Sarah Berkemer (Ecole Polytechnique - Palaiseau, FR)

Joint work of Sarah Berkemer, Ariel B Lindner, Yann Ponty, Carla Tous-Mayol

Various genetic disorders are caused by expansion of short tandem repeats as they aggregate in cells and form so-called RNA droplets or foci. However, themolecular mechanisms of RNA foci formation remains unclear. The aim of being able to design RNA tandem repeats and model RNA foci formation is twofold: it will help understand the mechanisms and therapies related to genetic disorders such as Huntigton's disease but at the same time serve as a method to spatial engineering inside cells as RNA droplets cause a liquid-liquid phase separation which can serve as process isolation and help to organize proteins and multienzyme pathways without fine-tuning RNA expression levels.

Phase-separating RNA molecule complexes are contructed from small repeating sequences, e.g. triplet repeats. Visualization of RNA foci is conducted by tagging droplets with e.g. GFP and corresponding adapters such as the MS2 aptamer.

Previous studies successfully showed the formation of RNA foci using various types of RNA triplet repeats and even longer repeat sequences where the formation of G-quadruplexes seems to be an important part for the interaction between two tandem repeat RNAs [1, 2, 3, 4].

Existing studies could experimentally show which RNA triplets are the most successful in forming RNA foci, however, the structure of RNA foci and their dynamics are not yet understood. Additionally, the liquid-liquid phase separation opens numerous possibilities for spatial engineering inside the cells, but we still lack the knowledge of structural and chemical properties of the RNA droplets and the space inside the foci. By designing RNA molecules that form droplets, we need to take into account interactions of more than two RNA sequences as well as possible interactions with binding proteins. Hence, we aim to develop design strategies for interacting short tandem RNA repeats and explore properties of RNA droplets and their formation.

References

1 Haotian Guo et al., `https://www.biorxiv.org/content/10.1101/2020.07.02.182527v2.full`
2 Jain and Vale, 2017, `https://doi.org/10.1038/nature22386`
3 Nguyen, Hori and Thirumalai 2021: `https://doi.org/10.1101/2021.02.20.432119`
4 Isiktas et al., 2022 `https://www.biorxiv.org/content/10.1101/2022.04.11.487960v1.full`

5.7 Scaling and Limits of DNA Strand Displacement Computing

Harold Fellermann (Newcastle University, GB)

Since it has been shown that DNA strand displacement (DSD) reactions can implement arbitrary chemical reaction networks, they have become a popular substrate for molecular computing applications. One question worth asking if whether and where there exists upper limits on the size and complexity of realistically achievable DSD circuits. To address this question, I am presenting as example application the design for a molecular QR code generator that displays a dedicated QR code for any configuration of n molecular input DNA strands. By increasing the number of inputs, the complexity of the circuit increases in a superexponential manner, as well as our current attempts to tame the number of required DSD gates that implement the requred function. The second part of the talk presents experimental results on the scalability of DSD circuits and limits that arise from toehold occlusion or partially complementary toeholds. Motivation for this study is the realization that reversible toehold binding imposes an upper limit on toehold length of typically six to eight nucleotides. This in turn puts a hard limit on the number of distinct toeholds a circuit can employ. While the number of distinct nucleotide sequences is still quite large, crosstalk appears in system with significantly fewer toehold sequences once the sequence of supposedly distinct toeholds becomes similar enough to cause undesired interactions. We have systematically analyzed noise in DSD systems caused by crosstalk between signals with single and double mismatches. Our main result is that toehold occlusion might occur already in systems one order of magnitude below the theoretical upper limit to toehold domains.

5.8 Generative modelling of riboswitches with restricted Boltzmann machines

Jorge Fernández de Cossío Díaz (ENS - Paris, FR)

Joint work of Jorge Fernández de Cossío Díaz, Andrea Di Gioacchino, Simona Cocco, Rémi Monasson, Bruno Sargueil, Yann Ponty, Bertrand Marchand, Pierre Hardouin, Francois-Xavier Lyonnet

Restricted Boltzmann machines (RBM) are energy-based latent variable generative models, consisting of two layers, that can offer interpretable representations of complex data. Recently they have been applied to modelling protein sequence data. In this talk, I will present evidence suggesting RBM are effective generative models of structured RNA. In particular, I consider the SAM riboswitch family, which regulates expression of downstream bacterial mRNAs by adopting competing structural conformations in response to the presence of a cellular metabolite. The RBM automatically infers relevant statistical features from the sequence data, such as conservation patterns, complementarity constraints consistent with the secondary structure, and the presence of a pseudoknot. The functionality of designed sequences has been validated experimentally by SHAPE mapping.

5.9 Get away from Plug and Pray: Synthetic Riboswitches - Applications and open Problems

Sven Findeiß (Universität Leipzig, DE)

Joint work of Sven Findeiß, Mario Mörl, Peter F. Stadler

I will talk about the collaborative projects with the group of Mario Mörl (Biochemistry department at Leipzig University) on transcription termination regulating riboswitches and how we put tRNA processing under ligand control. The presentation will summarize how the corresponding design models have been developed, implemented, and analyzed *in silico*, as well as the biochemical investigations *in vitro* and *in vivo*. I will not only show the success story but the main emphasis will be on the problems we faced, how we solved them, and especially the issues that remain.

References

1 Wachsmuth M. et al., 2013, `https://doi.org/10.1093/nar/gks1330`
2 Findeiß S. et al., 2018, `https://doi.org/10.1016/j.ymeth.2018.04.036`
3 Günzel C., et al., 2020, `https://doi.org/10.1080/15476286.2020.1816336`
4 Ender A. et al., 2021, `https://doi.org/10.1093/nar/gkaa1282`

5.10 Designing RNA during the DNA Origami revolution

Cody Geary (Aarhus University, DK)

Joint work of Cody Geary, Guido Grossi, Ewan K. S. McRae, Paul W. K. Rothemund, Ebbe S. Andersen
Main reference Cody Geary, Guido Grossi, Ewan K. S. McRae, Paul W. K. Rothemund, Ebbe S. Andersen: "RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds". Nat. Chem. 13, 549–558 (2021).
URL https://doi.org/10.1038/s41557-021-00679-1

RNA is the punk-brother of DNA. While DNA plays by rules, RNA is more rebellious. The diverse structural features of RNA that make it a powerfully-functional molecule in biology also make it difficult to tame and rationally-design.

In contrast to engineered DNA nanostructures such as DNA origamis, natural RNA molecules in cells must fold under non-equilibrium conditions; the RNAs fold continuously while the strand is still emerging from the polymerase. While design of staple strands to produce DNA origami nanostructures can be easily automated by simple algorithms, producing a single-stranded RNA origami requires the entire sequence of the RNA to be designed by inverse folding, which is computationally much more challenging.

Our RNA design software ROAD begins with a random starting sequence, and over many iterations mutates that sequence to improve its folding into a target fold. ROAD uses both positive and negative design cycles to perform a gradient descent based on an adapting scoring function. The strategy is based on *in vitro* selection methods where the selection conditions gradually become more difficult over successive rounds.

5.11 Two design stories: probes for SARS-CoV-2 detection and CRISPR/Cas9 gRNAs

Jan Gorodkin (University of Copenhagen, DK)

Joint work of Jan Gorodkin, Mohsen Mohammadniaei, Ming Zhang, Jon Ashley, Ulf Bech Christensen, Jan Friis-Hansen, Rasmus Gregersen, Jan Grom Lisby, Thomas Lars Benfield, Finn Erland Nielsen
Main reference Mohsen Mohammadniaei, Ming Zhang, Jon Ashley, Ulf Bech Christensen, Lennart Jan Friis-Hansen, Rasmus Gregersen, Jan Gorm Lisby, Thomas Lars Benfield, Finn Erland Nielsen, Jens Henning Rasmussen, Ellen Bøtker Pedersen, Anne Christine Rye Olinger, Lærke Tørring Kolding, Maryam Naseri, Tao Zheng, Wentao Wang, Jan Gorodkin, Yi Sun: "A non-enzymatic, isothermal strand displacement and amplification assay for rapid detection of SARS-CoV-2 RNA". Nat Commun 12, 5089 (2021).
URL https://doi.org/10.1038/s41467-021-25387-9
Main reference Xi Xiang, Giulia I. Corsi, Christian Anthon, Kunli Qu, Xiaoguang Pan, Xue Liang, Peng Han, Zhanying Dong, Lijun Liu, Jiayan Zhong, Tao Ma, Jinbao Wang, Xiuqing Zhang, Hui Jiang, Fengping Xu, Xin Liu, Xun Xu, Jian Wang, Huanming Yang, Lars Bolund, George M. Church, Lin Lin, Jan Gorodkin, Yonglun Luo: "Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning". Nat Commun 12, 3238 (2021).
URL https://doi.org/10.1038/s41467-021-23576-0

I will present two design cases. The first case concern non-enzymatic isothermal strand displacement and amplification for rapid detection of SARS-CoV-2, which we accomplished through design of DNA probes that opens and binds to targeted locations of the SARS-CoV-2 genome. Through RNA folding considerations, we show why one of two probes are more successful and makes the detection possible. In the second case, design of CRISPR/Cas9 guide RNA (gRNA) are made from first generating cleavage efficiency data and subsequently train a deep learning-based neural network which has cutting-edge performance tested on independent data sets.

5.12 What can geometric combinatorics say about RNA design?

Christine Heitsch (Georgia Institute of Technology - Atlanta, US)

Joint work of Christine Heitsch, Svetlana Poznanović, et al.

Branching is a critical characteristic of RNA design, yet can be challenging to validate with thermodynamic optimization approaches. Using mathematical methods (convex polytopes and their normal fans), we can improve prediction accuracy on well-defined families while also illuminating why the general problem is so difficult.

5.13 Experiments in Deep Learning for RNA Secondary Structure Prediction

Ivo Hofacker (Universität Wien, AT)

Joint work of Christoph Flamm, Julia Wielach, Michael T. Wolfinger, Stefan Badelt, Ronny Lorenz, Ivo L. Hofacker
Main reference Christoph Flamm, Julia Wielach, Michael T. Wolfinger, Stefan Badelt, Ronny Lorenz, Ivo L. Hofacker: "Caveats to Deep Learning Approaches to RNA Secondary Structure Prediction", Frontiers in Bioinformatics, Vol. 2, 2022.
URL https://doi.org/10.3389/fbinf.2022.835422

Machine learning (ML) and in particular deep learning techniques have the potential to overcome shortcomings of current RNA secondary structure prediction methods, such as the inability to predict pseudoknots and poor treatment of non-canonical pairs. Several recent publications have proposed deep neural networks for RNA secondary structure prediction and reported excellent accuracies. However, these works build upon training sets that are derived from a relatively small number of RNA families and therefore do not properly represent the RNA structure space.

By folding random sequences using the RNAfold program of the ViennaRNA package, we can generate synthetic data sets that allow to test in detail which properties of the RNA folding map are easy or hard to learn for these networks. We find that structure features that are local in the base pairing matrix, such as stacks and interior loops, are easy to learn, while less local multi-loops are much harder. Most strikingly, the number of base pairs predicted by convolutional networks grows quadratically, rather than linearly, with sequence length.

Using inverse folding, we designed a further synthetic training set that contains the same structures as the widely used bpRNA data set, and therefore exhibits the same lack of structure diversity in spite of near perfect randomness of the sequences. Networks trained on this data set achieve excellent performance on sequence that have no similarity to training sequences but fold into structures well represented in the training set. Nevertheless, the networks perform poorly on sequence folding into novel structures. This suggests, that the excellent performance reported in the literature is largely due biases in the data sets, i.e. that training and test sets that exhibit the same overrepresentation of a few well studied RNA families and their structures.

5.14 BarMap-QA – Cotranscriptional folding with quality assurance

Felix Kühnl (Universität Leipzig, DE)

URL https://www.bioinf.uni-leipzig.de/Software/BarMap_QA/

The structure of an RNA molecule is often a crucial characteristic to be able to explain its biological function. While studying the thermodynmaically optimal (MFE) structure often yields important information, there are relevant cases where a computation of the MFE structure alone is not sufficient to understand a molecule's behaviour, for example in transcriptional riboswitches. Cotranscriptional folding simulations can thus be a helpful tool to gain a deeper understanding a given RNA.

In this talk, I present the software pipeline BarMap-QA, which relies on the BarMap framework by Hofacker et al., to simulate cotranscriptional folding of RNAs. The pipeline is not only very streamlined and easy to set up and run, but it also provides several quality measures to assess the quality of the conducted analysis and thus allow the user to optimally ballance computational efficiency against simulation accuracy for a specific use case.

References

1 Kühnl F et al., accepted, `https://doi.org/10.1101/2020.01.06.895607`

5.15 Discovering RNA Self-Reproducers By In Silico And In Vitro Screening

Philippe Nghe (ESPCI - Paris, FR)

Joint work of Camille Lambert, Vaitea Opuu, Francesco Calvanese, Martin Weigt, Matteo Smerlak, Philippe Nghe

The RNA world hypothesis proposes that RNAs carry catalytic activity necessary for primordial evolution. A first necessary condition for evolution is reproduction. Whether self-reproduction is rare or common in the space of RNA sequences is central to assess the plausibility of this scenario. To date, two ribozymes have been shown to autocatalytically sustain their self-reproduction in the laboratory, starting from RNA oligomers: the Azoarcus ribozyme derived from the group I intron family (Hayden and Lehman 2006) and a fragmented ligase (Lincoln and Joyce 2009). In this project, we assess the probability of self-reproducing RNAs in sequence space by using as a starting point the Azoarcus ribozyme that can autocatalytically self-reproduce. We show that combining *in silico* and *in vitro* screening allows for the discovery of a large number of artificial self-reproducing ribozymes. For this, the strategy consists of: i) Identifying natural self-reproducing GIIs; ii) Applying physics-based and machine learning methods to generate artificial candidates for self-reproduction; iii) Testing designed sequences for self-reproduction using high-throughput sequencing; v) characterizing the representative self-reproducers. We find that generative models that combine statistical signatures from pair correlations and secondary structure prediction are efficient at producing functional ribozymes more than 60 nucleotides away from the original sequence, whereas random mutations destroy activity after only a few.

These methods interpolate the natural diversity found in group I introns, from which self-reproducers can be successfully re-engineered. This overall shows that self-reproduction is not an exceptional property of a few laboratory-made RNAs, but is relatively widespread in the sequence space.

5.16 Physical modeling of RNA polymorphism

Samuela Pasquali (University Paris-Diderot, FR)

Joint work of Samuela Pasquali, Konstantin Roeder

RNA molecules are characterized by the existence of a multitude of stable states that that result in a frustrated energy landscape, where the observed structures depend sensibly on experimental conditions and can depend on the initial, unfolded, structure. Using both atomistic and coarse-grained physical models for RNAs, combined with enhanced sampling methods, we investigate the energy landscape of these systems to understand what are the most relevant structures in the different conditions. Using a few significant examples we show how the combination of these methods allowed us to rationalize the experimental evidence showing the concurrent existence of multiple states [1, 2]. The coarse-grained model we develop [3] is also a useful starting point to couple simulations with experimental data, moving toward intergrative modeling. We have recently developed a simulation technique allowing to bias MD coarse-grained simulations with SAXS data on-the-fly [4], and a theoretical framework to perform fast constant pH simulations where we can model the system considering the exchange of charges with the solvent [5]. These developments allow us to account for the environment to obtain reasonable structures to then be studied more thoroughly with high-resolution modeling.

References

1 K Röder, G Stirnemann, AC Dock-Bregeon, DJ Wales, S Pasquali, Structural transitions in the RNA 7SK 5' hairpin and their effect on HEXIM binding, Nucleic Acids Research 48 (1), 373-389 (2020)

2 K Röder, AM Barker, A Whitehouse, S Pasquali, Investigating the structural changes due to adenosine methylation of the Kaposi's sarcoma-associated herpes virus ORF50 transcript, PLOS Computational Biology 2022, 18(5):e1010150, doi: 10.1371/journal .pcbi.1010150, PMID: 35617364

3 T. Cragnolini, Y. Laurin, P. Derreumaux, S. Pasquali, The coarse-grained HiRE-RNA model for de novo calculations of RNA free energy surfaces, folding, pathways and complex structure predictions, J. Chem. Theory Comput., 11, 3510 (2015)

4 L Mazzanti, L Alferkh, E Frezza, S Pasquali, Biasing RNA coarse-grained folding simulations with Small–Angle X–ray Scattering (SAXS) data, BioXiv doi: 10.1101/2021.03.29.437449 (2021)

5 S. Pasquali, E. Frezza, F.L. Barroso da Silva, Coarse-grained dynamic RNA titration simulations, Interface Focus 9: 20180066 (2019)

5.17 RNA dynamics: one basepair at a time

Katja Petzold (Karolinska Institute - Stockholm, SE)

Joint work of Petzold, Katja and the entire PetzoldLab (alumni and current)

Many functions of RNA depend on rearrangements in secondary structure that are triggered by external factors, such as protein or small molecule binding. These transitions can feature on one hand localized structural changes in base-pairs or can be presented by a change in chemical identity of e.g. a nucleo-base tautomer [1]. We use and develop R1ρ relaxation-dispersion NMR methods [2] for characterizing transient structures of RNA that exist in low abundance (populations <10%) and that are sampled on timescales spanning three orders of magnitude (µs to s).

The characterization of transient structures in microRNA miR-34a targeting the mRNA of Sirt1 [3] will be discussed and a first glimpse into ribosomal dynamics will be provided. We have trapped these short-lived states and characterized their structure and impact on function.

References

1. I.J. Kimsey, K. Petzold, B. Sathyamoorthy, Z.W. Stein and H.M. Al-Hashimi Nature, 519 (7543), pp 315-320, 2015
2. J. Schlagnitweit, E. Steiner, H. Karlsson and K. Petzold, Chemistry – A European Journal; 24(23):6067-6070, 2018. M. Marušić, J. Schlagnitweit and K. Petzold, ChemBioChem 2019, 20, 2685-2710
3. L. Baronti, I. Guzzetti†, P. Ebrahimi†, S. Friebe Sandoz†, E. Steiner†, J. Schlagnitweit, B. Fromm, L. Silva, C. Fontana, A. Chen and K. Petzold, Nature 2020, 583, 139-144

5.18 Minimalistic RNA inverse folding

Yann Ponty (Ecole Polytechnique - Palaiseau, FR)

Joint work of Yann Ponty, Jozef Hales, Alice Héliou, Jan Manuch, Ladislav Stacho, Sebastian Will, Stefan Hammer
Main reference Jozef Hales, Alice Héliou, Ján Manuch, Yann Ponty, Ladislav Stacho: "Combinatorial RNA Design: Designability and Structure-Approximating Algorithm in Watson-Crick and Nussinov-Jacobson Energy Models", Algorithmica, Vol. 79(3), pp. 835–856, 2017.
URL https://doi.org/10.1007/s00453-016-0196-x

We consider two minimalistic instances of RNA Design, restricting our attention to a minimal instance of RNA inverse folding based on a simplified energy model, where any canonical base pair contributes equally to the free-energy. This greatly simplifies the study of algorithmic questions, hopefully enabling new (exact? efficient?) solutions to design problems that are usually approached in a heuristic fashion.

First, we consider the problem of counting/sampling sequences that are simultaneously compatible with a collection of secondary (2D) structures. Valid sequence assignments turn out to be in bijection, up to trivial symmetry, to independent sets of a compatibility graph, built as the union of base pairs from all structures. As all graphs can be obtained as unions of 2D structures, this implies #P-hardness of the counting problem. Yet, the problem can be solved using an DP algorithm that is fixed parameter tractable for the tree-width of the graph. The associated algorithm can be further generalized to compute the partition function for generic constraints, and represents the engine of our declarative framework InfraRed for sequence sampling.

Next, we consider the inverse folding problem which starts from a single target 2D structure, and consists in finding a sequence that folds uniquely into the target with respect to base pair maximization. We first provide a complete characterization for designable structures without unpaired bases. More generally, we characterize extensive classes of (non-)designable structures, and prove the closure of the set of designable structures under the stutter operation. Finally, we consider a structure-approximating relaxation of the design, given a structure S (avoiding 2 basic undesignable motifs) transforms S into a designable structure by adding at most one base-pair to each helix. For all designable structures, a sequence can be generated in linear time, suggesting this relaxed version of design may be easier that the rigid version of the problem.

5.19 Challenges in designing RNA non-canonical modules

Vladimir Reinharz (University of Montreal, CA)

RNA structures depends largely on the geometry of interactions between its nucleotides. While the classical canonical/wobble interactions drive the folding of the major helices, there is a wide variety of different shapes that can connect through hydrogen bonds any nucleotide to any other. They have been classified by Leontis-Westhof into 12 non-canonical families. Graph algorithms have allowed to automatically retrieve all conserved network of non-canonical interactions in all known RNA structures. This work has exhibited the modularity and composability of theses structures. Nonetheless, most of them don't have any associated thermodynamic parameter and it is still unknown if their folding is opportunistic or actually pushed for by these interactions. We ask as questions: What would be a rational scheme to design novel sequences folding in these shapes? How much of the context must be taken into account to ensure the correct folding? And how can chemical modifications enable unique modules?

5.20 Learning to Design RNA

Frederic Runge (Universität Freiburg, DE)

Machine learning (ML) and especially deep learning (DL) approaches recently achieved remarkable results in different domains of life sciences. While such methods have entered many areas of molecular research, the field of RNA design still largely lacks deep learning-based approaches. To close this gap, we present two machine learning based approaches to tackle two different problems related to the field of RNA design. We present an automated deep reinforcement learning (AutoRL) approach that is capable of generating RNA sequences that fold into a desired secondary structure (inverse RNA folding) while often requiring only very few shots to yield a solution. Due to the sensitivity of deep RL algorithms to their hyperparameter settings and the lack of similar work in the field, we use a meta-optimization approach to automatically find the best RL setting for solving the problem. Since inverse RNA Folding is fundamentally linked to RNA folding, we present a probabilistic Transformer for the secondary structure prediction problem. We show that our method outperforms previous work on a commonly used benchmark dataset from the literature and that it improves the quality of non-canonical base pair and pseudoknot predictions compared to previous work. Besides the advantages of a global reception due to self-attention compared to convolution neural networks, the probabilistic nature of our method allows to reconstruct structure ensembles learned from data.

5.21 Differential SHAPE probing to screen computationally designed RNA and to detect pseudoknot and non-canonical interactions

Bruno Sargueil (Paris Descartes University, FR)

Joint work of Bruno Sargueil, Pierre Hardouin, Francois-Xavier Lyonnet, Elisa Frezza, Benoit Masquida, Yann Ponty, Sebastian Will, Simona Cocco, Rémy Monasson, Jorge Cossio, Andrea di Giocchino
Main reference Grégoire De Bisschop, Delphine Allouche, Elisa Frezza, Benoît Masquida, Yann Ponty, Sebastian Will, Bruno Sargueil: "Progress toward SHAPE Constrained Computational Prediction of Tertiary Interactions in RNA Structure", Non-Coding RNA, Vol. 7(4), 2021.
URL https://doi.org/10.3390/ncrna7040071

The development of reliable RNA design processes requires experimental validation. RNA structure modelling from chemical probing experiments has made tremendous progress, however accurately predicting large RNA structures is still challenging for several reasons. In particular interactions such as pseudoknots and non-canonical base pairs which are not captured by the available incomplete thermodynamic model are hardly predicted efficiently. To identify nucleotides involved in pseudoknots and non-canonical interactions, we scrutinized the SHAPE reactivity of each nucleotide of a benchmark RNA under multiple conditions. We show that probing at increasing temperature was remarkably efficient at pointing to non-canonical interactions and pseudoknot pairings. The SHAPE probing technology was then use to screen for RNA computationally designed to interact with a small molecule

5.22 ENSnano: a 3D modeling software for designing complex DNA/RNA nanostructures

Nicolas Schabanel (ENS - Lyon, FR)

Joint work of Nicolas Levy, Allan Mills, Julie Finkel, Gaëtan Bellot, Nicolas Schabanel
URL http://www.ens-lyon.fr/ensnano/

Since the 1990s, increasingly complex nanostructures have been reliably obtained out of self-assembled DNA strands: from "simple" 2D shapes to 3D gears and articulated nano-objects, and even computing structures. The success of the assembly of these structures relies on a fine tuning of their structure to match the peculiar geometry of DNA helices. Various softwares have been developed to help the designer. These softwares provides essentially four kind of tools: an abstract representation of DNA helices (e.g. cadnano, scadnano, DNApen, 3DNA, Hex-tiles); a 3D view of the design (e.g., vHelix, Adenita, oxDNAviewer); fully automated design (e.g., BScOR, Daedalus, Perdix, Talos, Athena), generally dedicated to a specific kind of design, such as wireframe origamis; and coarse grain or thermodynamical physics simulations (e.g., oxDNA, MrDNA, SNUPI, Nupack, ViennaRNA,...). MagicDNA combines some of these approaches to ease the design of configurable DNA origamis.

We present our first step in the direction of conciliating all these different approaches and purposes into one single reliable GUI solution: the first fully usable version (design from scratch to export) of our general purpose 3D DNA nanostructure design software ENSnano. We believe that its intuitive, swift and yet powerful graphical interface, combining 2D and 3D editable views, allows fast and precise editing of DNA nanostructures. It also handles editing of large 2D/3D structures smoothly, and imports from the most common solutions. Our software extends the concept of grids introduced in cadnano; grids allows to abstract and articulated the different parts of a design. ENSnano also provides new design tools which speeds up considerably the design of complex large 3D structures, most notably: a 2D split view, which allows to edit intricate 3D structures which cannot easily be mapped in a 2D view, and a copy & repeat functionality, which takes advantage of the grids to design swiftly large repetitive chunks of a structure. ENSnano has been validated experimentally, as proven by the AFM images of a DNA origami entirely designed in ENSnano.

ENSnano is a light-weight ready-to-run independent single-file app, running seamlessly in most of the operating systems (Windows 10, MacOS 10.13+ and Linux), it thus does not require the installation of any other softwares such as Matlab, Maya or Samson. Precompiled versions for Windows and MacOS are ready to download on ENSnano website. In the coming months, we will add new features to our software to extend its capacities in the various directions discussed in this article. We decided to release now this first version of our software as its 3D and 2D editing interface is meeting our usability goals. Because of its stability and ease of use, we believe that ENSnano should find already its place in anyone's design chain, when precise editing of a larger nanostructure is needed.

Furthermore, we propose a new method for designing curved origamis that deviates radically from the pattern-based previous approaches. We have developed a new model for DNA double helices curved in 3D that allows us to directly position the DNA double helices constituting the desired shape in the 3D interface of our software ENSnano. The crossovers positions are then simply deduced from the 3D positions of the nucleotides, as predicted by our model. This geometry-based interactive approach shortcuts the tedious process of manually coming up with a pattern suited for the desired curvature, and furthermore allows to deal transparently with structure whose curvature varies continuously. We also propose an innovative 2D representation synchronizing curved parallel double helices without relying on insertions or deletions, by automatically adapting the cell width for each nucleotide in the array representation.

We provide experimental data validating our curvy DNA model by successfully annealing two DNA origamis conceived thanks to two new DNA design methods. The first origami consists in a 6-helices bundle following an interactively created bezier curve whose curvature gets as low as 4.7nm. The second is an asymetrical Möbius torus whose DNA strands are routed along 2 spiraling helices covering its whole surface. This new spiraling technique, allowed by our DNA curvy model, enables to grasp xovers within a continuous range which results in an easier-to-design and smoother surface. Both of our designs folded as is, without any need to redesign their xover schemes.

5.23 Single-Stranded Architectures for RNA Co-Transcriptional Folding

Shinnosuke Seki (The University of Electro-Communications - Tokyo, JP)

Joint work of Daria Pchelina, Nicolas Schabanel, Shinnosuke Seki, Guillaume Theyssier
Main reference Daria Pchelina, Nicolas Schabanel, Shinnosuke Seki, Guillaume Theyssier: "Oritatami Systems Assemble Shapes No Less Complex Than Tile Assembly Model (ATAM)", in Proc. of the 39th International Symposium on Theoretical Aspects of Computer Science, STACS 2022, March 15-18, 2022, Marseille, France (Virtual Conference), LIPIcs, Vol. 219, pp. 51:1–51:23, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.
URL https://doi.org/10.4230/LIPIcs.STACS.2022.51

Oritatami (folding in Japanese) is a mathematical model of computation by co-transcriptional folding we proposed in 2016 and have been studying, primarily on its computational power. In this model, RNA co-transcriptional folding is generalized so that the bases (called "beads" herein) can be of arbitrarily defined, finitely-many types that may have arbitrary affinities with each other (rather than just the four bases in RNA with their fixed set of affinities), but restricted on the 2D plane. In this talk, we present the latest universal oritatami architecture that enables us to compute all computable functions (Turing universality) co-transcriptionally, with particular emphasis on simplicity of mechanisms it employs to read/write a bit, to store information, and to merge computational paths (erasure).

5.24 Coarse-grained modeling for RNA nanotechnology

Petr Sulc (Arizona State University - Tempe, US)

Nucleic acid nanotechnology uses designed DNA or RNA strands that self-assemble into larger complexes and nanodevices. Computer modeling and simulations can provide crucial insights into function and design of such nanostructures. However, the sizes (up to thousands of base pairs) and timescales of their assembly (minutes to hours) of such nanodevices presents major challenge for modeling approaches. Here, we will present a coarse-grained model, oxDNA/oxRNA, specifically designed to simulate DNA and RNA nanotechnology, and we will demonstrate its application to RNA strand displacement reaction, a key mechanism in active nanotechnology devices which has recently been also identified to occur during RNA folding *in vivo*. We will then discuss applications of our modeling platform for inverse design of multicomponent nanostructure assemblies: how to design individual nucleic acid building blocks that self-assemble reliable into target mutlicomponent structure while avoiding kinetic traps and alternative free-energy minima? We show that through combination of multiscale modeling and mapping of the inverse design problem to Boolean Satisfiability Problem (SAT), it is possible to design nanostructures that assemble large-scale 3D assemblies, opening ways to use nucleic acids to biotemplated manufacturing.

5.25 Persuading tRNA to jump over stop codons

Andrew Torda (Universität Hamburg, DE)

Joint work of Andrew E. Torda, Marco C. Matthies

Can one persuade a half-artificial tRNA to bind to a stop codon, pretend to be a tRNA-Ala and incorporate an alanine residue in a growing protein ? If so, you might be on the way to alleviating a disease caused by an unwanted stop codon.

If you want to design an RNA sequence, you want a series of real nucleotides at the end of the day, but you may well go through some non-physical mixed states along the way. You can represent a base as some fraction of A plus C plus.. If you have an energy model, you can take the derivative of energy with respect to the composition at each site. This lets you use gradient-based methods to optimise your sequence.

We used the program DSS-Opt to find our artificial tRNA sequences, although this was no longer a de novo problem. The tRNA does not just have to fold correctly. It has to be able to convince an amino-acyl synthetase to charge it and then sneak past a host of recognition factors before a ribosome would consider taking it seriously. This means our calculations were far from de novo design. Only about 45% of the bases were actually optimised.

About half a dozen candidates were tested for charging by an alanine amino acyl-tRNA synthetase and then for stop-codon read-through with a luciferase assay. The winner of this was fed to an antibiotic-stalled ribosome and the structure solved by cryo-EM (acquisition code 7B5K). A bouncing baby half-designed tRNA smiled at the authors from the coordinates.

You could either view this as a triumph of design or you could say, less than half the sites in the molecule were actually chosen.

5.26 Kinetic features of RNA-RNA interactions

Maria Waldl (Universität Wien, AT)

Interactions between RNAs are an essential mechanism in gene regulation. State-of-the-art computational genome-wide screens predict targets of regulatory RNAs based on thermodynamic stability but largely neglect kinetic effects. To overcome this limitation, we propose novel models of RNA-RNA interaction dynamics. On this basis we can improve our understanding of general principles that govern RNA-RNA interaction formation and improve target prediction tools.

While the dynamics of secondary structure formation of single RNAs have been successfully modeled using transition systems between conformations, analogous approaches for RNA-RNA interaction quickly lead to infeasibly large systems. Therefore, we propose reducing the interaction system to the direct trajectories (shortest paths) from possible first contacts to full hybridization. This key idea enables studying general principles and relevant features of the interaction formation as well as model details; e.g. the relative speed of intra- and intermolecular folding. Specifically, we isolate kinetic effects by comparing experimentally confirmed interactions from Salmonella and E. coli to a randomized background with similar thermodynamic properties.

These experiments indicate that native interactions are kinetically favored. Moreover, folding trajectories often look remarkably different depending on the site of the initial contact. Based on a machine learning classifier, we were able to identify a combination of interaction features that provide most information on the behavior of native RNA-RNA interactions. These features can be exploited to filter target predictions.

Due to the design of our RNA kinetics model, features like energy barriers can be computed efficiently. This enables refining genome-wide target predictions through kinetic criteria. Beyond these immediate practical improvements, we shed light on general principles like the long-debated influence of the accessibility of the initial contact site.

In the context of this seminar I would like to present this direct path models for interaction formation as well as the kinetic features that we identified and discuss how such features could extend current RNA design strategies.

5.27 Infrared: A sampling framework for RNA design... and beyond

Sebastian Will (Ecole Polytechnique - Palaiseau, FR)

Joint work of Sebastian Will, Yann Ponty, Hua-Ting Yao
URL http://www.lix.polytechnique.fr/ will/Software/Infrared/

Infrared is a modeling framework for efficient targeted sampling and optimization. It was originally developed for implementing complex sequence design approaches with multiple objectives and side constraints, e.g. design of sequences with multiple RNA target structures while controlling the GC-content (RNARedPrint). Due to its declarative, compositional application programming/modeling interface, Infrared allows extending existing design tools to solve very specific design tasks, e.g. optimizing codon-usage while targeting RNA structures and (possibly) additional constraints. In the same way, it enables rapid development of completely new design tools like RNAPOND (and, due to its generality, even methods beyond design, e.g. alignment of RNAs with pseudoknots). A main feature of the system is its automatic adaptation to the complexity of the declaratively modeled task. For this purpose, the system implicitly derives fixed- parameter-tractable sampling and optimization algorithms using tree-decomposition. The talk outlines main properties and background of the system, its elementary usage, and presents concrete examples of design applications.

5.28 Forbidden RNA motifs and the cardinality of secondary structure space

Hua-Ting Yao (Universität Wien, AT)

Joint work of Hua-Ting Yao, Cédric Chauve, Mireille Regnier, Yann Ponty
Main reference Hua-Ting Yao, Cédric Chauve, Mireille Regnier, Yann Ponty: "Exponentially Few RNA Structures Are Designable", in Proc. of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, BCB '19, p. 289–298, Association for Computing Machinery, 2019.
URL https://doi.org/10.1145/3307339.3342163

The problem of RNA design attempts to construct RNA sequences that perform a predefined biological function, identified by several additional constraints. One of the foremost objectives of RNA negative design is that the designed RNA sequence should adopt a predefined target secondary structure preferentially to any alternative structure, according to a given metrics and folding model. It was observed in several works that some secondary structures are undesignable, i.e. no RNA sequence can fold into the target structure while satisfying some criterion measuring how preferential this folding is compared to alternative conformations.

We show that the proportion of designable secondary structures decreases exponentially with the size of the target secondary structure, for various popular combinations of energy models and design objectives. This exponential decay is, at least in part, due to the existence of forbidden motifs, which can be generically constructed, and jointly analyzed to yield asymptotic upper bounds on the number of designable structures. Moreover, we define a lower bound of the structural ensemble defect. We show that, across uniformly distributed secondary structures, such a lower bound has a Normal limiting distribution with the expected value and the variance both linear to the size of the secondary structure.

Participants

- Ebbe Sloth Andersen
Aarhus University, DK
- Maciej Antczak
Poznan University of Technology, PL
- Stefan Badelt
Universität Wien, AT
- Danny Barash
Ben Gurion University – Beer Sheva, IL
- Sarah Berkemer
Ecole Polytechnique – Palaiseau, FR
- Anne Condon
University of British Columbia – Vancouver, CA
- Harold Fellermann
Newcastle University, GB
- Jorge Fernández de Cossío Díaz
ENS – Paris, FR
- Sven Findeiß
Universität Leipzig, DE
- Christoph Flamm
Universität Wien, AT
- Cody Geary
Aarhus University, DK
- Jan Gorodkin
University of Copenhagen, DK
- Christine Heitsch
Georgia Institute of Technology – Atlanta, US
- Ivo Hofacker
Universität Wien, AT
- Felix Kühnl
Universität Leipzig, DE
- István Miklós
ELKH – Budapest, HU
- Philippe Nghe
ESPCI – Paris, FR
- Cyrille Merleau Nono Saha
MPI für Mathematik in den Naturwissen. – Leipzig, DE
- Samuela Pasquali
University Paris-Diderot, FR
- Katja Petzold
Karolinska Institute – Stockholm, SE
- Yann Ponty
Ecole Polytechnique – Palaiseau, FR
- Vladimir Reinharz
University of Montreal, CA
- Lorenz Ronny
Universität Wien, AT
- Frederic Runge
Universität Freiburg, DE
- Bruno Sargueil
Paris Descartes University, FR
- Nicolas Schabanel
ENS – Lyon, FR
- Shinnosuke Seki
The University of Electro-Communications – Tokyo, JP
- Petr Sulc
Arizona State University – Tempe, US
- Marta Szachniuk
Poznan University of Technology, PL
- Andrew Torda
Universität Hamburg, DE
- Maria Waldl
Universität Wien, AT
- Sebastian Will
Ecole Polytechnique – Palaiseau, FR
- Hua-Ting Yao
Universität Wien, AT

Report from Dagstuhl Seminar 22382

Machine Learning for Science: Bridging Data-Driven and Mechanistic Modelling

Philipp Berens*[1], Kyle Cranmer*[2], Neil D. Lawrence*[3], Ulrike von Luxburg*[4], and Jessica Montgomery*[5]

1 **Universität Tübingen, DE.** philipp.berens@uni-tuebingen.de
2 **University of Wisconsin – Madison, US.** kyle.cranmer@nyu.edu
3 **University of Cambridge, GB.** ndl21@cam.ac.uk
4 **Universität Tübingen, DE.** luxburg@informatik.uni-tuebingen.de
5 **University of Cambridge, GB.** jkm40@cam.ac.uk

Abstract

This report documents the programme and the outcomes of Dagstuhl Seminar 22382 "Machine Learning for Science: Bridging Data-Driven and Mechanistic Modelling".

Today's scientific challenges are characterised by complexity. Interconnected natural, technological, and human systems are influenced by forces acting across time- and spatial-scales, resulting in complex interactions and emergent behaviours. Understanding these phenomena – and leveraging scientific advances to deliver innovative solutions to improve society's health, wealth, and well-being – requires new ways of analysing complex systems.

The transformative potential of AI stems from its widespread applicability across disciplines, and will only be achieved through integration across research domains. AI for science is a rendezvous point. It brings together expertise from AI and application domains; combines modelling knowledge with engineering know-how; and relies on collaboration across disciplines and between humans and machines. Alongside technical advances, the next wave of progress in the field will come from building a community of machine learning researchers, domain experts, citizen scientists, and engineers working together to design and deploy effective AI tools.

This report summarises the discussions from the seminar and provides a roadmap to suggest how different communities can collaborate to deliver a new wave of progress in AI and its application for scientific discovery.

Seminar September 18–23, 2022 – http://www.dagstuhl.de/22382

2012 ACM Subject Classification Computing methodologies → Machine learning; Computing methodologies → Artificial intelligence

Keywords and phrases machine learning, artificial intelligence, life sciences, physical sciences, environmental sciences, simulation, causality, modelling

Digital Object Identifier 10.4230/DagRep.12.9.150

* Editor / Organizer

Machine Learning for Science: Bridging Data-Driven and Mechanistic Modelling, *Dagstuhl Reports*, Vol. 12, Issue 9, pp. 150–199
Editors: Philipp Berens, Kyle Cranmer, Neil D. Lawrence, Ulrike von Luxburg, and Jessica Montgomery
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Philipp Berens
Kyle Cranmer
Neil D. Lawrence
Ulrike von Luxburg
Jessica Montgomery

Today's scientific challenges are characterised by complexity. Interconnected natural, technological, and human systems are influenced by forces acting across time- and spatial-scales, resulting in complex interactions and emergent behaviours. Understanding these phenomena – and leveraging scientific advances to deliver innovative solutions to improve society's health, wealth, and well-being – requires new ways of analysing complex systems.

Artificial intelligence (AI) offers a set of tools to help make sense of this complexity. In an environment where more data is available from more sources than ever before – and at scales from the atomic to the astronomical – the analytical tools provided by recent advances in AI could play an important role in unlocking a new wave of research and innovation. The term AI today describes a collection of tools and methods, which replicate aspects of intelligence in computer systems. Many recent advances in the field stem from progress in machine learning, an approach to AI in which computer systems learn how to perform a task, based on data.

Signals of the potential for AI in science can already be seen in many domains. AI has been deployed in climate science to investigate how Earth's systems are responding to climate change; in agricultural science to monitor animal health; in development studies, to support communities to manage local resources more effectively; in astrophysics to understand the properties of black holes, dark matter, and exoplanets; and in developmental biology to map pathways of cellular development from genes to organs. These successes illustrate the wider advances that AI could enable in science. In so doing, these applications also offer insights into the science of AI, suggesting pathways to understand the nature of intelligence and the learning strategies that can deliver intelligent behaviour in computer systems.

Further progress will require a new generation of AI models. AI for science calls for modelling approaches that can: facilitate sophisticated simulations of natural, physical, or social systems, enabling researchers to use data to interrogate the forces that shape such systems; untangle complicated cause-effect relationships by combining the ability to learn from data with structured knowledge of the world; and work adaptively with domain experts, assisting them in the lab and connecting data-derived insights to pre-existing domain knowledge. Creating these models will disrupt traditional divides between disciplines and between data-driven and mechanistic modelling.

The roadmap presented here suggests how these different communities can collaborate to deliver a new wave of progress in AI and its application for scientific discovery. By coalescing around the shared challenges for AI in science, the research community can accelerate technical progress, while deploying tools that tackle real-world challenges. By creating user-friendly toolkits, and implementing best practices in software and data engineering, researchers can support wider adoption of effective AI methods. By investing in people working at the interface of AI and science – through skills-building, convening, and support for interdisciplinary collaborations – research institutions can encourage talented researchers to develop and adopt new AI for science methods. By contributing to a community of research and practice, individual researchers and institutions can help share insights and

expand the pool of researchers working at the interface of AI and science. Together, these actions can drive a paradigm shift in science, enabling progress in AI and unlocking a new wave of AI-enabled innovations.

The transformative potential of AI stems from its widespread applicability across disciplines, and will only be achieved through integration across research domains. AI for science is a rendezvous point. It brings together expertise from AI and application domains; combines modelling knowledge with engineering know-how; and relies on collaboration across disciplines and between humans and machines. Alongside technical advances, the next wave of progress in the field will come from building a community of machine learning researchers, domain experts, citizen scientists, and engineers working together to design and deploy effective AI tools.

2 Table of Contents

3 Introduction: bridging data driven and mechanistic modelling

The 21st century has been characterised as the century of complexity.[1] Shifting social, economic, environmental, and technological forces have created increasingly interconnected communities, affected by "wicked" problems in domains such as health, climate, and economics [1]. This complexity is reflected in today's scientific agenda: whether in natural, physical, medical, environmental, or social sciences, researchers are often interested in the dynamics of complex systems and the phenomena that emerge from them.

Science has always proceeded through the collection of data. Through their experiments and observations, researchers collect data about the world, use this data to develop models or theories of how the world works, make predictions from those models, then test those predictions, leading to further refinements to the model and the underpinning theory. Digitisation of daily activities–in the lab, and elsewhere – means that researchers today have access to more data from a greater range of sources than ever before. In parallel, more sophisticated tools to collect data have opened new scales of scientific inquiry, from detailed patterns of gene expression to light signals from other galaxies. Data proliferation is both a signal of the complexity of today's environment, and an opportunity to make sense of such complexity.

Advances in artificial intelligence (AI) have produced new analytical tools to make sense of these data sources. The term "AI" today describes a collection of methods and approaches to create computer systems that can perform tasks that would typically be associated with "intelligent" behaviour in living systems.[2] In this document, the term AI is used broadly, to refer to algorithmic decision-making systems that combine data, mathematical models, and compute power to make predictions about the world.

AI is already unlocking progress across research disciplines:

- In Earth sciences, it is helping researchers investigate how different parts of the Earth's biosphere interact, and are affected by climate change.[3]
- In climate science, it supports modelling efforts to reconstruct historical climate patterns, enabling more accurate predictions of future climate variability.[4]
- In agricultural science, it is helping farmers access faster diagnoses of animal diseases, enabling more effective responses.[5]
- In astrophysics, it is advancing understandings of the nature of dark matter and its role in the Universe.[6]
- In developmental biology, it is generating insights into the genetic processes that shape how cells develop and differentiate into specialist roles.[7]
- In environmental science, it allows researchers to analyse the features of natural environments more accurately, aiding land and resource managers.[8]
- In neuroscience, it can help model how different neural circuits fire to deliver different behaviours in animals.[9]

---

[1] This quote is attributed to Stephen Hawking, in an interview with the San Jose Mercury News in January 2000.

[2] While not the only branch of the field, machine learning is the approach to AI that has delivered many of the recent advances in AI. Machine learning is an approach to AI in which models process data, learning from that data to identify patterns or make predictions. In this document, the terms machine learning and AI are used interchangeably.

[3] These examples are inspired by talks given at the Dagstuhl Seminar; these are provided later in the document. This example is inspired by Markus Reichstein's talk.

[4] This example is inspired by Ieva Kazlauskaitė's talk.

[5] This example is inspired by Dina Machuve's talk.

[6] This example is inspired by Siddharth Mishra-Sharma's talk.

[7] This example is inspired by Maren Büttner's talk.

[8] This example is inspired by Christian Igel's talk.

[9] This example is inspired by Jakob Macke's talk.

The diversity of these successes illustrates the transformative potential of AI for research across the natural, physical, social, medical, and computer sciences, arts, humanities, and engineering. By enabling researchers to extract insights from a greater volume of data, drawn from a wider variety of sources, and operating across multiple dimensions and scales, AI could unlock new understandings of the world. In so doing, AI could influence the conduct of science itself. AI-enabled analytical tools mean researchers can now generate sophisticated simulations of natural or physical systems, creating "digital siblings" of real-world systems that can be used for experimentation and analysis. Machine learning models that combine the ability to learn adaptively from data with the ability to make structured predictions reflecting the laws of nature can help researchers untangle the web of cause-effect relationships that drive the dynamics of complex systems. AI-assisted laboratory processes could increase the efficiency of experiments, and support researchers to develop and test new hypotheses.

Achieving this potential will require advances in the science of AI, the design of AI systems that serve scientific goals, and the engineering of such systems to operate safely and effectively in practice. These advances in turn rely on interdisciplinary collaborations that connect domain expertise to the development of machine learning models, and feed the insights generated by such models back into the domain of study. As interest in the potential of AI to drive a new wave of research grows, the challenge for the field is to identify technical and operational strategies to realise this potential. In the process, new questions arise about the future of "AI for science"; whether this will emerge as a distinct field, characterised by its own research agenda and priorities, or whether its benefits can be best achieved through separate, domain-focused sub-fields, which seek to integrate AI into business-as-usual across research disciplines.

In response, this document proposes a roadmap for "AI for science". Synthesising insights from recent attempts to deploy AI for scientific discovery, it proposes a research agenda that can help develop more powerful AI tools and the areas for action that can provide an enabling environment for their deployment. It starts by exploring core research themes – in simulation, causality, and encoding domain knowledge – then draws from these ideas to propose a research agenda and action plan to support further progress. The ideas presented are inspired by discussions at "Machine Learning for Science: Bridging Mechanistic and Data Driven Modelling Approaches", a Dagstuhl Seminar convened in September 2022. Abstracts from the talks given at the seminar are shown throughout this document. These talks and the discussions they provoked should be credited for the ideas that have shaped it. Thank you to the speakers and participants for their thoughtful contributions to both the seminar and the development of this work.

References

1 Horst W. J. Rittel and Melvin M. Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973. Reprinted in N. Cross, ed. Developments in design methodology, pp. 135–44. Chichester: J. Wiley & Sons, 1984.

4 Snapshots of AI in science

Across domains, AI is being deployed to advance the frontiers of science. The snapshots below introduce some current areas of research in AI for science, and explore the issues raised by these research projects. Across these snapshots, some common themes emerge:

- How can researchers most effectively combine observations, data-driven models, and physical models to enhance understanding of complex systems? To answer this question,

methods are needed to integrate different types of model, operating across different levels of granularity, while managing the impact of the uncertainties that emerge when a machine learning model is integrated in a wider system. New approaches to simulation and emulation can support progress in tackling these challenges, alongside new strategies for examining the robustness or performance of machine learning models.
- How do the outputs from an AI system align with what researchers already know about the world, and how can such systems help uncover causal relationships in data? Advances in causal machine learning are needed to connect the laws and principles already established in many areas of research with data-driven methods.
- How can AI be integrated into the scientific process safely and robustly? Effective integration will rely on the ability to encode domain knowledge in AI systems, the design of interfaces that facilitate interaction between humans and AI, and the development of mechanisms for sharing knowledge and know-how about how to use AI in practice.

4.1 In Earth sciences

The Earth is a complex system,[10] comprised of terrestrial, marine, and atmospheric biospheres that interact with each other and are shaped by biological, chemical, and physical processes that exchange energy across scales from the molecular to the planetary. It is also a unique system: researchers have yet to discover other planets that replicate its dynamics. Studies of the Earth system therefore rely on observations and physical models, which describe the dynamics of energy exchange from first principles and use those principles to build models of the Earth's sub-systems. As climate change perturbs this complex system, it is increasingly important to have accurate models that can be used to analyse how the Earth will respond to increasing carbon dioxide levels. The challenge for Earth system science is to build more complex models that represent the web of relationships between biospheres under changing conditions, without generating overwhelming uncertainties and while generating actionable insights that can be used by individuals, organisations, and policymakers to understand the localised impact of changing environmental conditions [1].

For example, how much carbon dioxide is absorbed by different biospheres can be affected by diverse factors including volume and type of vegetation cover, water and drought stress in different areas, and local temperature, which have implications for how carbon dioxide contributes to climate change. Researchers have access to data that describes local uptake of carbon dioxide by some ecosystems, such as tropical rainforest, European beech forest, or Mediterranean savanna, for example, but lack sufficient observational coverage to scale from these local observations to accurate global representations of carbon exchange. One response to this challenge is to leverage data-driven models to knit together the different mechanistic models that describe (for example) carbon, water, and energy cycles in different biospheres.

By starting with observational data and combining this with physics-informed modelling, researchers can leverage machine learning to create simulations that can generate new understandings of how complex systems function. Taking this approach, the FLUXNET project combines observed data on carbon emissions from different sources to generate a data-driven picture of global carbon dynamics. By combining data across scales to establish a statistical model of global carbon dynamics, this project can generate simulations of how the Earth

[10] This example is inspired by Markus Reichstein's talk, the abstract for which is provided later in this document.

breathes [2]. The ability to integrate across scales and combine models of different Earth subsystems can also contribute to wider efforts to build a "digital twin" of the Earth, with the aim of better understanding the implications of climate change across biospheres and communities.

As the Earth's climate changes,[11] researchers anticipate that local environmental conditions will change and extreme weather events will increase. Understanding the impact of these changes is important for those seeking to develop appropriate responses, for example developing environmental management plans or planning human activities.

How a landscape responds to changing environmental conditions will vary depending on the local climate, characteristics of the terrain (vegetation type, for example), and human activities in the area. Under changing climate conditions, as extrapolation beyond known limits becomes necessary, the assumptions or abstractions that form the basis of a model can be rendered invalid. Relying solely on either mechanistic descriptions of the system – the impact of temperature on plant growth, for example[12] – or statistical models could result in inaccuracies. Machine learning can help respond to this challenge, through the creation of hybrid models that combine an understanding of the physical laws with model parameters learned from data. Researchers often already have access to known physical parameters for a system (for example, the equations that govern how water evaporates to air). These parameters can be fed into a machine learning model that will learn other patterns. Known equations specify the chemical and physical processes; machine learning can then help elucidate the other biological forces at play. Integrating this physical structure in the model helps make it both more interpretable to the domain scientists and more reliable in its predictions. The resulting model can accurately forecast the impact of climate change on the features of local landscapes, operating within the bounds set by the laws of physics [4].

Ice loss[13] has been the greatest contributor to sea-level rise in recent decades [5]. Large volumes of fresh water are stored as ice: NASA estimates that if all the world's glaciers and ice sheets melted, sea levels globally would rise by over 60 metres, flooding all coastal cities [6]. Researchers can estimate the contribution that melting ice makes to sea level rise through mechanistic models that describe the underlying physical processes (that turn ice to water) and through observational data about the velocity of ice sheet movement. Machine learning could offer a toolkit to make these models more accurate, connecting ice sheet models to ocean and atmospheric models, and integrating different data types in hybrid mechanistic-data models.

Efforts to build such models, however, illustrate the complexity of designing tools to meet domain needs. Projects in this space have considered emulating the ice sheet system – or its individual components – to see if models could be run faster; though successful methodologically, it has not been clear that such efforts address a clear research need. Another approach is to use machine learning to streamline simulations, for instance by identifying the most effective level of granularity for different models (is a spatial breakdown of 5km or 10km more interesting?). An important lesson from such collaborations is the specificity of domain needs: machine learning is a tool for research, but just because researchers have a hammer,

[11] This example is inspired by Markus Reichstein's talk, the abstract for which is provided later in this document.
[12] Under conditions of extreme temperature, patterns of stomatal opening and closing in plants changes. See, for example [3].
[13] This example is inspired by Ieva Kazlauskaitė's talk, the abstract for which is provided later in this document.

does not mean every research problem is a nail. Effectively deploying machine learning for research requires both suitable AI toolkits and an understanding of which toolkits are best deployed for which challenges.

4.2 In environmental and agricultural sciences

Poultry farming[14] is a vital source of income and food for many communities in Tanzania. 4.6 million households in the country raise approximately 36 million chickens, but despite the importance of this activity, poultry farming suffers from relatively low productivity due to the prevalence of disease. Efforts to tackle poultry diseases such as Salmonella, Newcastle disease, and coccidiosis are held back by the accessibility of diagnostic processes and lack of data. Diagnosis currently requires lab analysis of droppings, which can take 3-4 days. Once disease is confirmed, farmers often lose their entire farm's flock.

Farm-level tests and diagnostics could increase the effectiveness of disease surveillance and treatment, giving farmers rapid access to information about the diseases affecting their flock and action plans about how to manage outbreaks. With mobile phones ubiquitous across the country – there are almost 49 million mobile phone subscriptions in Tanzania – there are opportunities for new uses of local data to detect disease outbreaks.

By collecting images of droppings from farms, researchers have been creating a dataset to train a machine learning system that can identify the symptoms of these diseases. Fecal images are taken on farms, annotated with diagnostic information from agricultural disease experts and the results of lab tests, then used to train an image recognition system to automate the diagnosis process [7]. System robustness and accuracy is vital, given the significant implications of a positive diagnosis, and careful design is necessary to incentivise farmers to make use of the app.

Collaboration with experts from different domains is central to developing this system. Input from farmers is needed to collect data and test the system in practice; from veterinary pathologists to help annotate the data and ensure the system's accuracy; and from technologists to develop an AI system that is effective in deployment as an app on mobile phones. These collaborations also open opportunities for new forms of citizen science, as farmers and local communities are engaged in efforts to develop and maintain an open toolkit for disease diagnosis, providing a gateway for communities to take ownership of machine learning as a tool to serve their needs.

Trees and forests[15] play a crucial role in maintaining healthy ecosystems. Despite this, an estimated ten million hectares of forest are lost globally each year due to reforestation, with only around half of this balanced by tree-planting efforts [8]. Africa experienced an annual rate of forest loss of approximately 3.9 million hectares per year from 2010-2020. This loss has implications for biodiversity and people, with trees a vital contributor to ecosystem services such as carbon storage, food provision, and shelter. In this shifting landscape, understanding the number and distribution of trees is important for the development of forestry management plans and for understanding the carbon storage implications of changes to land use.

[14] This example is inspired by Dina Machuve's talk, the abstract for which is provided later in this document.
[15] This example is inspired by Christian Igel's talk, the abstract for which is provided later in this document.

To estimate the number and biomass of trees in the West African Sahara and Sahel, researchers have used satellite imagery of 90,000 trees from 400 sampling sites to create a labelled dataset for use in machine learning. Using an image segmentation tool to identify the location of trees, an automated system was able to count the number of trees, with domain experts guiding the system to distinguish trees from surrounding vegetation. This tree count can then be used to estimate the biomass of trees in the area, and predict the amount of carbon they store; the prediction is generated using allometric calculations, which translate the properties of the tree to its carbon storage potential. In this approach, machine learning measures the properties of the ecosystem from satellite images, then these properties are used to feed mechanistic models that describe the ecosystem's physical functions [9]. This opens the possibility of new tools to estimate tree cover, leveraging these insights for more effective environmental management. However, in the process, care is needed to manage the type and nature of the uncertainties created by different modelling approaches. Different allometric models, for example, can be more or less suited to different types of tree cover [10], meaning that the method for estimating biomass from satellite imagery can be subject to biases when applied across a large area. A small error in the calculation of the biomass from one tree can have a cumulatively large effect when that method is scaled to country-level. The type and nature of such uncertainties need to be considered when a machine learning model is used within a wider system.

Vector borne diseases[16] account for more than 17% of diseases in people and over 700,000 deaths annually [11]. Changes to the climate and patterns of land use, amongst other factors, are bringing human populations into contact with new vectors of disease. In Africa, for example, populations of mosquitoes carrying malaria that might previously have been found mainly in rural areas are spreading into cities.

Tools to characterise building features from satellite imagery have already been developed and made available for use.[17] Leveraging these to analyse multi-scale data – from household to city-level–researchers are investigating how the built environment influences people's risk of contracting mosquito-borne disease. For example, it has been found that the prevalence of mosquitos in an area is related to the type of roofing used in construction; metal roofing tends to be associated with lower mosquito prevalence, potentially due to the high temperatures they attract during the day [13]. These insights can be deployed by policymakers in the development of appropriate policy responses [14].

Decisions made on the basis of insights generated by machine learning models will be influenced by the assumptions made in those models. In the context of housing, for example, the decision about which type of housing to identify as "at risk" or which building materials to flag as "problematic" may have significant consequences for individuals or communities. When those decisions are assimilated within a model or analysis before a downstream "policy decision", the implications for those communities of different courses of action may be obscured, creating a risk of marginalising or disadvantaging individuals or groups. The assumptions are built into the model, and how visible those assumptions are made to different user groups, can have significant social and scientific consequences.

[16] This example is inspired by Christian Igel's talk, the abstract for which is provided later in this document.
[17] For example: [12].

4.3 In physical sciences

Understanding the nature of dark matter[18] is one of the biggest unsolved challenges of particle physics today. The matter that researchers can measure using cosmological observations makes up about 5% of the Universe [15]. While not directly observable, evidence for the existence of dark matter can be found in a variety of phenomena not otherwise accounted for by currently known laws of physics: stars rotate around galaxies faster than might be expected; the pattern of fluctuations in primordial microwave observations indicate that there were sources of gravitation in the early Universe beyond ordinary matter; light bends around galaxy clusters due to gravitational effects from dark matter.

Despite knowing that dark matter exists and that it plays an important role in how the Universe formed, its particle composition or properties remains unclear. Investigating these properties is the focus of large-scale experimental studies, for example in particle colliders.[19] A variety of data could contain information about the properties of dark matter, from studies of cosmic rays, cosmic microwave radiation, properties of stars, gravitational lensing studies, and more. These datasets are complex: they are typically high-dimensional, represent complex relationships between the micro-physics and macro-phenomenon in a system, and may contain artefacts or noise from the instruments used to collect them. To make use of this data, researchers need to account for this complexity and tether their models to assumptions about physical processes.

The challenge for machine learning in astro-particle physics research is to extract insights about the particle composition of dark matter from the macroscopic patterns that can be observed in the Universe. For example, gravitational lensing is a phenomenon in which the pathway of light traveling through the Universe is deflected due to the influence of gravity from an intervening mass, distorting how this background light is observed [17]. Gravitational lensing effects arising from dark matter clumps ("substructure") could hold information about the structure of dark matter at a microscopic level. To infer the presence of substructure of these lensing systems, researchers need models that describe the effect of dark matter, ordinary matter, and the wider environment while simultaneously modelling the form of the background light, which can be a morphologically-complex galaxy. By letting a machine learning model, like a neural network, describe the complex background light source, it is possible to make predictions about how the light might appear after being lensed with and also without the impact of dark matter clumps. By performing many simulations considering various possibilities, researchers can compare these with observations from telescopes and understand which dark matter theories are compatible with the data.

Rapid progress in this field is generating a variety of models and approaches. In its next wave of development, further research is needed to test how trustworthy these methods are, by assessing their performance in generating physically plausible results and robust constraints on the properties of dark matter and other forms of new physics [18].

How particles move[20] across their environment is a shared area of interest for many domains. In chemistry, for example, researchers are often interested in how molecules diffuse, and where they end up distributed, based on the physical forces that shape their movement

[18] This example is inspired by Siddharth Mishra Sharma's talk, as well as insights from Gilles Louppe's talk, the abstracts for which are provided later in this document.
[19] For example: [16]
[20] This example is inspired by Francisco Vargas's talk, the abstract for which is provided later in this document.

over time. The analogy of particle movement can also be applied as an abstraction of larger scale physical processes, such as in agent-based models for crowd simulation.[21] In these systems the initial system state is represented in an initial probability distribution, the scientific objective can then also be represented as a target distribution. The dynamics underpinning this diffusion are formalised mathematically in the Schrödinger bridge problem. This long-standing problem is concerned with finding the most likely paths along which particles move from their starting distribution to their distribution at a defined point in time, based on experimentally-observed start and end positions. In general, finding analytic solutions to the Schrödinger bridge problem is intractable, but machine learning tools are providing new approaches for finding approximate numerical solutions that can be deployed across domains [21].

4.4 In biological sciences

The development and differentiation of cells into tissues and organs[22] is a complicated process, shaped by hormonal and genetic influences on cell growth [22]. Advances in genomics have allowed researchers to characterise the genetic material of different organisms; more recent progress in single-cell genomics extends this ability to the single-cell level, unlocking detailed analysis of how genetic activity determines cellular function.

Single-cell RNA studies examine how ribonucleic acids (RNA) shape cellular properties and development pathways. The RNA profiles created by genetic sequencing techniques allow researchers to identify which genes are active in a cell. The question for the field today is how to move from these single-cell analyses to an atlas of cell development that shows how cells specialise and form tissues or organs.

By combining statistical and machine learning techniques, researchers can reconstruct the gene dynamics – which genes are activated at which time – that influence cell development [23]. Cells in the small intestine, for example, undergo a pattern of differentiation that takes them from their base state to highly specialised units, able to variously secrete mucus, absorb nutrients, or respond to hormones. By studying what genes are expressed in a cell at an early stage, researchers can predict how the cell will specialise and identify which genetic changes are associated with that specialisation, opening opportunities to treat intestinal diseases [24].

Building these models relies on effective data management. Lab processes can inject artefacts into datasets, for example batch effects arising from how cells were grown or harvested for study, which need to be removed from data before analysis. Effective data correction maintains biologically-relevant information, while removing noise from the data. A variety of tools exist for this correction, including regression models, dimensionality reduction, graph methods, and deep learning. For domain researchers to be able to identify the tools that are useful for them, benchmarking studies are vital in identifying the most effective data integration method for their purpose [25]. However, there remain open questions about how best to benchmark the performance of a system when there are complex pipelines of analysis involved. Understanding the end-to-end nature of an analytical pipeline can be difficult, and new approaches to assessing performance may be needed.

[21] Examples of agent-based models for crowd simulation include: [19, 20].
[22] This example is inspired by Maren Büttner's talk, the abstract for which is provided later in this document.

To understand how the brain works,[23] neuroscientists develop mathematical models that describe the activity of individual neurons, and how these connect across brain networks. Models on the mechanistic level take the form of differential equations. These models are based on experimental data, from experiments that examine how neurons respond to different signals or perturbations. To build a computational model from this data, it is first necessary to find which factors influence how a neuron acts, creating a set of parameters that determine how the model works. This process of finding parameters is often labour-intensive, relying on trial-and-error, which limits researchers' ability to scale models across complex neural networks. Machine learning can help streamline that model definition process, by predicting which models are more likely to be compatible with data. By automatically identifying model parameters, researchers can rapidly develop simulations of complex structures, such as brains or nervous systems in different animals [26].

References

1 Tim Summers, Erik Mackie, Risa Ueno, Charles Simpson, J. Scott Hosking, Tudor Suciu, Andrew Coburn, and Emily Shuckburgh. Localized impacts and economic implications from high temperature disruption days under climate change. *Climate Resilience and Sustainability*, 1(2):e35, 2022. `doi:10.1002/cli2.35`.

2 Christian Beer, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, M. Altaf Arain, Dennis Baldocchi, Gordon B. Bonan, A. Bondeau, A. Cescatti, G. Lasslop, A. Lindroth, M. Lomas, S. Luyssaert, H. Margolis, K. W. Oleson, O. Roupsard, E. Veenendaal, N. Viovy, C. Williams, F. I. Woodward, and D. Papale. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*, 329(5993):834–838, 2010. `doi:10.1126/science.1184984`.

3 Renée M. Marchin, Diana Backes, Alessandro Ossola, Michelle R. Leishman, Mark G. Tjoelker, and David S. Ellsworth. Extreme heat increases stomatal conductance and drought-induced mortality risk in vulnerable plant species. *Global Change Biology*, 28(3):1133–1146, 2022. `doi:10.1111/gcb.15976`.

4 Christian Requena-Mesa, Markus Reichstein, Miguel Mahecha, Basil Kraft, and Joachim Denzler. Predicting landscapes from environmental conditions using generative networks. In *Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41*, pages 203–217. Springer, 2019. `doi:10.1007/978-3-030-33676-9_14`.

5 IPCC. IPCC special report on the ocean and cryosphere in a changing climate. IPCC Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2019.

6 NASA. Understanding sea level,. available from: `https://sealevel.nasa.gov/understanding-sea-level/global-sea-level/ice-melt`.

7 Dina Machuve, Ezinne Nwankwo, Neema Mduma, and Jimmy Mbelwa. Poultry diseases diagnostics models using deep learning. *Frontiers in Artificial Intelligence*, page 168, 2022. `doi:10.3389/frai.2022.733345`.

8 H. Ritchie and M. Roser. Forests and deforestation. Published online at OurWorldInData.org, `https://ourworldindata.org/forests-and-deforestation`, 2021.

9 Martin Brandt, Compton J. Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, Laurent Kergoat, Ole Mertz, Christian Igel, Fabian Gieseke, Johannes Schöning, Sizhuo Li, Katherine Melocik, Jesse Meyer, Scott Sinno, Eric Romero, Erin Glennie, Amandine Montagu, Morgane Dendoncker, and Rasmus Fensholt. An unexpectedly large count of trees in the west african sahara and sahel. *Nature*, 587(7832):78–82, 2020. `doi:10.1038/s41586-020-2824-5`.

[23] This example is inspired by Jakob Macke's talk, the abstract for which is provided later in this document.

10 Pierre Hiernaux, Hassane Bil-Assanou Issoufou, Christian Igel, Ankit Kariryaa, Moussa Kourouma, Jérôme Chave, Eric Mougin, and Patrice Savadogo. Allometric equations to estimate the dry mass of sahel woody plants mapped with very-high resolution satellite imagery. *Forest Ecology and Management*, 529:120653, 2023. doi:10.1016/j.foreco.2022.120653.

11 Luis Hernandez-Triana and Ssuzanna Bell. Taking the sting out of vector borne diseases. APHA Science Blog, available at: https://aphascience.blog.gov.uk/2022/07/06/vector-borne-diseases/, 2022.

12 John Quinn. Mapping africa's buildings with satellite imagery. Google Research Blog, available at: https://ai.googleblog.com/2021/07/mapping-africas-buildings-with.html, 2021.

13 Steve W Lindsay, Musa Jawara, Julia Mwesigwa, Jane Achan, Nabie Bayoh, John Bradley, Balla Kandeh, Matthew J. Kirby, Jakob Knudsen, Mike Macdonald, et al. Reduced mosquito survival in metal-roof houses may contribute to a decline in malaria transmission in sub-saharan africa. *Scientific reports*, 9(1):7770, 2019. doi:10.1038/s41598-019-43816-0.

14 Royal Danish Academy. New research to combat malaria mosquitoes in african metropolises. Available at: https://royaldanishacademy.com/news/ny-forskning-skal-bekaempe-malariamyg-i-afrikanske-storbyer, 2022.

15 NASA. Dark energy, dark matter. Available at: https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy. Last accessed March 4th 2023.

16 The ATLAS Collaboration et al. The ATLAS experiment at the CERN large hadron collider. *Jinst*, 3:S08003, 2008. doi:10.1088/1748-0221/3/08/S08003.

17 Siddharth Mishra-Sharma and Ge Yang. Strong lensing source reconstruction using continuous neural fields. Technical report, 2022. URL: https://arxiv.org/abs/2206.14820.

18 Cora Dvorkin, Siddharth Mishra-Sharma, Brian Nord, V. Ashley Villar, Camille Avestruz, Keith Bechtol, Aleksandra Ćiprijanović, Andrew J. Connolly, Lehman H. Garrison, Gautham Narayan, and Francisco Villaescusa-Navarro. Machine learning and cosmology. Technical report, 2022. doi:10.48550/ARXIV.2203.08056.

19 Fumiyasu Makinoshima and Yusuke Oishi. Crowd flow forecasting via agent-based simulations with sequential latent parameter estimation from aggregate observation. *Scientific Reports*, 12(1):1–13, 2022. doi:10.1038/s41598-022-14646-4.

20 Nicolas Malleson, Kevin Minors, Le-Minh Kieu, Jonathan A. Ward, Andrew West, and Alison Heppenstall. Simulating crowds in real time with agent-based modelling and a particle filter. *Journal of Artificial Societies and Social Simulation*, 23(3):3, 2020. doi:10.18564/jasss.4266.

21 Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil D. Lawrence. Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021. doi:10.3390/e23091134.

22 K. V. Krishnamurthy, Bir Bahadur, S. John Adams, and Padma Venkatasubramanian. Development and organization of cell types and tissues. *Plant Biology and Biotechnology: Volume I: Plant Diversity, Organization, Function and Improvement*, pages 73–111, 2015. doi:10.1007/978-81-322-2286-6_3.

23 Laleh Haghverdi, Maren Büttner, F. Alexander Wolf, Florian Buettner, and Fabian J. Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10):845–848, 2016. doi:10.1038/nmeth.3971.

24 Anika Böttcher, Maren Büttner, Sophie Tritschler, Michael Sterr, Alexandra Aliluev, Lena Oppenländer, Ingo Burtscher, Steffen Sass, Martin Irmler, Johannes Beckers, Christoph Ziegenhain, Wolfgang Enard, Andrea C. Schamberger, Fien M. Verhamme, Oliver Eickelberg, Fabian J. Theis, and Heiko Lickert. Non-canonical Wnt/PCP signalling regulates intestinal stem cell lineage priming towards enteroendocrine and paneth cell fates. *Nature cell biology*, 23(1):23–31, 2021. doi:10.1038/s41556-020-00617-2.

25 Malte D. Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F. Müller, Daniel C Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, and Fabian J. Theis. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods*, 19(1):41–50, 2022. doi:10.1038/s41592-021-01336-8.

26 Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A. Haddad, Tim P. Vogels, David S. Greenberg, and Jakob H. Macke. Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, 9:e56261, 2020. doi:10.7554/eLife.56261.

5 Talks given during this seminar session

5.1 Machine-learning-model-data-integration for a better understanding of the Earth System

Markus Reichstein (MPI für Biogeochemistry – Jena, DE)

The Earth is a complex dynamic networked system. Machine learning, i.e. derivation of computational models from data, has already made important contributions to predict and understand components of the Earth system, specifically in climate, remote sensing and environmental sciences. For instance, classifications of land cover types, prediction of land-atmosphere and ocean-atmosphere exchange, or detection of extreme events have greatly benefited from these approaches. Such data-driven information has already changed how Earth system models are evaluated and further developed. However, many studies have not yet sufficiently addressed and exploited dynamic aspects of systems, such as memory effects for prediction and effects of spatial context, e.g. for classification and change detection. In particular new developments in deep learning offer great potential to overcome these limitations. Yet, a key challenge and opportunity is to integrate (physical-biological) system modelling approaches with machine learning into hybrid modelling approaches, which combines physical consistency and machine learning versatility. A couple of examples are given with focus on the terrestrial biosphere, where the combination of system-based and machine-learning-based modelling helps our understanding of aspects of the Earth system.

5.2 Poultry Diseases Diagnostics Models using Deep Learning

Dina Machuve (DevData Analytics – Arusha, TZ)

Coccidiosis, Salmonella, and Newcastle are the common poultry diseases that curtail poultry production if they are not detected early. In Tanzania, these diseases are not detected early due to limited access to agricultural support services by poultry farmers. Deep learning techniques have the potential for early diagnosis of these poultry diseases. In this study, a deep Convolutional Neural Network (CNN) model was developed to diagnose poultry diseases by classifying healthy and unhealthy fecal images. Unhealthy fecal images may be symptomatic of Coccidiosis, Salmonella, and Newcastle diseases. We collected 1,255 laboratory-labeled fecal images and fecal samples used in Polymerase Chain Reaction diagnostics to annotate the laboratory-labeled fecal images. We took 6,812 poultry fecal photos using an Open Data Kit. Agricultural support experts annotated the farm-labeled fecal images. Then we used a baseline CNN model, VGG16, InceptionV3, MobileNetV2, and Xception models. We trained models using farm and laboratory-labeled fecal images and then fine-tuned them. The test set used farm-labeled images. The test accuracies results without fine-tuning were 83.06% for the baseline CNN, 85.85% for VGG16, 94.79% for InceptionV3, 87.46% for MobileNetV2, and 88.27% for Xception. Finetuning while freezing the batch normalization layer improved model accuracies, resulting in 95.01% for VGG16, 95.45% for InceptionV3, 98.02% for MobileNetV2, and 98.24% for Xception, with F1 scores for all classifiers above 75% in all four classes. Given the lighter weight of the trained MobileNetV2 and its better ability to generalize, we recommend deploying this model for the early detection of poultry diseases at the farm level. There are open questions about the deployment of the model at the farm level and potential areas for further research.

5.3 Simulation-based approaches to astrophysics dark matter searches

Siddharth Mishra-Sharma (MIT – Cambridge, US)

We are at the dawn of a data-rich era in astrophysics and cosmology, with the capacity to extract useful scientific insights often limited by our ability to efficiently model complex processes that give rise to the data rather than the volume and nature of observations itself. I will describe recent progress in applying mechanistic forward modeling techniques to a range of astrophysical observations with the goal of searching for signatures of new physics, in particular the nature of dark matter. These leverage developments in machine learning-aided inference, e.g. using simulation-based inference as well as differentiable probabilistic programming, while encoding domain knowledge, in order to maximize the scientific output of current as well as future experiments.

5.4 Single-cell transcriptomics

Maren Büttner (Helmholtz Zentrum München, DE & Universität Bonn, DE)

Cells are the fundamental units of life. Understanding cellular processes is a basis for improving human health, disease diagnosis and monitoring. The advent of single-cell transcriptomics (scRNA-seq) allows characterizing the gene expression patterns of entire organs and organisms at single cell resolution. The human genome encodes more than 30.000 genes, and high-throughput scRNA-seq methods create samples with tens of thousands of cell measurements. The analysis of such data requires a variety of methods from the machine learning field, e.g. dimensionality reduction techniques from PCA to variational autoencoders, graph-based clustering, classification of cell types, trajectory inference and causal inference of gene regulation to understand cell fate decision making. To date, scRNA-seq is a widely applied research technique, which has the potential for standard application in the clinics. My presentation focusses on current approaches for large-scale scRNA-seq data, current open questions, and implications for human health.

5.5 Estimating ecosystem properties: Combining machine learning and mechanistic models

Christian Igel (University of Copenhagen, DK)

Joint work of Christian Igel, Martin Brandt, Rasmus Fensholt, Compton J. Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, Laurent Kergoat, Ole Mertz, Fabian Gieseke, Sizhuo Li, Katherine Melo
Main reference Martin Brandt, Compton J. Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, Laurent Kergoat, Ole Mertz, Christian Igel, Fabian Gieseke, Johannes Schöning, Sizhuo Li, Katherine Melocik, Jesse Meyer, Scott Sinno, Eric Romero, Erin Glennie, Amandine Montagu, Morgane Dendoncker, Rasmus Fensholt: "An unexpectedly large count of trees in the West African Sahara and Sahel". Nature 587, 78–82 (2020).
URL https://doi.org/10.1038/s41586-020-2824-5

Progress in remote sensing technology and machine learning algorithms enables scaling up the monitoring of ecosystems. This leads to new knowledge about their status and dynamics, which will be helpful in land degradation assessment (e.g., deforestation), in mitigating poverty (e.g., food security, agroforestry, wood products), and in managing climate change (e.g., carbon sequestration).

We apply deep learning for the mapping of individual trees and forests. Tree crowns are segmented in satellite imagery using fully convolutional neural networks. This provides detailed measurements of the canopy area and of the distribution of trees within and outside forests. Allometric equations are applied to estimate the biomasses (and thereby the stored carbon) of the individual trees. We use iterative gradient-based optimization of the allometric models and suggest techniques such as jackknife+ for quantifying the uncertainty of the model predictions. Tree biomass can also be directly inferred from LiDAR (laser imaging, detection, and ranging) measurements using 3D point cloud neural networks. This leads to highly accurate results without requiring a digital elevation model.

In a new project, we consider risk assessment of vector-borne diseases based on deep learning and remote sensing. Malaria risk is related to the housing conditions, for example, the type of roofing material, which can be determined from satellite images.

5.6 Partial differential equations and Variational Bayes

Ieva Kazlauskaite (University of Cambridge, GB)

Inverse problems involving partial differential equations (PDEs) are widely used in science and engineering. Although such problems are generally ill-posed, different regularisation approaches have been developed to ameliorate this problem. Among them is the Bayesian formulation, where a prior probability measure is placed on the quantity of interest. The resulting posterior probability measure is usually analytically intractable. The Markov Chain Monte Carlo (MCMC) method has been the go-to method for sampling from those posterior measures. MCMC is computationally infeasible for large-scale problems that arise in engineering practice. Lately, Variational Bayes (VB) has been recognised as a more computationally tractable method for Bayesian inference, approximating a Bayesian posterior distribution with a simpler trial distribution by solving an optimisation problem. The talk covered some recent experiences of applying Bayesian inference, generative models and probabilistic programming languages in the context of learning material properties in civil engineering and in ice sheet and ice core modelling. The main shortcomings of PPLs and differentiable problems were highlighted.

5.7 The Schrödinger bridge problem

Francisco Vargas (University of Cambridge, GB)

Recent works in diffusion-based models have been achieving competitive results across generative modelling and inference, in this presentation we propose to explore a unifying framework based on Schrodinger bridges to explore/explain diffusion-based methodology. The Schrödinger bridge problem (SBP) finds the most likely stochastic evolution between two probability distributions given a prior (reference) stochastic evolution. Recently SBP based methodology has made its way into generative modelling, sampling, and inference. In this talk we propose the exploration of a unifying framework for the aforementioned works based on the renowned IPF/Sinkhorn algorithm. The motivation behind this is to cast a unifying lens via the Schrodinger perspective relating inference, sampling and transport, in a way that we can leverage many of the useful techniques and heuristics from each field to benefit each other.

6 Building effective simulations

6.1 Moving upstream

Science proceeds through hypothesis, observation and analysis. For hundreds of years, researchers have advanced the frontiers of knowledge by collecting data, compressing those observations into a model, then computing that model to create representations of how the world works, generating new insights about natural and physical phenomena and theories about the systems from which those phenomena emerge in the process [1]. These mathematical models rely on numerical methods: algorithms that help solve mathematical problems where no analytical solution is available. Today, data collection and the basic computational tasks involved in its analysis – linear algebra, optimisation, simulation, and so on – remain consistent features of the scientific process. Progress in machine learning, however, has changed the modelling landscape.

"AI for science" offers a data-centric approach to modelling and simulating the world. Operating alongside the traditional mathematical models that are central to many disciplines, machine learning provides data-centric analytical methods that can be integrated across the scientific pipeline, for example enabling sophisticated simulations of real-world systems. These simulations can be used to inform model development, test hypotheses and shape areas of research focus, or unlock insights from complex data.

6.2 Nurturing a diversity of approaches

Simulations are a well-established tool for scientific discovery. Their fundamental task is to allow data sampling from a model where the differences between simulation and the real world are reduced as far as feasible, to enable experimentation or testing of the impact of different perturbations, while allowing some measure of simplification of the system. Effective simulators allow researchers to move from theory to an understanding of what data should look like.

Domains such as particle physics, protein folding, climate science, and others, have developed complex simulations that use known theories and parameters of interest to make predictions about the system of study. AI for science can be brought in to speed up some of these through surrogate models. Machine learning can complement "traditional" approaches to scientific simulation, adding components that model the most uncertain elements of a system to strongly mechanistic models that might otherwise be too restrictive in their assumptions.

Much early excitement surrounding AI for science was rooted in the reverse process, asking: instead of starting with theory, could researchers instead start with the large amounts of data available in many areas of research and, from that data, build an understanding of what an underpinning theory might be? Given a set of observations, is it possible to find parameters for a model that result in simulations that reflect the measured data? Such simulation-based inference (SBI) offers the opportunity to generate novel insights across scientific disciplines.

To enable such analysis, machine learning methods are needed that can extract insights from high-dimensional, multi-modal data, in ways that are labour- and compute-efficient [2]. The field of probabilistic numerics offers a way to flexibly combine information from mechanistic models with insights from data, solving numerical problems through statistical approaches [3]. Operationalising these methods to create effective data-driven simulations requires balancing different model characteristics. The model's parameters must be specified to a sufficient level of granularity to describe the real-world system, while operating at a level of abstraction that is amenable to analysis and computation; almost all models are "wrong" or falsifiable because of this, but some level of abstraction is necessary to make them useful for analysis. The simulation must also be designed to be robust, and able to generate inferences that align with real-world observations.

6.3 Truth, truthiness, and interfacing with the real world

The excitement underpinning AI for science stems from the aspiration to unearth new understandings of the world, leveraging data to advance the frontiers of knowledge. While subject to their own limitations, the scientific community has developed checks and balances to scrutinise new knowledge and maintain the rigour of scientific inquiry. Recent years have seen a variety of challenges or benchmarks emerge in the machine learning community that have come to represent the field's expected standards of performance from algorithms on defined tasks. However, these standards do not necessarily align with the expectations of domain researchers [4]. As data-centric simulations are integrated into scientific process, machine learning researchers must consider their responsibility in maintaining the integrity of the domains into which they are deployed, raising the question: what guardrails are needed to ensure researchers can be confident in the outputs from machine learning-enabled simulations?

A variety of diagnostic tests can help. Core to many of these diagnostics is analysis of whether a model is computationally faithful. In short: the inferences generated by a simulation should reflect those from observations [4]. One approach to checking this alignment is to consider the consistency of distributions from inferred and observed datasets. If the model is a good fit, the data it generates should broadly match the data observed through experimentation.

Underpinning these diagnostics is a fundamental question about how to manage uncertainty, in a context where different failure modes have different implications. Put simply: when a model fails, is it worse to be over-confident in its results, or over-conservative? In the scientific context, over-confidence seems more likely to result in negative outcomes, whether through giving misleading interpretations or results or driving lines of enquiry in unproductive directions. Machine learning methods can be designed for conservatism, reducing the risk of false positives.

Implementing a schedule of model building, computing, critiquing, and repeating can refine this process. One lesson from experiences of building machine learning-enabled simulations is that there can be a disconnect between how machine learning approaches inference and model building, and how the same task is approached by domain scientists. From a domain perspective, model building seems naturally an iterative process: collect data, fit a model, find errors or areas for improvement, update the model, and so on. This iterative process is guided by expert intuition and knowledge; deep understanding of the system under study and how it responds to perturbation. Machine learning research has developed practices for prior elicitation – using domain knowledge to shape the structure of probabilistic models – but the nuances of this domain intuition are often not easily captured a priori, instead emerging when models fail as an informal sense of what "feels" like it should be true. This qualitative input is vital in building effective simulations. It requires close collaboration, which in turn requires an investment of time and energy from domain communities, generated through mutual trust, incentives, and long-term relationship-building.

6.4 Connecting simulation to practice

Computational tools are central to the effective deployment of machine learning-enabled simulation. The function and form of such tools must align with the requirements of the community deploying them. Designing computational systems to match user needs – and work effectively in practice – requires both effective software engineering and close collaboration with domain groups that can articulate the requirements and expectations of those working in the field. To remain effective over the longer-term, such systems must leverage effective software engineering practices, including embedding version control and building interfaces that work with other models and systems. Those practices, and the software systems that emerge from them, must be designed for the needs of those using the system, drawing from existing best practices in software engineering, but adapting those practices to reflect the needs of the domain for deployment.

Constructing computational tools requires a mix of technical insight and craft skill – of knowledge and know-how. Tools produced by the machine learning community differ in their usefulness on different problems: some work well for certain tasks, but not for others. Without access to such craft skills, those outside the "AI for science" community can find it challenging to determine which tools to use for which purposes, reducing the generalisability of existing methods and approaches. This challenge becomes particularly visible when practitioners are tightly integrated into the analysis pipeline, such as in applications in developmental biology, in the developing world, and in data-centric engineering. Widening access to the field will require user guides that characterise which simulations are effective for which tasks or purposes, supported by case studies or user stories that help demystify how machine learning can work in practice.

6.5 Directions

Machine learning typically requires an explicit representation of a likelihood, but these are often difficult to compute. Further advances in SBI are necessary to allow researchers to identify model parameters from data.

- Techniques such as likelihood-free inference can enhance existing Bayesian methods for inferring posterior estimations [5].
- Building surrogate models,[24] using Bayesian approaches for simulation planning to optimise information gain,[25] or deploying emulations [8] can also enhance the efficiency of simulations.
- Probabilistic numerics offers a route to develop statistically-optimal algorithms that are amenable to comprehensive uncertainty quantification, leveraging Gaussian Process-based Ordinary Differential Equation (ODE) solvers to pursue simulation as an inference problem [9].

Operationalising these approaches will also require new toolkits to support implementation of probabilistic numerical methods.[26]

Computational faithfulness – alignment of inferred parameters with scientific knowledge – can be achieved through:

- Diagnostic checks in the self-consistency of the Bayesian joint distribution, which measure the scientific quality of the regions computed by Bayesian SBI methods [4, 11]. Checking for self-consistency gives a sense whether the model is "good enough" (ie whether the inference engine gives a good sense of the posterior).
- Enforcing conservative neural ratio estimation through binary classifier specification, producing more conservative posterior approximations [12].
- Hybrid modelling, which combines machine learning components learned from data with the mechanistic components specified by existing domain knowledge [13].
- Further study of the impact of model misspecification could also help generate new robustness diagnostic checks [14].

"Digital twins" have recently received much attention as a tool to exploit sophisticated simulations. In Earth sciences, for example, ambitious efforts to develop a digital twin of the Earth propose to allow more accurate forecasting, visualisation, or scenario-testing of the impact of climate change and efforts to mitigate it.[27] The challenge is to integrate different models or components of a system – for example, connecting atmospheric models, with land models, with models of human behaviour – in a way that represents the complete

[24] See above, and [6].
[25] See, for example, [7].
[26] See, for example, the previous Dagstuhl meeting on this topic: `https://www.probabilistic-numerics.org/meetings/2021_Dagstuhl/` and [10].
[27] For example: [15].

Earth system. That requires consideration of the different levels of granularity with which these different models operate: economic models of human behaviour, for example, operate with different assumptions and levels of enquiry in comparison to physical models of ocean circulation. The full range of granularities becomes apparent when considering that specific applications, such as disease monitoring on poultry farms, sit within the wider ecosystem of the natural and built environment. A digital twin needs to make choices about what levels of granularity it is operating at, from the scale of the poultry farm to the planet. The questions that emerge from such ambitions is: what level of granularity is helpful or necessary to deliver effective results? And what interfaces between diverse models might be possible?

References

1 David M. Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232, 2014. doi:10.1146/annurev-statistics-022513-115657.

2 Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020. doi:10.1073/pnas.1912789117.

3 Philipp Hennig, Michael A. Osborne, and Hans P. Kersting. *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press, 2022.

4 Joeri Hermans, Arnaud Delaunoy, François Rozet, Antoine Wehenkel, and Gilles Louppe. A trust crisis in simulation-based inference? your posterior approximations can be unfaithful. Technical report, 2021. URL: https://arxiv.org/abs/2110.06581.

5 Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458, 07 2019. doi:10.1093/mnras/stz1960.

6 Alexander Lavin, Hector Zenil, Brooks Paige, David Krakauer, Justin Gottschlich, Tim Mattson, Anima Anandkumar, Sanjay Choudry, Kamil Rocki, Atılım Güneş Baydin, et al. Simulation intelligence: Towards a new generation of scientific methods. Technical report, 2021. URL: https://arxiv.org/abs/2112.03235.

7 Kyle Cranmer, Lukas Heinrich, Tim head, and Gilles Louppe. Active sciencing. Available from https://github.com/cranmer/active_sciencing, 2017.

8 Jan Boelts, Jan-Matthis Lueckmann, Richard Gao, and Jakob H. Macke. Flexible and efficient simulation-based inference for models of decision-making. *eLife*, 11:e77220, 2022. doi:10.7554/eLife.77220.

9 Hans Kersting. *Uncertainty-Aware Numerical Solutions of ODEs by Bayesian Filtering*. PhD thesis, Eberhard Karls Universität Tübingen, 2021. URL: https://hanskersting.github.io/publication/phd-thesis/.

10 Jonathan Schmidt, Nicholas Krämer, and Philipp Hennig. A probabilistic state space model for joint inference from differential equations and data. *Advances in Neural Information Processing Systems*, 34:12374–12385, 2021. URL: https://proceedings.neurips.cc/paper/2021/hash/6734fa703f6633ab896eecbdfad8953a-Abstract.html.

11 Siddharth Mishra-Sharma. Inferring dark matter substructure with astrometric lensing beyond the power spectrum. *Machine Learning: Science and Technology*, 3(1):01LT03, 2022. doi:10.1088/2632-2153/ac494a.

12 Arnaud Delaunoy, Joeri Hermans, François Rozet, Antoine Wehenkel, and Gilles Louppe. Towards reliable simulation-based inference with balanced neural ratio estimation. Technical report, 2022. URL: https://arxiv.org/abs/2208.13624.

13 Antoine Wehenkel, Jens Behrmann, Hsiang Hsu, Guillermo Sapiro, Gilles Louppe, and Jörn-Henrik Jacobsen. Robust hybrid learning with expert augmentation. Technical report, 2022. URL: https://arxiv.org/abs/2202.03881.

14 Patrick Cannon, Daniel Ward, and Sebastian M. Schmon. Investigating the impact of model misspecification in neural simulation-based inference. Technical report, 2022. URL: https://arxiv.org/abs/2209.01845.
15 European Commission. Destination Earth – new digital twin of the Earth will help tackle climate change and protect nature. Available at https://ec.europa.eu/commission/presscorner/detail/en/IP_22_1977, 2022.

7 Talks given during this seminar session

7.1 Information from data and compute in scientific inference

Philipp Hennig (Universität Tübingen, DE)

Simulations are central to scientific inference. Simulators are typically treated as black boxes, with the inference loop wrapped around them. This approach is convenient for the programming scientists, but can be highly inefficient. Probabilistic numerical methods represent computational and empirical data in the same language, which allows for inference from mechanistic knowledge and empirical data in one combined step. I will argue that scientific computing needs to embrace such new computational paradigms to truly leverage ML in science, which also requires rethinking scientific codebases.

7.2 ODE filters and smoothers: probabilistic numerics for mechanistic modelling

Hans Kersting (INRIA – Paris, FR)

Probabilistic numerics (PN) unifies statistical and numerical approximations by formulating them in the same language of statistical (Bayesian) inference. For ODEs, a well-established probabilistic numerical method is ODE filters and smoothers which can help to deal more aptly with uncertainty in mechanistic modeling. In the first half of this talk, we will first introduce PN and then present ODE filters/smoothers as a specific instance of PN. In the second half, we will discuss how ODE filters/smoothers can improve mechanistic modeling in the natural sciences and present a recent application of inferring the parameters of real-word dynamical system.

7.3 Four short stories on simulation-based inference

Jakob Macke (Universität Tübingen, DE)

Many fields of science make extensive use of simulations expressing mechanistic forward models, requiring the use of simulation-based inference methods. I will share experiences and lessons learned from four applications: Describing the dynamics and energy consumptions of neural networks in the stomatogastric ganglion; inferring parameters of gravitational wave models; optimising single-molecule localisation microscopy, and building computational models of the fly visual system. I will try to convey some thoughts on the challenges and shortcomings of current approaches.

7.4 Towards reliable simulation-based inference and beyond

Gilles Louppe (University of Liège, BE)

Modern approaches for simulation-based inference build upon deep learning surrogates to enable approximate Bayesian inference with computer simulators. In practice, the estimated posteriors' computational faithfulness is, however, rarely guaranteed. For example, Hermans et al., 2021 have shown that current simulation-based inference algorithms can produce posteriors that are overconfident, hence risking false inferences. In this talk, we will review the main inference algorithms and present Balanced Neural Ratio Estimation (BNRE), a variation of the NRE algorithm designed to produce posterior approximations that tend to be more conservative, hence improving their reliability.

7.5 Modeling the data collection process: My journey

Thomas G. Dietterich (Oregon State University – Corvallis, US)

In this talk, I will describe three examples of my attempts to integrate subject-matter knowledge with machine learning. The first example involves predicting grasshopper infestations. I will sketch the methodology in which we first modeled the life cycle of the grasshoppers to capture the factors that affect their population. Unfortunately, most variables of interest were not measured, so we used the model to guide the construction of proxy variables. Ultimately, this project did not succeed, but it is hard to determine whether this is due to modeling problems or to the chaotic nature of the biological phenomenon.

8 Connecting data to causality

8.1 Causality in science and data

Most scientific endeavours have a causal element: researchers want to characterise how a system works, why it works that way, and what happens when it is perturbed. How researchers identify cause-and-effect relationships varies across domains. For some disciplines, the process of hypothesis design – data collection – model development provides the core structure for interrogating how a system works. In others, where experimentation is more difficult, researchers may rely on natural experiments and observations to compare the response of a system under different conditions. Those studying the Earth system, for example, have little scope to replicate planetary conditions, so instead rely on observational data and modelling to identify the impact of different interventions. These different approaches, however, share a modelling approach in which researchers provide variables to create structural, causal models.

In contrast, machine learning proceeds by learning representations or rules from data, based on statistical information, rather than structured rules about how a system works (such as physical laws). Causal inference – the ability to identify cause-and-effect relationships in data – has been a core aim of AI research, in service of both wider ambitions to replicate intelligence in machines and efforts to create AI systems that are robust in deployment. However, in many respects efforts to integrate causal inference into AI systems have yet to deliver [1].

An apocryphal story in AI tells of efforts by US researchers during the 1980s to train a computer system that could distinguish between images of tanks from the US and USSR. The resulting system delivered high accuracy on its training data, but failed repeatedly in practice. The system was subsequently found to be classifying images based on their resolution and background features – is the image grainy? Does it contain snow? – rather than the tanks themselves. It found patterns in the data that were co-incident, rather than causal. That same error has real-world implications for the AI systems deployed today. In medical sciences, AI systems trained to detect collapsed lungs from medical images have been proven inaccurate, after the model was found to have learned to detect the tube inserted into the lung to enable a patient to breath as a response to its collapse, rather than the physical features of the lung itself [2]. In medical sciences, deployment of such systems could put patient care at risk. In social sciences, these AI design and data bias failures can combine to marginalise vulnerable populations [3].

Conversely, an understanding of the structures within data can improve the accuracy of machine learning analyses. In exoplanet discovery, for example, machine learning is used as a tool to detect variations in light signals from large-scale astronomical datasets. The movement of exoplanets around stars results in periodic changes to the light signals from those stars, as the planet obscures them in its transit. Machine learning can detect those signals and predict where exoplanets might be located, but the data is often noisy. Noticing that the structure of this noise was consistent across a number of stars, which were too distant from each other to be interacting, researchers concluded that instrumentation effects were distorting the data, and developed a method to model those effects and remove them from exoplanet predictions. The result was an efficient method for exoplanet identification that subsequently contributed to the discovery of the first potentially habitable planet [4].

8.2 Causal models as a route to advancing the science of AI and AI for science

Many of these errors in misdiagnosing cause-effect relationships arise from a core assumption in many machine learning methods: that data follows an independent and identical distribution (IID). In practice, almost all data from real-world, or complex, systems will violate this assumption, given the interconnectedness of different variables. The task of causality in machine learning is to create models that can manage this violation, distinguishing between patterns in data that simply co-occur and patterns that are causal. The resulting AI systems would be able to solve a task in many different environments, based on an understanding of the fundamental causal mechanisms in a system [5]. They would be more robust in deployment, being less likely to make incorrect predictions as the environment in which they operate changes, and could be more efficient to train and deploy. They would also represent a step towards replicating human- or animal-like intelligence, being able to solve a task in many different environments.

In these regards, causal machine learning offers a route to balancing the widespread utility of statistical modelling with the strengths of physical models. Causality allows models to operate at a level of abstraction beyond strongly mechanistic approaches, such as those based on differential equations, moving along a continuum from mechanistic to data-driven modelling. They provide researchers with the ability to make accurate predictions under conditions of dataset shift (enable out of distribution generalisation); can provide insights into the physical processes that drive the behaviour of a system; unlock progress towards AI systems that "think" in the sense of acting in an imagined space; while also leveraging insights that can be learned from data, but not otherwise detected.[28] They also offer opportunities to explore counterfactuals in complex systems, asking what the impact of different interventions could have been, opening a door to the development of simulation-based decision-making tools.[29]

Achieving this potential requires technical developments in a number of directions, but can also yield more effective AI systems. Such systems would:

- Be able to operate on out of distribution data, performing the task for which they are trained in environments with varying conditions.
- Be able to learn how to perform a task based on relatively few examples of that task in different conditions, or be able to rapidly adapt what they have learned for application in new environments through transfer, one-shot, or lifelong learning approaches.
- Support users to analyse the impact of different interventions on a system, providing explanations or ways of attributing credit to different actions.
- Respond to different ways of transmitting information between individuals and groups, enabling effective communication with their users or other forms of cultural learning.

8.3 From methods to application

Achieving the level of technical sophistication required for causal modelling requires careful model design, based on close collaboration between machine learning and domain scientists. The process of specifying what to represent in a causal machine learning system involves

[28] For reference, see the table on page 11 of reference [4].
[29] Such tools may have particular relevance in policy. For example: [6].

a series of "micro-decisions" about how to construct the model, negotiated by integrating machine learning and domain expertise. In this regard, causal machine learning can be a positive catalyst for deeper interdisciplinary collaboration; model construction can be a convening point for sharing understandings between domains. However, the level of detail required can also be in tension with efforts to promote widespread adoption of AI methods across research. The availability of easy-to-use, off-the-shelf AI tools has been an enabler for adoption in many domains. The hand-crafted approach inherent to current causal methods renders them less accessible to non-expert users. Part of the challenge for the field is to make such methods more broadly accessible through open-source toolkits or effective software engineering practices.

This tension between specification and learning also highlights the importance of nurturing a diversity of methods across the spectrum from data-driven to mechanistic modelling. The domain (or, how much prior knowledge is available and what knowledge should be included), research question of interest, and other practical factors (including, for example, compute budget), will shape where along this spectrum researchers wish to target their modelling efforts.

While pursuing practical applications, advances in causal inference could help answer broader questions about the nature of intelligence and the role of causal representations in human understanding of how the worlds work. Much of human understanding of the world arises from observing cause and effect; seeing what reaction follows an intervention – that an object falls when dropped, for example – in a way that generalises across circumstances and does not require detailed understanding of mathematical or physical laws. Integrating this ability into machine learning would help create systems that could be deployed on a variety of tasks. The process of building causal machine learning forces researchers to interrogate the nature of causal representations – What are they? How are they constructed from the interaction between intelligent agents and the world? By what mechanism can such agents connect low-level observations to high-level causal variables? – which may in turn support wider advances in the science of AI.

8.4 Directions

Causality in machine learning is a long-standing and complex challenge. In the context of scientific discovery, learning strategy, model design, and encoding domain knowledge all play a role in helping identify cause-effect relationships.

Different learning strategies can improve the "generalisability" of machine learning, increasing its performance on previously unseen tasks, based on learning underlying structure of a task or environment in ways that can contribute to broader understandings of causality. Such learning strategies include:

- Transfer learning, taking learning from one task or domain and applying it in another.
- Multi-task learning, enabling a system to solve multiple tasks in multiple environments.
- Adversarial learning, to reduce the vulnerability of models to performance degradation on out-of-distribution data.
- Causal representation learning, defining variables that are related by causal models [4].
- Reinforcement learning strategies that reward agents for identifying policies based on invariances over different conditions.

Across these new learning approaches, attempts to establish causal mechanisms are also prompting progress in machine learning theory, through statistical formulations of core principles [7].

Combining different methods can also enhance the functionality of an AI system. For example:

- Neural ODEs have been shown to identify causal structures in time series data [8].
- Describing causal effects as objective functions in constrained optimisation problems can deliver a form of stochastic causal programming [9].
- Technical interventions [10] can constrain or optimise a model towards causal outcomes. As with simulation design, diagnostic checks can also help identify cause-effect relationships by examining model outputs against "reality criteria",[30] which compare outputs to real-world results.

There are also a variety of approaches to representing existing scientific knowledge in machine learning models, notably by specifying the assumptions made about the world through symmetries, invariances, and physical laws (see Figure 1).

References

1 Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021. doi:10.1109/JPROC.2021.3058954.

2 Johannes Rueckel, Lena Trappmann, Balthasar Schachtner, Philipp Wesp, Boj Friedrich Hoppe, Nicola Fink, Jens Ricke, Julien Dinkel, Michael Ingrisch, and Bastian Oliver Sabel. Impact of confounding thoracic tubes and pleural dehiscence extent on artificial intelligence pneumothorax detection in chest radiographs. *Investigative Radiology*, 55(12):792–798, 2020. doi:10.1097/RLI.0000000000000707.

3 Jesse Emspak. How a machine learns prejudice. Scientific American Blogs, available at: https://www.scientificamerican.com/article/how-a-machine-learns-prejudice/, 2016.

4 Bernhard Schölkopf. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 765–804. 2022. doi:10.1145/3501714.3501755.

5 Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, Cambridge, MA, USA, 2017.

6 Atalanti Mastakouri and Bernhard Schölkopf. Causal analysis of covid-19 spread in germany. *Advances in Neural Information Processing Systems*, 33:3153–3163, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/205e73579f21c2ed134dbd6ce7e4a1ea-Abstract.html.

7 Siyuan Guo, Viktor Tóth, Bernhard Schölkopf, and Ferenc Huszár. Causal de Finetti: On the identification of invariant causal structure in exchangeable data. Technical report, 2022. URL: https://arxiv.org/abs/2203.15756.

8 Hananeh Aliee, Fabian J. Theis, and Niki Kilbertus. Beyond predictions in neural odes: Identification and interventions. Technical report, 2021. URL: https://arxiv.org/abs/2106.12430.

9 Kirtan Padh, Jakob Zeitler, David Watson, Matt Kusner, Ricardo Silva, and Niki Kilbertus. Stochastic causal programming for bounding treatment effects. Technical report, 2022. URL: https://arxiv.org/abs/2202.10806.

10 Martin Emil Jakobsen and Jonas Peters. Distributional robustness of k-class estimators and the pulse. *The Econometrics Journal*, 25(2):404–432, 2022. doi:10.1093/ectj/utab031.

11 Michael Stadler and Peter Kruse. *Über Wirklichkeitskriterien*. Suhrkamp, Frankfurt am Main, Germany, 1990.

[30] Including syntactic, semantic, and pragmatic elements: [11].

9 Talks given during this seminar session

9.1 Causality, causal digital twins, and their applications

Bernhard Schölkopf (MPI für Intelligente Systeme – Tübingen, DE)

1. Desiderata for causal machine learning: work with (and benefit from) non-IID data, multi-task/multi-environment, sample-efficient, OOD, generalisation from observation of marginals, interventional.
2. Modelling taxonomy: differential equations, causal models, statistical models.
3. How to get from one level to the next.
4. How to transfer between statistical models that share the same underlying causal model.
5. The assumption of independent causal mechanisms (ICM) (for example, invariance/autonomy) and sparse mechanism design.
6. How to derive the arrow of time from ICM and algorithmic information theory.
7. Statistical formulation of ICM: causal de Finetti.
8. Application to exoplanet discovery and Covid-19 vaccine scenarios.
9. Causal representations as (a) causal digital twins and (b) AI models.

9.2 Invariance: From Causality to Distribution Generalization

Jonas Peters (University of Copenhagen, DK)

Assume that we observe data from a response Y and a set of covariates X under different experimental conditions (or environments). Rather than focusing on the model that is most predictive, it has been suggested to take into account the invariance of a model. This can help us to infer causal structure (Which covariates are causes of Y?) and find models that generalize better (How well does the model perform on an unseen environment?). We show a few applications of these general principles and discuss first steps towards understanding the corresponding theoretical guarantees and limits.

9.3 Can we discover dynamical laws from observation?

Niki Kilbertus (TU München, DE & Helmholtz AI München, DE)

I will start with a brief introduction to identifiability of ODE systems from a unique continuous or discrete observed solution trajectory. Then, I will provide an overview of modern approaches to inferring dynamical laws (in the form of ODEs) from observational data with a particular focus on interpretability and symbolic methods. Finally, I will describe our recent attempts and results at inferring scalar ODEs in symbolic form from a single irregularly sampled, noisy solution trajectory.

9.4 Invariances and equivariances in machine learning

Soledad Villar (Johns Hopkins University – Baltimore, US)

In this talk, we give an overview of the progress in the last few years by several research groups in designing machine learning methods that repeat physical laws. Some of these frameworks make use of irreducible representations, some make use of high-order tensor objects, and some apply symmetry enforcing constraints. Our work shows that it is simple to parameterise universally approximating functions that are equivariant under actions of the Euclidean, Lorentz, and Poincare group at any dimensionality. The key observation is that $O(d)$-equivariant (and related group-equivariant) functions can be universally expressed in terms of a lightweight collection of dimensionless scalars (scalar products and scalar contractions of the scarla, vector, and tensor inputs). We complement our theory with numerical examples that show that the scalar-based method is simple and efficient, and mention ongoing work on cosmology simulations.

9.5 Divide-and-Conquer Equation Learning with R2 and Bayesian Model Evidence

Bubacarr Bah (AIMS South Africa – Cape Town, ZA)

Deep learning is a powerful method for tasks like predictions and classification, but lacks interpretability and analytic access. Instead of fitting up to millions of parameters, an intriguing alternative for a wide range of problems would be to learn the governing equations from data. Resulting models would be concise, parameters can be interpreted, the model can adjust to shifts in data, and analytic analysis allows for extra insights. Common challenges are model complexity identification, stable feature selection, expressivity, computational feasibility, and scarce data. In our work, the mentioned challenges are addressed by combining existing methods in a novel way. We choose multiple regression as a framework and argue how a surprisingly large space of model equations can be captured. For feature selection, we exploit the computationally cheap coefficient of determination (R2) to loop through millions of models, and by using a divide-and-conquer strategy, we are able to rule out remaining models in the equation class. Final model selection is achieved by exact values of the Bayesian model evidence with empirical priors, which is known to identify suitable model complexity without relying on mass data. Random polynomials, and a couple of chaotic systems are used as examples.

10 Encoding domain knowledge

10.1 Where's My [Science] Jetpack?

Humans have a long history of imagining futures where human progress is accelerated by intelligent machines. Embedded in these visions for the future are aspirations that AI

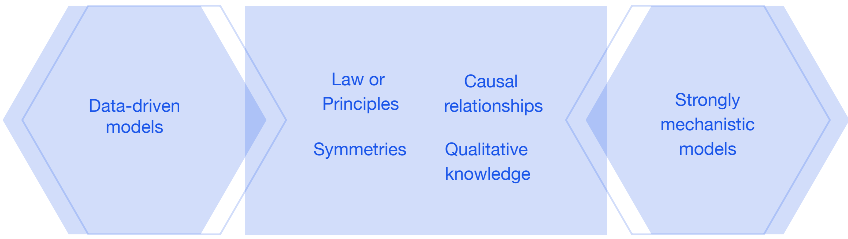


Figure 1 Models along a spectrum from classical i.i.d models to strongly mechanistic differential equation models introduce aspects of causality and symmetries to create a continuum between mechanistic and data-driven worlds. Statistical or data-driven models are weakly mechanistic (i.e. they include smoothness assumptions or similar).

can be a faithful servant, easing daily activities or enhancing human activities [16]. As with many emerging technologies, the reality of AI today looks different to these Sci-Fi futures.[31] Practical experiences of deploying AI highlights a range of potential failure modes, often rooted in insufficient contextual awareness, misspecification of user needs, or misunderstanding of environmental dynamics [14].

Today's science builds on thousands of years of attempts to understand the world, which can be leveraged to design AI that serves scientific goals. The result should be a collaborative endeavour between humans and machines. Researchers need the analytical power of AI to make sense of the world, while AI needs input from human understandings of the domain in which it is deployed to function effectively; both need well-designed human-machine interfaces to make this collaboration work. In this context, effective integration of domain knowledge into AI systems is vital, and three (broad) strategies have emerged to facilitate this encoding: algorithmic design; AI integration in the lab; and effective communication and collaboration.

10.2 Encoding domain knowledge through model design

Traditional modelling approaches make use of well-defined rules or equations that explain the dynamics of the system under study. The laws of physics, for example, describe how energy moves through a system, based on conservation principles. These laws are complemented by mathematical symmetries that arise from our abstract representations of physical objects and describe what features of an object remain consistent, despite changes or transformations in a system [17]. There may also be known invariances in a system: factors that do not change under any perturbations or that change in a defined way [11]. Building on this existing knowledge, and connecting to efforts to generate causal understandings of the world through machine learning, an area of growing interest has been the design of machine learning models that respect these rules or symmetries.

The principle underpinning this design strategy is that it is possible to move across a continuum from statistical (data-driven) models to strongly mechanistic models, creating hybrid systems whose outputs should be constrained by what is physically feasible, while also leveraging insights from data (Figure 1).

[31] The title of this section is inspired by: `https://www.fantasticfiction.com/w/daniel-h-wilson/where-s-my-jetpack.htm`

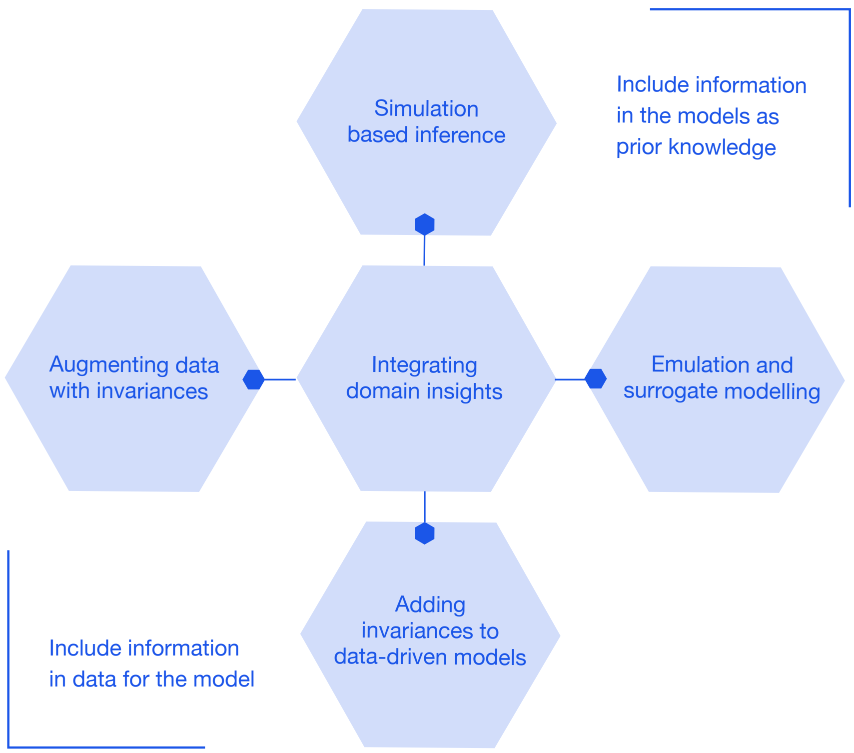


Figure 2 Strategies for integrating domain insights: including information in data and including information as prior knowledge.

At one end of that continuum, mechanistic models would obey known laws or principles in a strongly deterministic way; at the other, statistical models encode fewer assumptions and rely more on data [10]. The addition of invariances and symmetries, alongside other forms of domain knowledge, allows bridging between these two model classes (Figure 1). Models that describe how much heat is absorbed by the oceans under conditions of climate change, for example, should obey the laws of thermodynamics and energy conservation. By encoding the domain knowledge that has yielded these fundamental laws, such as the conservation of momentum or energy, researchers can ensure the outputs of a machine learning model will have a physically allowable expression. This encoding can come from integrating equations, symmetries, or invariances into model design. These encodings constrain the operation of a machine learning system to align with the known dynamics of physical systems. The resulting models might be expected to produce more accurate results, with smaller generalisation errors, and with better out-of-distribution generalisation.

10.3 Scientific centaurs

Complementing modelling strategies to encode scientific knowledge are deployment strategies to use AI in the lab. The lab has long provided a physical hub for collaboration and knowledge-generation, its function and form having remained broadly consistent across centuries of scientific progress. Today, the digitisation of experimental equipment and laboratory processes offers opportunities to integrate AI in experimental design and create new virtual labs.

By combining data from measurement devices, simulations of laboratory processes, and computational models of research or user objectives, these virtual labs provide a digital sibling of in-person research activities that can be used to optimise such activities. In drug discovery, for example, virtual labs could accelerate the testing and analysis processes that identify candidate drugs from potential drug targets. Instead of relying on physical testing of such starting molecules, multiple rounds of virtual testing can rapidly simulate the processes of drug design, manufacture, testing, and analysis to assess which starting molecules are more (or less) likely to be viable candidate drugs [8]. As a result, AI can help accelerate the research process.

Advances in machine learning methods to enable effective simulations, causal modelling, and encoding pre-existing domain insights – while packaging such methods into usable toolkits – are all necessary foundations for such digital siblings. Moving from virtual laboratory to "AI assistants" requires further advances in AI system design to create AI agents that can elicit guidance or input from their domain experts. Such agents would not only provide useful intuitions for scientific modelling, but would serve as "scientific sidekicks", actively helping researchers to drive their research.

This new type of AI assistant would combine the ability to model the research problem of interest with the ability to model the goals and preferences of their expert users, even when the user themselves might not be able to clearly articulate those goals. As a starting point, these systems would need to support forms of user interaction that can extract user knowledge, leveraging this to identify appropriate courses of action. To operate in contexts where user goals might be uncertain and user behaviour might change in response to the outputs of the AI system, these AI sidekicks will need insights from cognitive science, studies of team decision-making, and new learning strategies based on limited examples. The sophisticated user modelling so-created would unlock new forms of human-AI collaboration; scientific centaurs that combine both human and machine intelligence [3].

10.4 Enabling communication across domains

Underpinning these efforts to integrate pre-existing knowledge into the design and deployment of AI systems is a feedback loop between domain and machine learning research, in which each elicits from and feeds into the other. This loop requires the ability to exchange knowledge and insights across disciplines through interdisciplinary collaboration and communication.

Matching model to user need requires shared understandings of the research question at hand, the constraints – whether from data, compute, funding, or time and energy available – that affect different collaborators, and the user needs of the domain environment. While AI researchers might be tempted to develop complex models, showcasing assorted theoretical and methodological advances in the field, from a domain perspective, a relatively "simple" model may seem preferable. Collaborators need to be able to mutually explore what is possible, while also considering what is useful.

To complete the loop, outputs from machine learning models need to feed back into the application domain: insights from AI need to be accessible in ways that allow the transfer of learning from model to user. This implies some level of explainability. It is not sufficient for an AI system to produce highly accurate results; those results must also be interpretable by a domain researcher. As the complexity of AI systems increases, however, understanding why these systems have produced a particular result becomes increasingly challenging. While not an issue for all machine learning methods, this complexity often results in difficulties explaining the functioning of AI systems.

In response, AI researchers have developed a variety of different methods to interrogate how AI systems work, or why a particular output has been produced. Again, to understand which of these methods is desirable in the context of a scientific application, researchers must collaborate closely with domain experts. In the context of pharmaceutical experiments where the aim is to measure how many target cells are killed off at different dosages of a drug (or drug combination), for example, researchers might be seeking to "sense-check" how different drug dosages affect the model, before investigating specific drugs more rigorously. In astronomical studies, researchers are often working with high-dimensional datasets with many confounding correlations. For example, gravitational waves are ripples in space-time catalysed by the movement of massive bodies in space, such as planets or stars [13]. These invisible phenomena are studied at observatories across the world,[32] based on models to describe wave signals and the "noise" generated by instruments that measure them [4]. Measurements of gravitational waves can be used to infer the properties of black holes that create them, such as their location, mass, and spin, using simulation-based inference to characterise the source of a wave, given the data that detects it. To make such methods more efficient than existing analytical tools, researchers need to take into account the structure that sits underneath it: for example, gravitational wave detectors are located across the globe, and their location affects the angle at which they detect waves hitting the Earth. This structure can be exploited through data sampling strategies to help make machine learning more efficient [4]. An alternative, however, is to use deterministic models that already reflect relevant physical laws [2]. Across these approaches, software packages play an important role in enabling communication and dissemination of methods for wider use.[33]

10.5 Directions

New modelling approaches and mathematical innovations offer exciting opportunities to integrate domain knowledge, symmetries and invariances into AI systems [18]. Integration can be achieved in different ways:

- Data augmentation can help exploit invariances and symmetries, resulting in improved model performance, by including in the data domain knowledge for a model to ingest.
- Symmetries can be embedded in the design of deep learning systems, for example by using the same convolutional filters in different locations of an image, CNNs can leverage translation and rotation symmetries.
- Latent force models allow representations of known symmetries alongside probabilistic factors, enabling integration of mechanistic models with unknown forces [1, 19].
- Architectural features can restrict model focus to outputs that satisfy symmetries, for example using weight sharing, irreducible representations, or invoking symmetries as constraints.[34]
- Loss functions can be deployed to penalise predictions that fail to satisfy physical constraints or symmetries.

In the process, emerging mathematical questions include: how can AI learn invariances from data? And is it possible to quantify the performance gain achieved through this?

[32] See, for example, the LIGO project. Information available at: `https://www.ligo.caltech.edu`
[33] See, for example: `https://lscsoft.docs.ligo.org/bilby/`
[34] See, for example: [9, 12, 6]

Research to develop AI assistants in the lab raises interesting questions about learning strategies and human-machine collaboration. These AI agents would need to be able to learn how to assist another agent, in a multi-agent decision-making scenario, where goals might be unclear, uncertain, or changeable. To tackle this challenge:

- Decision-making with delayed reward or zero-shot learning can help agents solve tasks when there is little or nothing known about the reward function, and no previous behaviour to learn from.
- Interactive knowledge elicitation [15], combining prior knowledge from cognitive science with learning from data [7], and generative user models [5] can support more effective interactions between user and machine.

Across these areas, care is needed in the design of the points of interaction between human and AI system. A core question here is: how can AI researchers extract domain knowledge from relevant experts and integrate it into a machine learning model? Insights from human-machine interaction studies and collaborative decision-making systems are necessary to create effective interfaces between human and machine, based on factors such as:

- What forms of visualisation are helpful for human users?
- What types of interpretability or explainability are needed for a user to achieve their desired interactions?
- What might be the unintended consequences of human-machine interaction, such as over-confidence in results or over-reliance on the AI system?
- What "theory of mind" is needed to anticipate how human users might be likely to respond to an AI system?

A challenge in these interactions is that much of the relevant knowledge held by the domain expert might be qualitative: an intuition of how a system works, developed over a long period of study, rather than quantifiable insights.

References

1 Mauricio A. Álvarez, David Luengo, and Neil D. Lawrence. Linear latent force models using Gaussian processes. *TPAMI*, 35(11):2693–2705, 5 2013. `doi:10.1109/TPAMI.2013.86`.

2 Erik Bodin, Zhenwen Dai, Neill Campbell, and Carl Henrik Ek. Black-box density function estimation using recursive partitioning. In *International Conference on Machine Learning*, volume 139, pages 1015–1025. PMLR, 2021. URL: `https://proceedings.mlr.press/v139/bodin21a.html`.

3 Mustafa Mert Çelikok, Frans A. Oliehoek, and Samuel Kaski. Best-response bayesian reinforcement learning with bayes-adaptive pomdps for centaurs. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS) 2022*, 2022. URL: `https://ui.adsabs.harvard.edu/abs/2022arXiv220401160M/abstract`.

4 Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.*, 127:241103, Dec 2021. `doi:10.1103/PhysRevLett.127.241103`.

5 Sebastiaan De Peuter, Antti Oulasvirta, and Samuel Kaski. Toward ai assistants that let designers design. Technical report, 2021. URL: `https://arxiv.org/abs/2107.13074`.

6 Nadav Dym and Haggai Maron. On the universality of rotation equivariant point cloud networks. Technical report, 2020. URL: `https://arxiv.org/abs/2010.02449`.

7 Antti Kangasrääsiö, Jussi PP Jokinen, Antti Oulasvirta, Andrew Howes, and Samuel Kaski. Parameter inference for computational cognitive models with approximate bayesian computation. *Cognitive science*, 43(6):e12738, 2019. `doi:10.1111/cogs.12738`.

8 Arto Klami, Theodoros Damoulas, Ola Engkvist, Patrick Rinke, and Samuel Kaski. Virtual laboratories: Transforming research with ai. Technical report, 2022. URL: https://www.techrxiv.org/articles/preprint/Virtual_Laboratories_Transforming_research_with_AI/20412540.
9 Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, volume 80, pages 2747–2755. PMLR, 2018. URL: https://proceedings.mlr.press/v80/kondor18a.html.
10 Neil D. Lawrence. Introduction to learning and inference in computational systems biology. In Neil D. Lawrence, Mark Girolami, Magnus Rattray, and Guido Sanguinetti, editors, *Learning and Inference in Computational Systems Biology*, chapter 1. MIT Press, Cambridge, MA, 2010. URL: https://mitpress.mit.edu/9780262013864/learning-and-inference-in-computational-systems-biology/.
11 Julia Ling, Reese Jones, and Jeremy Templeton. Machine learning strategies for systems with invariance properties. *Journal of Computational Physics*, 318:22–35, 2016. doi:10.1016/j.jcp.2016.05.003.
12 Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. Technical report, 2018. URL: https://arxiv.org/abs/1812.09902.
13 NASA. What is a gravitational wave? Available at: https://spaceplace.nasa.gov/gravitational-waves/en/.
14 Andrei Paleyes, Raoul-Gabriel Urma, and Neil D. Lawrence. Challenges in deploying machine learning: a survey of case studies. *ACM Computing Surveys*, 55(6):1–29, 2022. doi:10.1145/3533378.
15 Iiris Sundin, Tomi Peltola, Luana Micallef, Homayun Afrabandpey, Marta Soare, Muntasir Mamun Majumder, Pedram Daee, Chen He, Baris Serim, Aki Havulinna, Caroline Heckman, Giulio Jacucci, Pekka Marttinen, and Samuel Kaski. Improving genomics-based predictions for precision medicine through active elicitation of expert knowledge. *Bioinformatics*, 34(13):i395–i403, 06 2018. doi:10.1093/bioinformatics/bty257.
16 The Royal Society. Ai narratives: portrayals and perceptions of artificial intelligence and why they matter. Available at: https://royalsociety.org/topics-policy/projects/ai-narratives/, 2018.
17 Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. *Advances in Neural Information Processing Systems*, 34:28848–28863, 2021. URL: https://proceedings.neurips.cc/paper/2021/hash/f1b0775946bc0329b35b823b86eeb5f5-Abstract.html.
18 Soledad Villar, Weichi Yao, David W Hogg, Ben Blum-Smith, and Bianca Dumitrascu. Dimensionless machine learning: Imposing exact units equivariance. Technical report, 2022. URL: https://arxiv.org/abs/2204.00887.
19 Wil Ward, Tom Ryder, Dennis Prangle, and Mauricio Alvarez. Black-box inference for non-linear latent force models. In *International Conference on Artificial Intelligence and Statistics*, volume 108, pages 3088–3098. PMLR, 2020. URL: http://proceedings.mlr.press/v108/ward20a.html.

11 Talks given during this seminar session

11.1 Virtual laboratories for science, assisted by collaborative AI

Samuel Kaski (Aalto University, FI)

I introduced two ideas: virtual laboratories for science, aiming to introduce an interface between algorithms and domain science that enables AI-driven scale advantages, and AI-based 'sidekick' assistants, able to help other agents research their goals, even when they are not able to yet specify the goal explicitly, or it is evolving. Such assistants would ultimately be able to help human domain experts run experiments in the virtual laboratories. I invited researchers to join the virtual laboratory movement, both domain scientists in hosting a virtual laboratory in their field and methods researchers in contributing new methods to virtual laboratories, simply by providing compatible interfaces in their code. For developing the assistants, I introduced the basic problem of agents that are able to help other agents reach their goals, also in zero-short settings, formulated the problem, and introduced solutions in the simplified setting of prior knowledge elicitation, and in AI-assistted decision and design tasks.

11.2 Making data analysis more like classical physics

David W. Hogg (New York University, US)

The laws of physics are very structured: They involve coordinate-free forms, they are equivariant to a panoply of group actions, and they can be written entirely in terms of dimensionless, invariant quantities. We find that many existing machine-learning methods can be very straightforwardly modified to obey the rules that physical law must obey; physics structure can be implemented without big engineering efforts. We also find that these modifications often lead to improvements in generalization, including out-of-sample generalization, in natural-science contexts. We have some intuitions about why.

The second example is work by Dan Sheldon on analysis of doppler radar to extract bird biomass and motion. The radar measures the radial velocity modulo a constant (i.e., the velocity wraps around to zero). Previous work had attempted to "unwrap" the data using heuristics. Dan instead incorporated the modulus operation into the likelihood function and then developing an algorithm for maximizing this somewhat nasty likelihood. The result has revolutionized radar analysis and has been deployed in the BirdCast product from the Cornell Lab of Ornithology.

The third example is the species occupancy model introduced by MacKenzie et al (2002). When human observers conduct wildlife surveys, they may fail to detect a species even though the species is present. The occupancy model combines this detection probability with a habitat model. However, the expressiveness of the two models (detection and habitat) must be carefully controlled. Rebecca Hutchinson and I learned this when we tried to replace the linear logistic regression models with boosted trees.

In all cases, downstream use of the estimates that come from such data collection models must be aware of the measurement uncertainties. How can we correctly quantify those uncertainties and incorporate them in the downstream analysis? Maybe there are lessons ecologists can learn from physicists?

11.3 Latent force models

Mauricio A. Álvarez (University of Manchester, GB)

A latent force model is a Gaussian process with a covariance function inspired by a differential operator. Such a covariance function is obtained by performing convolution integrals between Green's functions associated with the differential operators, and covariance functions associated with latent functions. Latent force models have been used in several different fields for grey box modelling and Bayesian inversion. In this talk, I will introduce latent force models and several recent works in my group where we have extended this framework to non-linear problems.

11.4 Translating mechanistic understandings to stochastic models

Carl Henrik Ek (University of Cambridge, GB)

Statistical learning holds the promise of being the glue that allows us to improve knowledge parametrised explicitly by a mechanistic model with implicit knowledge through empirical evidence. Statistical inference provides a narrative of how to integrate these two sources of information leading to an explanation of the empirical evidence in "light" of the explicit knowledge. While the two sources of knowledge are exchangeable in terms of predictive performance they are not if our focus is that of statistical learning as a tool for science where we want to derive new knowledge.

In this talk we will focus on challenges associated with translating our mechanistic understanding into stochastic models such that they can be integrated with data. In particular, we will focus on the challenges of translating composite knowledge. We will show how these structures and the computational intractabilities they lead to make knowledge discovery challenging.

The perceived "success" of machine learning comes from application where we have large volumes of data such that only simple and generic models are needed in order to regularise the problem. This means that much of the progress that have been made with predictive models are challenging to translate into useful mechanisms for scientific applications. In this talk we will focus on challenges associated with translating our mechanistic understanding into stochastic models such that they can be integrated with data. In specific we will focus on the challenges of translating composite knowledge. We will show how these structures and the computational intractabilities they lead to makes knowledge discovery challenging. We will discuss properties that we desire from such structures and highlight the large gap that exists with current inference mechanism.

12 A research agenda in AI for science

"AI for science" sits at a nexus of disciplines, methods, and communities. Both AI and "science" (broadly defined) share a core interest in learning from data. From this interest emerge different research directions: for AI, questions about the nature of intelligence and how to understand the learning process in humans and machines; for science, the outputs of this learning process are the focus, with the aim of adding new knowledge about natural, physical, and social systems. A distinctive feature of the emerging "AI for science" agenda is the ability to move between these worlds, using AI to drive progress in science and taking inspiration from science to inspire progress in AI. The result is a continuum of modelling approaches along a spectrum from strongly mechanistic to statistical models, which allow researchers to introduce or operate at different levels of abstraction.

The AI for science community therefore combines the ambitions of AI research with domain-specific goals to advance the frontiers of research and innovation in their discipline, with an engineering focus on designing systems that work in deployment, while operating across scales from the nano- to the interstellar. From these interfaces emerges a research agenda that – if successful – promises to accelerate progress across disciplines. Inspired by discussions at the Dagstuhl Seminar, a list of research questions arising from this agenda is given in Annex 2. These span three themes:

Building AI systems for science: Attempts to deploy AI in the context of scientific discovery have exposed a collection of gaps in current machine learning and AI capabilities. Further work is needed to develop the technical capabilities that will allow AI to be used more effectively in research and innovation; developing those capabilities also offers opportunities to contribute to wider attempts to deliver sophisticated AI systems. Areas for progress include:

- Advancing methods, software and toolkits for high-quality simulation and emulation, which integrate effective uncertainty quantification and leverage advances in machine learning robustness to ensure they operate safely and effectively.
- Detecting scientifically meaningful structure in data, through advances in causal machine learning.
- Encoding domain knowledge in AI systems through integration of scientific laws, principles, symmetries, or invariances in machine learning models, and through virtual, autonomous systems to make research more effective.

Combining human and machine intelligence: Effective deployment of AI in science requires effective interactions between human, domain and machine intelligence across all stages of the deployment pathway. AI systems can be made more effective by integrating pre-existing knowledge about the system of study, but mechanisms are needed to extract and encode that knowledge. Effective interfaces are also required in the reverse direction. Translating the outputs of AI analysis to increased human capability requires an understanding of what insights are relevant, how they are best communicated, and the cultural environment that shapes the conduct of science. Areas for progress include:

- Designing interfaces between humans and machines or AI agents that can extract, formalise, and assimilate knowledge that domain researchers have acquired, including tacit knowledge, and that communicate new knowledge back to the user as actionable insights.
- Building mechanisms for explainability that allow researchers to interrogate why and how an AI system delivered a particular result, with the explanations provided being tailored to user need.

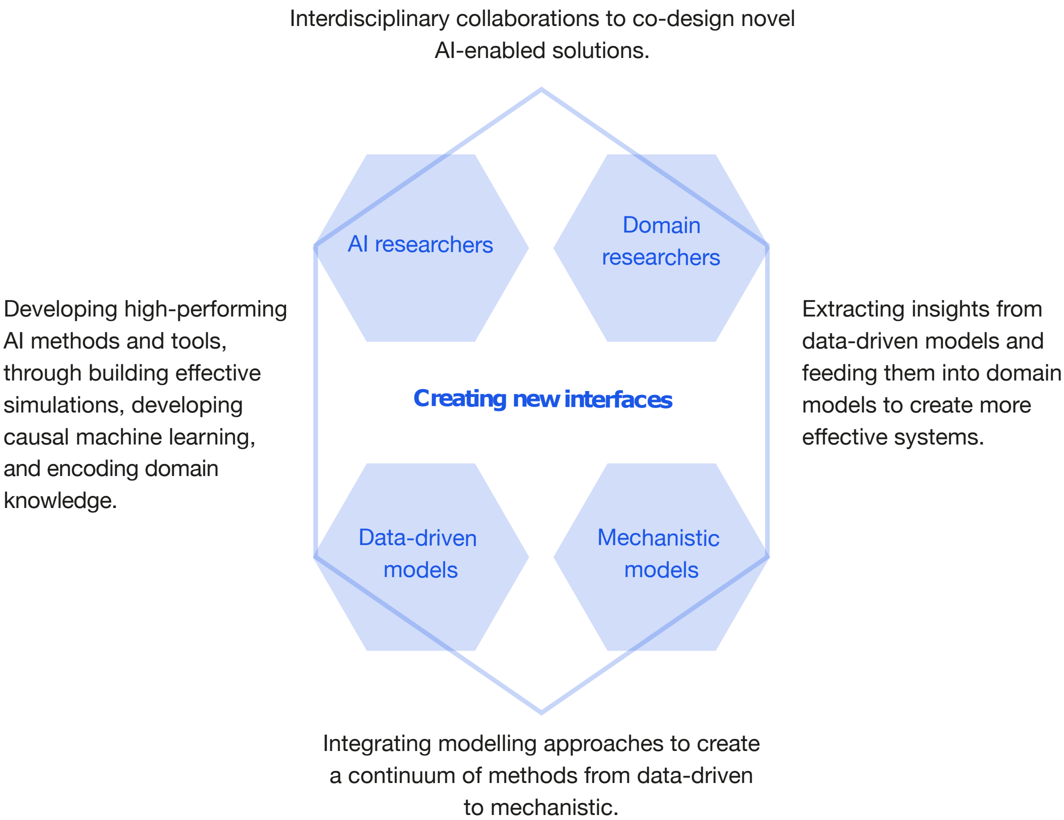


Figure 3 Interfaces between machine learning and domain researchers, and between data-driven and mechanistic models.

- Accelerating the pace of knowledge creation and use, through systems that mine the existing research knowledge base or that automate repetitive or time-consuming elements of the research process.

Influencing practice and adoption: By learning from recent experiences of deploying AI for science, the field has an opportunity to promote wider uptake and progress in both scientific domains and in AI research. This requires capturing both the knowledge that the community has already generated, about how to design AI systems, and the know-how about how to overcome practical challenges that accompanies it, while taking action to grow the community of researchers excited about the potential of AI in science. Areas for progress include:

- Supporting new applications, through challenge-led research programmes that promote interdisciplinary collaborations and support co-design of AI systems to help tackle scientific challenges.
- Developing toolkits and user guides that allow researchers to understand which AI tools are suitable for which purposes, and how to deploy those tools in practice.
- Sharing skills and know-how, through community outreach that disseminates knowledge and know-how in how to use AI.

Together, these areas for action highlight the importance of interfaces – between researchers and between modelling approaches – in shaping the development of AI for science (Figure 3).

13 Accelerating progress in AI for science

Building on the impressive advances that machine learning has already supported in many domains, widespread adoption of AI for research has the potential to catalyse a new wave of innovations that in turn could drive greater health, wealth, and wellbeing. The question facing researchers, funders, and policymakers today is how to harness that potential. The challenge is to build capability across the research landscape, connect areas of expertise to areas of need, and to accelerate the transfer of successful ideas between domains.

The experiences of deploying AI for science described in this document, and the research agenda that results from these experiences, suggest a roadmap for action. That roadmap charts a pathway to create an enabling environment for AI in science, by advancing research that delivers AI methods to support scientific discovery, building tools and resources to make AI accessible, championing interdisciplinary research and the people pursuing it, and nurturing a community at the interface of these different domains. Progress across these areas can unlock scientific and methodological advances in AI for science, while also helping answer an emerging question about whether there exists a core discipline of "AI for science". The shared themes and interests that emerge from research projects at the interface of AI and scientific domains suggest that there is potential for "AI for science" to surface as a distinct speciality in computer science. In parallel, domain-specific efforts to drive the adoption of AI as an enabler of innovation are also needed to deliver the benefits of AI for scientific discovery.

13.1 Advance new methods and applications

Efforts to deploy AI in the context of research have highlighted cross-cutting challenges where further progress in AI methods and theory is needed to create tools that can be used more reliably and effectively in the scientific context. Effective simulations are needed to study the dynamics of complex systems; causal methods to understand why those dynamics emerge; and integration of domain knowledge to relate those understandings to the wider world. While elements of these research challenges are shared with other fields – topics such as robustness, explainability, and human-machine interaction also come to the fore in fields such as AI ethics, for example – they share an intersection in the use of AI for science, in the context of efforts to bridge mechanistic and data-driven modelling.

Alongside these "AI" challenges are a collection of "science" challenges, where researchers, policymakers and publics have aspirations for AI to deliver real-world benefits.[35] Such challenges offer the opportunity to accelerate progress in AI, while facilitating interdisciplinary exchanges, and opening the field to input from citizen science or other public engagement initiatives. In developing these research missions, care is needed to define cross-cutting questions or challenges that broaden scientific imaginations, rather than restricting them. The process of converting a complicated scientific problem into something tractable with AI necessarily involves some narrowing of focus; to be successful, mission-led innovation efforts must achieve this focus without losing meaning, or creating benchmarks that misrepresent the complexity of the real-world challenge.

Defining shared challenges could help rally the AI for science community and drive progress

[35] See, for example: the EU's Innovation Missions `https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en` and UN SDG's `https://sdgs.un.org/goals`

in both methods and applications of AI in science. There already exists examples of how such challenges can build coalitions of researchers across domains from which the field can draw inspiration. These include the GREAT08 project, which developed image analysis techniques to study gravitational lensing [1]; the Open Problems in Single Cell Biology challenge, which convened the machine learning community to make progress in Multimodal Single-Cell Data Integration;[36] and the SENSORIUM challenge, focused on advancing understandings of how the brain processes visual inputs.[37] In pursuing this agenda, researchers can leverage well-established protocols in open-sourcing materials and sharing documentation to help ensure research advances are rapidly and effectively disseminated across disciplines. The result should be more effective methods, and an agile research environment where researchers can flex methods across disciplines.

13.2 Invest in tools and toolkits

Complementing these efforts to build and share knowledge, well-designed software tools can help make accessible the craft skills (or know-how) that make AI for science projects successful. Modelling is a core component of all AI for science projects. In some aspects, the task for the field can be thought of as charting a path between the statistician, whose effectiveness comes from proximity to the domain but whose methods struggle to scale, and the mathematician, whose tools are adopted across domains but with some loss of meaning as the distance between method-generator and adopter increases.

The energy already invested in building effective machine learning models can be leveraged for wider progress across domains through investment in toolkits that support the generalisation of effective approaches. Wide-spectrum modelling tools could offer "off the shelf" solutions to common AI for science research questions. The challenge for such toolkits is to create an effective interface between tool and user. Connecting with the field of human-computer interaction could generate design insights or protocols to help create more effective human-AI interfaces.

Best practices in software engineering can help, through documentation that supports users to successfully deploy modelling tools. User guides – or taxonomies of which models are best suited for which purposes and under what circumstances – can also help make accessible to non-expect users the accumulated know-how that machine learning researchers have gained through years of model development and deployment.

A related engineering challenge is that of data management and pipeline-building. To interrogate how a model works, why a result was achieved, or whether an AI system is working effectively, researchers often benefit from being able to track which data contributed to which output. The data management best practices that allow such tracking need to be embedded across AI for science projects. Data management frameworks – such as the FAIR data principles – have already been developed with the intention of making data more available, and useful, for research. Further investment is now needed in efforts to implement those principles in practice.

Investment in these foundational tools and resources can help build understanding of which AI methods can be used and for what purposes, lowering the barriers to adopting AI methods across disciplines.

[36] For further information, see: `https://openproblems.bio/neurips_2021/`
[37] For further information, see: `https://sensorium2022.net/home`

13.3 Build capability across disciplines

Central to progress in both research and toolkit engineering is the availability of talented researchers with a passion for advancing science through AI. People matter at all stages of the AI development and deployment pipeline. Successful projects rely on researchers who are motivated to work at the interface of different domains; collaborators who can explain and communicate core concepts in their work across disciplinary boundaries; engineers who can translate the needs of different users into AI toolkits; and convenors that can inspire wider engagement with the AI for science agenda.

Building these capabilities requires multiple points of engagement. Domain researchers need access to learning and development activities that allow them to understand and use foundational methods in machine learning, whether as formal training or through the availability of tutorials or user guides. AI researchers need access to the scientific knowledge that should shape the methods they develop, the skills to translate their advanced knowledge to materials that can be shared for wider use, and the capacity to dedicate time and resource to learning about domain needs.[38] Both need skills in communication, organisation, and convening to operate across disciplines. Without such capability-building, disciplines risk remaining siloed; domains developing unrealistic expectations about what AI can deliver in practice, and AI losing touch with the scientific questions that are most meaningful to domains.

Institutional incentives shape how individuals engage (or not) with such interdisciplinary exchanges. Interdisciplinary research often takes longer and lacks the outlets for recognition available to those working in single domains, affecting both the motivation of and opportunities for career progression that are open to those working at the interface of different disciplines. Much of the engineering work required to make data and AI accessible beyond a specific project and useful to a wider community is also traditionally unrecognised by academic incentive structures. Aligning individual and institutional incentives in support of interdisciplinarity is a long-standing challenge in research, and one that becomes more critical to address in the context of developments in AI. In this context, there may be new opportunities to recognise and reward successes in AI for science, whether through new fellowships, prizes, or ways of promoting the work done by those at this interface.

13.4 Grow communities of research and practice

The areas for action described above feed into and from each other. Progress in research and application can be leveraged to inspire a generation of researchers to pursue interdisciplinary projects; effective toolkits can make such progress more likely; skills-building initiatives can prime researchers to be able to use these toolkits; and so on, to create an environment where researchers and research advances transition smoothly across disciplines, leading to a rising AI tide that lifts all disciplines. Communities of research and practice are the backdrop for creating such positive feedback loops.

A collection of AI for science initiatives are already building links across the research landscape. The Machine Learning for Science Cluster of Excellence at the University of Tübingen is leveraging the strength of its local ecosystem in AI to drive wider progress in

[38] A comparison here can be drawn with the development of statistics as an enabling discipline for many domains: statisticians have devoted time to understanding domain practices and integrating their work within those practices, often dedicating significant resource to understand the nature of the datasets with which they are working, before introducing modelling ideas.

research and innovation;[39] the Accelerate Programme for Scientific Discovery at the University of Cambridge is building bridges across disciplines, building a community passionate about opportunities in AI for science;[40] the University of Copenhagen's SCIENCE AI Centre provides a focal point for AI research and education in its Faculty for Science;[41] New York University's Center for Data Science hosts interdisciplinary faculty pursuing innovative research and education;[42] the University of Wisconsin-Madison's American Family Insurance Data Science Institute is developing strategic partnerships to accelerate the use of data science in research;[43] new investments by Schmidt Futures across a network of research institutions are supporting new postdoctoral fellowships at the interface of AI and sciences [2]. Together, these initiatives demonstrate the appetite for progress in AI for science.

There is an opportunity today to leverage these emerging interests into a wider movement. Existing initiatives can drive capability-building, by making training and user guides open, reaching out to engage domain researchers in skills-building activities, and fostering best practice in software and data engineering across disciplines. The links they establish across research domains can form the basis of new communication channels, whether through discussion forums, research symposia, or newsletters to share developments at the interface of AI and science. These communications can be deployed to raise the profile of people and projects at this interface, celebrating successes, sharing lessons, and demonstrating the value of interdisciplinary work. Together, they can help develop an infrastructure for AI in science.

That infrastructure may also benefit from new institutional interventions to address long-standing challenges in interdisciplinary AI. New journals could provide an outlet to publish and recognise high-quality AI for science research, bringing in contributions from multiple disciplines and helping translate lessons across areas of work. Membership organisations could help foster a sense of belonging and community for researchers working at the interface of AI, science, and engineering, developing career pathways and incentives. Efforts to convene across disciplines can also catalyse new connections and collaborations.

Emerging from these efforts is a paradigm shift in how to drive progress in science. Historically, a small number of foundational texts have been the catalyst that changed how researchers studied the world; Newton's Principia; Darwin's Origin of Species; and so on. For much of its modern history, scientific knowledge has been transmitted through textbooks; canonical descriptions of the current state of knowledge. Today, the transformative potential of AI is driven by its pervasiveness; its impact in science will be achieved through integration across disciplines. This integration requires widespread mobilisation, convening machine learning researchers, domain experts, citizen scientists, and affected communities to shape how AI technologies are developed and create an amenable environment for their deployment. It takes a community.

[39] Programme website available at: `https://uni-tuebingen.de/en/research/core-research/cluster-of-excellence-machine-learning/home/`.
[40] Programme website available at: `https://acceleratescience.github.io`.
[41] Programme website available at: `https://ai.ku.dk`.
[42] Programme website available at: `https://cds.nyu.edu`.
[43] Programme website available at: `https://datascience.wisc.edu/institute/`.

13.5 AI and science: building the interface

Advances in AI have disrupted traditional ways of thinking about modelling in science. Where researchers might previously have conceptualised models as mechanistic – reflecting known forces in the world – or data-driven, the "AI for science" methods that are emerging today reject this separation. They are both, combining insights from mechanistic and data-driven methods, integrating methods to create something new. What follows from these developments is a spectrum of modelling approaches, which researchers can deploy flexibly in response to the research question of interest.

Today, the field of AI for science is characterised by intersections. Between AI and scientific domains; between science and engineering; between knowledge and know-how; between human and machine. It operates across disciplinary boundaries, across scales from the atomic to the universal, and across both the mission to understand intelligence and the quest to deploy human intelligence to understand the world. Emerging from these missions is a continuum of models and methods that allow researchers to work across domains, extracting the knowledge that humans have acquired, and levels of inquiry, enhancing that knowledge and returning it in actionable form.

As both a domain itself and an enabler of other disciplines, the power of AI in science lies in its ability to convene diverse perspectives in ways that accelerates progress across research areas. AI for science is a rendezvous point. Its next wave of development will come from taking strength from its diversity, and bringing more people into its community.

Acknowledgments

The Accelerate Programme for Scientific Discovery would like to thank Schmidt Futures for its continuing support, and the donation that enables its work.

References

1 Sarah Bridle, Sreekumar T Balan, Matthias Bethge, Marc Gentile, Stefan Harmeling, Catherine Heymans, Michael Hirsch, Reshad Hosseini, Mike Jarvis, Donnacha Kirk, et al. Results of the great08 challenge: an image analysis competition for cosmological lensing. *Monthly Notices of the Royal Astronomical Society*, 405(3):2044–2061, 2010. doi:10.1111/j.1365-2966.2010.16598.x.

2 Schmidt Futures. Schmidt futures launches $148m global initiative to accelerate ai use in postdoctoral research. Available at https://www.schmidtfutures.com/schmidt-futures-launches-148m-global-initiative-to-accelerate-ai-use-in-\postdoctoral-research/ , 2022.

A Research questions arising from the "AI for science research agenda" discussion during the Dagstuhl Seminar

Building AI systems for science

- How can AI systems accurately generalise from finite observations? How can they detect causality or structure from finite observations?
- What is the computational cost of complexity, and what methods can help manage this?
- What forms of system calibration and uncertainty quantification are useful in the context of scientific discovery? Are theoretical guarantees necessary?

- What new forms of explainability or interpretability could facilitate the deployment of AI in science?
- How could AI support generalisation from a small number of observations? What methods could enable few- or one-shot learning?
- How can AI researchers build meaningful models from data to accurately represent causal mechanisms in the system of study? How can researchers identify the most effective model for their system of study?
- What does it mean to understand a model? How can researchers combine explainability with complexity?
- How can AI methods be made robust and easy to use in deployment by domain scientists?
- How can advances in simulation methods be applied in domains where the system at hand is less easily described by equations?
- What advances are needed to expand the use of simulations in science? How can AI help simulate laboratory experiments or environments, helping make more efficient different elements of the scientific process? How might this be expanding in the long-term, for example to planning experimental design or helping identify where data is missing?
- How can "digital siblings" be used to explore the impact of different interventions on complex systems?

Combining human and machine intelligence

- How can AI researchers best extract, formalise and assimilate the knowledge that domain researchers have acquired? What forms of knowledge representation can formalise scientific understandings of the world, translating these to objective functions for AI systems? What forms of human-AI engagement can make use of the "qualitative" knowledge – or intuitions about a system – that domain researchers have accumulated?
- How can AI capture the qualitative understanding that researchers have of their domain to more accurately or effectively characterise a system?
- How can AI be effectively deployed to mine the existing research knowledge base – for example, papers, databases, and so on – to extract new insights?
- Where can automation support research progress? Which elements of the scientific process could be automated, and where is human input vital?
- What forms of collaboration are needed to effectively specify helpful outputs from an AI system?
- How can insights from AI analysis be returned to researchers in an actionable way? What mix of AI design, engineering, social interaction, and education can make effective interfaces between domain researchers and AI systems?
- How can the outputs of AI systems be made interpretable for scientific users?
- How can AI researchers better understand and design for the forms of interpretability that resonate with domain researchers?
- What processes of collaboration or co-design can help describe what scientists "need to know" from an AI system?
- What best practices or methods can be deployed to effectively communicate uncertainty from AI systems to human users?

A.1 Influencing practice and adoption

- What are the craft skills in AI for science? What “know-how” is necessary to make AI work effectively in practice?
- What skills-building or forms of outreach can help take AI tools out of the AI community and into “the lab”?
- How has machine learning been used most effectively for research and innovation? What best practices, or lessons, do existing efforts in AI for science offer?
- Which AI tools are suitable for which purposes, disciplines, or experimental designs? Is it possible to create a taxonomy for science?
- Are there generalisable methods or conclusions that can be taken from domain-specific efforts to deploy AI for science?

Participants

- Mauricio A Álvarez
University of Manchester, GB
- Bubacarr Bah
AIMS South Africa –
Cape Town, ZA
- Jessica Beasley
Collective Next – Boston, US
- Philipp Berens
Universität Tübingen, DE
- Maren Büttner
Helmholtz Zentrum München, DE & Universität Bonn, DE
- Kyle Cranmer
University of Wisconsin –
Madison, US
- Thomas G. Dietterich
Oregon State University –
Corvallis, US
- Carl Henrik Ek
University of Cambridge, GB
- Stuart Feldman
Schmidt Futures – New York, US
- Asja Fischer
Ruhr-Universität Bochum, DE
- Philipp Hennig
Universität Tübingen, DE
- David W. Hogg
New York University, US
- Christian Igel
University of Copenhagen, DK
- Samuel Kaski
Aalto University, FI
- Ieva Kazlauskaite
University of Cambridge, GB
- Hans Kersting
INRIA – Paris, FR
- Niki Kilbertus
TU München, DE & Helmholtz AI München, DE
- Vidhi Lalchand
University of Cambridge, GB
- Neil D. Lawrence
University of Cambridge, GB
- Gilles Louppe
University of Liège, BE
- Dina Machuve
DevData Analytics – Arusha, TZ
- Jakob Macke
Universität Tübingen, DE
- Eric Meissner
University of Cambridge, GB
- Siddharth Mishra-Sharma
MIT – Cambridge, US
- Jessica Montgomery
University of Cambridge, GB
- Jonas Peters
University of Copenhagen, DK
- Aditya Ravuri
University of Cambridge, GB
- Markus Reichstein
MPI für Biogeochemistry –
Jena, DE
- Bernhard Schölkopf
MPI für Intelligente Systeme –
Tübingen, DE
- Francisco Vargas
University of Cambridge, GB
- Soledad Villar
Johns Hopkins University –
Baltimore, US
- Ulrike von Luxburg
Universität Tübingen, DE
- Verena Wolf
Universität des Saarlandes –
Saarbrücken, DE

Report from Dagstuhl Seminar 22391

Cognitive Robotics

Fredrik Heintz*[1], Gerhard Lakemeyer*[2], and Sheila McIlraith*[3]

1 **Linköping University, SE.** `fredrik.heintz@liu.se`
2 **RWTH Aachen University, DE.** `gerhard@kbsg.rwth-aachen.de`
3 **University of Toronto, CA.** `sheila@cs.toronto.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22391 on the topic of "Cognitive Robotics". Cognitive Robotics is concerned with endowing robots or software agents with higher level cognitive functions that involve reasoning, for example, about goals, perception, actions, the mental states of other agents, and collaborative task execution. The seminar is the latest event in a series of events on this topic that were initiated in 1998. With its roots in knowledge representation and reasoning, the program for this seminar was influenced by transformative advances in machine learning and deep learning, by recent advances in human-robot interactions, and by issues that arise in the development of trustworthy cognitive robotic systems. Reflective of this, the seminar featured sessions devoted to the following four themes: cognitive robotics and KR, verification of cognitive robots, human-robot interaction and robot ethics, and planning and learning. Each theme consisted of plenary talks, plenary discussions and working groups resulting in a research road map for the coming years. There was also a poster session where new or published results could be presented by the participants.

The seminar was very successful and well received by the participants thanks to the excellent environment for exchanging ideas provided by Schloss Dagstuhl.

Seminar September 25–30, 2022 – http://www.dagstuhl.de/22391

2012 ACM Subject Classification Computing methodologies → Cognitive robotics; Computing methodologies → Planning and scheduling; Computing methodologies → Machine learning; Human-centered computing → Human computer interaction (HCI); Computer systems organization → Robotics

Keywords and phrases Artificial Intelligence, Knowledge Representation and Reasoning, Cognitive Robotics, Verification, Human-robot Interaction, Robot Ethics, Machine Learning, Planning

Digital Object Identifier 10.4230/DagRep.12.9.200

1 Executive Summary

Fredrik Heintz (Linköping University, SE, `fredrik.heintz@liu.se`*)*
Gerhard Lakemeyer (RWTH Aachen University, DE, `gerhard@kbsg.rwth-aachen.de`*)*
Sheila McIlraith (University of Toronto, CA, `sheila@cs.toronto.edu`*)*

Cognitive Robotics is concerned with endowing robots or software agents with higher level cognitive functions that involve reasoning, for example, about goals, perception, actions, the mental states of other agents, collaborative task execution, etc. This research agenda has historically been pursued by describing, in a language suitable for automated reasoning, enough of the properties of the robot, its abilities, and its environment, to permit it to make high-level decisions about how to act. Such properties were typically encoded by a human, but with recent advances in machine learning, many of these properties, and the

* Editor / Organizer

determination of how to act, can be learned or adapted through experience. This in turn raises the question of how we can ensure that robots, or other intelligent agents, can be constructed in a manner that is compatible with human values and modes of interactions.

The Cognitive Robotics workshop series has been running since 1998 and includes a Dagstuhl Seminar held in 2010. While progress in Cognitive Robotics has undoubtedly been made over the past twenty years, it is fair to say that we are still far away from creating truly cognitive robots. In particular, the years since the previous Dagstuhl Seminar have seen tremendous progress in many areas that touch on the realisation of cognitive robots such as advances in human-robot interaction and machine learning.

This seminar featured sessions devoted to the following four themes:

Cognitive Robotics and KR: While knowledge representation and reasoning (KR) has played a role in robotic systems for many years, for example, by incorporating domain knowledge in the form of description logic-based ontologies or using automated planning systems for high-level robot control, obstacles remain, which prevent today's robots from benefiting from the true potential of KR. In this session we re-visited the state of the art of how KR is used in robotics and discussed challenges and possible benchmark problems that would demonstrate the need and benefit of KR techniques for cognitive robots. The session was organized by Michael Beetz, University of Bremen.

Verification of Cognitive Robots: Verification has been an active research area in formal methods for many years. It is also an important topic when it comes to cognitive robots, especially when it comes to achieving trustworthiness. However, the sheer complexity of the interplay between a robot's hard- and software components makes verification particularly challenging. In this session we discussed where we currently stand in terms of verifying cognitive robots and what challenges lie ahead. The session was organized by Michael Fisher, University of Manchester.

Human-robot Interaction and Robot Ethics: For cognitive robots to be useful in human environments, effective human-robot interaction (HRI) plays a crucial role. Besides the technological challenges such as multi-modal communication, ethical considerations have become more and more important. These range from robots observing norms and conventions to humans viewing robots as moral agents. In this session we discussed the many facets of robot ethics in the context of HRI and identified a number of future challenges and open problems. The session was organized by Matthias Scheutz, Tufts University.

Planning and Learning: While planning and learning have traditionally been separate research tracks in cognitive robotics, recent work has shown how action primitives that form the basis of planning can be learned from data without background knowledge, thus avoiding the need for hand-crafted solutions. In this sessions this work and related proposals were discussed and a roadmap with short- and long-term challenges was drawn up. The session was organized by Hector Geffner, ICREA and Universitat Pompeu Fabra, Spain. The format of the sessions varied and consisted of one or more plenary talks, plenary discussions and/or working groups. Working groups for all four themes discussed challenges and roadmaps for the future, and one representative of each group presented their findings on the last day of the seminar. Besides talks and discussions that centered around the four themes, the seminar also featured two invited talks by Luis Lamb, Universidade Federal Do Rio Grande Do Sul, on neurosymbolic AI and by Jan Peters, TU Darmstadt, on robot learning. In addition, a number of participants gave poster presentations on their research.

The organizers of the seminar wish to thank Schloss Dagstuhl for providing such an excellent environment for exchanging ideas on how to move the field of cognitive robotics forward.

2 Table of Contents

3 Overview of Talks

3.1 Knowledge Representation and Reasoning for Cognition-enabled Robot Manipulation

Michael Beetz (Universität Bremen, DE)

Robotic agents that can accomplish manipulation tasks with the competence of humans have been the holy grail for AI and robotics research for more than 50 years. However, while the fields made huge progress over the years, this ultimate goal is still out of reach. I believe that this is the case because the knowledge representation and reasoning methods that have been proposed in AI so far are necessary but still too abstract. In this talk I propose to endow robots with the capability to mentally "reason with their eyes and hands," that is to internally emulate and simulate their perception-action loops based on photo-realistic images and faithful physics simulations, which are made machine-understandable by casting them as virtual symbolic knowledge bases. These capabilities allow robots to generate huge collections of machine-understandable manipulation experiences, which they can then generalize into commonsense and intuitive physics knowledge applicable to open manipulation task domains. The combination of learning, representation, and reasoning will equip robots with an understanding of the relation between their motions and the physical effects they cause at an unprecedented level of realism, depth, and breadth, and enable them to master human-scale manipulation tasks. This breakthrough will be achievable by combining simulation and visual rendering technologies with mechanisms to semantically interpret internal simulation data structures and processes.

3.2 Online Replanning with Human-in-The-Loop for Non-Prehensile Manipulation in Clutter – A Trajectory Optimization based Approach

Tony Cohn (University of Leeds, GB)

Joint work of RafaelPapalla, Anthony G. Cohn, Mehmet R. Dogar

We are interested in the problem where a number of robots, in parallel, are trying to solve reaching through clutter problems in a simulated warehouse setting. In such a setting, we investigate the performance increase that can be achieved by using a human-in-the-loop providing guidance to robot planners. These manipulation problems are challenging for autonomous planners as they have to search for a solution in a high- dimensional space. In addition, physics simulators suffer from the uncertainty problem where a valid trajectory in simulation can be invalid when executing the trajectory in the real-world. To tackle these problems, we propose an online-replanning method with a human-in-the-loop. This system enables a robot to plan and execute a trajectory autonomously, but also to seek high-level suggestions from a human operator if required at any point during execution. This method aims to minimize the human effort required, thereby increasing the number of robots that can be guided in parallel by a single human operator. We performed experiments in simulation and on a real robot, using an experienced and a novice operator. Our results show a significant increase in performance when using our approach in a simulated warehouse scenario and six robots.

3.3 Joint Perceptual Learning and Natural Language Acquisition for Autonomous Robots

Tony Cohn (University of Leeds, GB)

Joint work of Muhannad Alomari, Fangjun Li, David C. Hogg, Anthony G. Cohn

In this work, the problem of bootstrapping knowledge in language and vision for autonomous robots is addressed through novel techniques in grammar induction and word grounding to the perceptual world. In particular, we demonstrate a system, called OLAV, which is able, for the first time, to (1) learn to form discrete concepts from sensory data; (2) ground language (n-grams) to these concepts; (3) induce a grammar for the language being used to describe the perceptual world; and moreover to do all this incrementally, without storing all previous data. The learning is achieved in a loosely-supervised manner from raw linguistic and visual data. Moreover, the learnt model is transparent, rather than a black-box model and is thus open to human inspection. The visual data is collected using three different robotic platforms deployed in real-world and simulated environments and equipped with different sensing modalities, while the linguistic data is collected using online crowdsourcing tools and volunteers. The analysis performed on these robots demonstrates the effectiveness of the framework in learning visual concepts, language groundings and grammatical structure in these three online settings.

3.4 Verifying Autonomous Systems

Michael Fisher (University of Manchester, GB)

Autonomy represents a step-change in systems development and requires new approaches to system architectures, to systems analysis and to effective usage.

In this presentation, I describe an approach that utilises the modularity and heterogeneity of (robotic) software architectures to provide a hybrid agent architecture. Then, a range of verification techniques can be applied to the different components, from formal verification applied to the core autonomous decision-making through to varieties of testing used in other parts of the system.

Finally, an important component is the use of runtime verification (or runtime monitoring) to check for anomolies and violations. Together, these mechanisms provide a basis for more relaible, transparent, trustworthy and verifiable autonomous systems.

3.5 Top-down Representation Learning for Acting and Planning

Hector Geffner (ICREA and Universitat Pompeu Fabra, ES)

Recent breakthroughs in AI have shown the remarkable power of deep learning and deep reinforcement learning. These developments, however, have been tied to specific tasks, and progress in out-of-distribution generalization has been limited. While it is assumed that these limitations can be overcome by incorporating suitable inductive biases in neural nets, this is left vague and informal, and does not provide meaningful guidance. In this talk, I articulate a different learning approach where representations are learned over domain-independent target languages whose structure and semantics yield a meaningful and strongly biased hypothesis space. The learned representations do not emerge then from biases in a low level architecture but from a general preference for the simplest hypothesis that explain the data. I illustrate this general idea by considering three learning problems in AI planning: learning general actions models, learning general policies, and learning general subgoal structures ("intrinsic rewards"). In all these cases, learning is formulated and solved as a combinatorial optimization problem although nothing prevents the use of deep learning techniques instead. Indeed, learning representations over domain-independent languages with a known structure and semantics provides an account of what is to be learned, while learning representations with neural nets provides a complementary account of how representations can be learned. The challenge and the opportunity is to bring the two approaches together.

3.6 Better Autonomy Through Uncertainty

Nick Hawes (Oxford University, GB)

Due to the challenges of perception and action, and inevitable inaccuracies in world modelling, the results of a robot's interactions with its environment are inherently stochastic. To successfully complete extended missions under such conditions it is therefore essential that autonomous robots use techniques from decision-making under uncertainty to plan goal-directed behaviour. In this talk I will give an overview of our recent work on planning under uncertainty for autonomous robots, drawing examples from mobile service robots, underwater vehicles, and quadrupeds.

3.7 Cognitive Robotics – A KR Perspective

Gerhard Lakemeyer (RWTH Aachen University, DE)

In this overview talk I address some of the main representation and reasoning techniques that have been used in robotic systems. On the representation side, these include simple databases (logical literals), description logics, and geometric or topological maps with semantic annotations. On the reasoning side, we find methods for temporal, spatial, and uncertainty reasoning as well as automated planning techniques. I also touch upon the need for execution monitoring and failure diagnosis. At the end of my talk I briefly introduce the RoboCup Logistics League, where robots interact with machines in a production logistics scenario and which can serve as a benchmark for applying KR in robotics, both in simulation and on real robots.

3.8 Learning and Reasoning in Neurosymbolic AI

Luis Lamb (Universidade Federal Do Rio Grande Do Sul, BR))

Neurosymbolic AI aims to bring together the statistical nature of machine learning and the logical essence of reasoning in AI systems. Recently, leading technology companies and research groups have put forward agendas for the development of the field, as modern AI systems require sound reasoning and improved explainability. In this talk, we highlight Neurosymbolic AI research results that led to applications and novel developments towards building richer AI systems. We summarize how the field evolved over the years and how it can potentially contribute to improved explainability and the effective integration of learning and reasoning in robust AI.

3.9 Learning Grounded Language for Human Interaction

Cynthia Matuszek (University of Maryland, Baltimore County, US)

Robots deployed today largely perform a predefined set of tasks in limited, controlled environments. In order to handle the complexity of human-centric spaces, it is necessary to learn about the world and tasks from human end users, and natural language is a key modality for such learning. Two high level approaches to understanding and learning from such language are, first, learning probabilistic grammars describing the perceptual state of the world and, second, learning directly from speech, without any textual intermediary. This talk describes work on using a combination of language and perceptual data to learn about how people describe objects in the world, with the long-term goal of understanding tasks and instructions presented in natural language by non-specialist end users. The importance of using speech directly is discussed, and the effectiveness of using featurized speech is compared to ASR-based approaches. Using speech not only improves performance on the language grounding task, but also reduces performance differences among different demographic groups, leading to more immediately deployable robotic systems.

3.10 Reward Machines: Formal Languages and Automata for Reinforcement Learning

Sheila McIlraith (University of Toronto, CA)

Reinforcement Learning (RL) is proving to be a powerful technique for building sequential decision-making systems in cases where the complexity of the underlying environment is difficult to model. Two challenges that face RL are reward specification and sample complexity. Specification of a reward function – a mapping from state to numeric value – can be challenging, particularly when reward-worthy behaviour is complex and temporally extended. Further, when reward is sparse, it can require millions of exploratory episodes for an RL agent to converge to a reasonable quality policy. In this talk I'll show how formal languages and automata can be used to represent complex non-Markovian reward functions. I'll present the notion of a Reward Machine, an automata-based structure that provides a normal form representation for reward functions, exposing function structure in a manner that greatly expedites learning. Finally, I'll also show how these machines can be generated via symbolic planning or learned from data, solving (deep) RL problems that otherwise could not be solved.

3.11 Model Learning for Plannning

Christian Muise (Queens University – Kingston, CA)

Model learning can primarily be characterized across three dimensions: (1) the input data format; (2) the output model components; and (3) the priors/partial models that we start with. Here, we explore two settings where model learning for planning has been studied.

First, we detail the Model Acquisition Toolkit (MACQ): a library dedicated to learning action theories from state traces of various forms. Each technique in the library comes with its own priors, but collectively the library provides the most comprehensive treatment to date of extracting action theories from discrete time series data.

The second work explores how strong priors influenced by planning concepts can aid in learning planning models from image pairs alone. By embedding strong notions of action representation into the learning architecture itself, we are able to learn action theories and state representations that can be given to off-the-shelf planners.

These are but two modern examples of how model learning is being explored in the context of planning.

3.12 Hardware Acceleration: Why, What, How, Use Cases?

Bernhard Nebel (Universität Freiburg, DE)

This talk does not report on research results, but rather on perspectives of how hardware acceleration can be exploited for automatic planning. Focusing on RPG-style heuristics, it is sketched how such heuristics estimators can be compiled into sequential circuits for moderately large planning tasks, which opens up the possibility to implement that on standard FPGAs. Since 80-90% of the compute time in planning systems is spent on computing heuristic estimates, this could result in a speedup of one order of magnitude.

3.13 Robot Learning: Quo Vadis?

Jan Peters (TU Darmstadt, DE)

Autonomous robots that can assist humans in situations of daily life have been a long standing vision of robotics, artificial intelligence, and cognitive sciences. A first step towards this goal is to create robots that can learn tasks triggered by environmental context or higher level instruction. However, learning techniques have yet to live up to this promise as only few methods manage to scale to high-dimensional manipulator or humanoid robots. In this talk, we investigate the challenges for robot learning from both the symbolic and subsymbolic perspective! We show how symbols can arise in a robot learning system and can used to further the general application of robot learning. We also discuss how classically disjunct approaches from first order insight can be used as inductive biases for faster learning using the simulation based approach. We describe the work in various robotic scenarios ranging from tactile manipulation to robot juggling.

3.14 HRI and Robot Ethics

Matthias Scheutz (Tufts University – Medford, US)

Robot ethics is no different from bioethics, information ethics, environmental ethics, etc. in that as a technology it has impact on humans human societies. It is different from all other technologies in that AI enables the development and deployment of autonomous systems that perceive their environment and determine their actions without human ado. AI/robot ethics thus raises the question of whether these systems can operate in human societies and interact with humans in a way that is ethical and acceptable to humans, not causing any harm. For this, robots need to be able to learn human norms from observations and instructions and follow them. When norm conflicts arise, they need to be able to determine the best course of action and justify their choices by appealing to principles used for their decisions. How to build a robotic architecture capable of all of this is the main challenge of ethical HRI!

3.15 Active Learning in Risky Environments: Exploring Deep-Sea Volcanoes and Ocean Worlds

Brian Williams (MIT – Cambridge, US)

Machine learning is a powerhouse in information rich environments. However, machine learning remains challenging when data is sparse, is costly to collect, and is dangerous and complex to acquire. As two examples, ocean exploration and subsea inspection use autonomous vehicles to perform information gathering, to answer questions about the environment. In these applications, communication is limited, vehicles need to be autonomous, environments are risky, and resources are constrained.

Our vision is to create systems that answer information queries by performing active learning in risky environments. These systems 1) generate information gathering plans that bound risk, while maximizing information with respect to a set of questions being asked, 2) continuously adapt plans based on what is observed and what remains unanswered and 3) incorporates informative measures and risk within operational plans, at multiple levels of abstraction.

The talk introduces a series of model-based agent programming paradigms that support this process of active learning in risky environments, starting with state and decision-theoretic programming. The talk then focuses on planning and learning methods that are needed to support two new programming paradigms – information theoretic and risk-aware programming. These approaches are demonstrated in the context of a 2019 ocean campaign, to explore the Columbo volcano in the Mediterranean Sea.

4 Poster Presentations

All participants provided a brief oral introduction and an overview of their research at the outset of the 5-day seminar. All participants were also given the opportunity to present their work in a poster session. The following is a list of participants who presented their research as posters.

Mohamed Behery and Gerhard Lakemeyer

Poster: *Assistive Robot Teleoperation Using Phase Switching Behavior Trees*
Authors: Mohamed Behery, Minh Trinh, Christian Brecher, Gerhard Lakemeyer
Related Publications: Not published yet.

Anthony G Cohn

Poster: *A framework for categorising AI evaluation instruments*
Authors: A G Cohn, José Hernández-Orallo, Julius Sechang Mboli, Yael Moros-Daval, Zhiliang Xiang, Lexin Zhou
Related Publications: `https://ceur-ws.org/Vol-3169/paper3.pdf`

Jasmin Grosinger

Poster: *Proactivity*
Author: Jasmin Grosinger
Related Publications: Not published yet.

Till Hofmann and Gerhard Lakemeyer

Poster: *Controlling Golog Programs against MTL Constraints*
Authors: Till Hofmann, Stefan Schupp, Gerhard Lakemeyer
Related Publications: Not published yet.

Mikhail Khodak

Poster: *Learning Algorithms and Learning Algorithms*
Authors: Mikhail Khodak
Related Publications: listed at the bottom of the poster.

Sven Koenig

Poster: *Multi-Agent Path Finding (MAPF) and Its Applications*
Authors: Many, as listed on poster
Related Publications: `http://idm-lab.org/project-p.html`

Yves Lespérance

Poster: *Plan Recognition in a High Level Belief-Based Programming Language*
Authors: Yves Lespérance, Alistair Scheuhammer, Yu Chen, and Petros Faloutsos

Setareh Maghsudi

Poster: *Multi-Agent Reinforcement Learning*
Authors: Setareh Maghsudi
Related Publications: based on several publications as listed in the poster,

Sheila McIlraith

Poster: *LTL and Beyond: Formal Languages for Reward Function Specification in Reinforcement Learning*
Authors: Alberto Camacho, Rodrigo Toro Icarte, Toryn Q. Klassen, Richard Valenzano, Sheila A. McIlraith
Related Publications: based on several publications as listed in the poster.

Bernhard Nebel

Poster: *The Complexity of MAPF on Directed Graphs & The Small Solution Hypothesis*
Authors: Bernhard Nebel
Related Publications: *The Small Solution Hypothesis for MAPF on Strongly Connected Directed Graphs is True*, arXiv:2210.04590.

Maayan Shvo

Poster: *Proactive Robotic Assistance via Theory of Mind*
Authors: Maayan Shvo, Ruthrash Hari, Ziggy O'Reilly, Sophia Abolore, Nina Wang, Sheila A. McIlraith
Related Publications: *Proactive Robotic Assistance via Theory of Mind*, IROS 2022.

5 Working groups

The seminar focused on four themes central to cognitive robotics, with one expert among the participants organizing a session around each theme: cognitive robotics and KR (Michael Beetz), verification of cognitive robots (Michael Fisher), HRI and robot ethics (Matthias Scheutz), and planning and learning (Hector Geffner). The format of the sessions varied and consisted of one or more plenary talks, plenary discussions and/or working groups. Working groups for all four themes discussed challenges and roadmaps for the future, and one representative of each group presented their findings on the last day of the seminar: Gerhard Lakemeyer (cognitive robotics and KR), Fredrik Heintz (verification of cognitive robots), Cynthia Matuszek (HRI and robot ethics), Christian Muise (planning and learning). Here is a summary.

5.1 Cognitive Robotics and KR

Knowledge Representation and Reasoning (KR) has been a concern in cognitive robotics for many years, beginning with the robot *Shakey* developed at SRI in the late sixties. While ontological knowledge, formalized using description logics, and automated planning systems, among other things, can be found in many robotic applications, KR has yet to play a central role in building cognitive robots. In this working group, we discussed and collected some of the challenges that remain in order to leverage the true potential of KR for cognitive robotics. The following lists the main findings and recommendations.

5.1.1 Challenges

- How does a robot know when system 2 is needed (meta cognition)?
- Finding suitable open-ended robotic tasks that demonstrate the need for KR.
- Industrial use cases, where humans and robots collaborate during production (issues in planning, HRI).
- Addressing problems with long-tail phenomena, which are best solved with commonsense.
- How to acquire commonsense for specific tasks.
- Standardization of KR formalisms would help with the uptake (as has happened with OWL).
- Creating a NELL (lifelong learned KB) for robots.
- How to control the complexity of a task? Compilation techniques?
- How can a robot be taught like a human or, how to transfer conceptual representations of a human to a robot?
- How to build a system that can perform a task after watching a video that shows how to do it. How to do it with tools different from those in the video.

- How to build systems that can introspect on their own actions and explain what they are doing.
- How to build robots with a theory of mind (going beyond traditional BDI, which does not consider action, perception, failures, uncertainty).

5.1.2 Reasons why KR is not yet central to robotics and possible ways to overcome this

- When working with robots, 90% of the time is spent on things other than KR. For roboticist, KR issues are often an afterthought, while KR people cannot grapple with the complexity of robots.
- KR for robotics is lacking a “playground” such as benchmarks suitable for testing/evaluating implemented systems. (Attempts like RoboCup Logistics in simulation were not taken up by the planning people because of the complexity, see also Multi-Agent Programming)
- Appropriate environments need to be developed (RoboCup?)
- Those need to be spread and advertized via tutorials at the KR and ICAPS conferences.
- Similarly, KR tools need to be created for use by roboticists.

5.1.3 Roadmap (5–10 years)

- Principled approaches to abstraction of perception.
- Goal reasoning for robots.
- Rationalizing existing implemented KR systems like KnowRob.
- A theory of explainable behavior and its realization in cognitive robots.
- A theory of mind for robots.

5.2 Verification of Cognitive Robots

Verification and validation of complex cognitive robots is very challenging and existing methods, mainly from formal verification, can only be applied to relatively simple cases. This section summarizes the challenges, directions for future research and provides a roadmap towards verification of cognitive robots.

5.2.1 Challenges

- Correct-by-design
- End-to-end verification
- Composing verified components into verified systems
- Combining partial/abstract offline verification with complete/detailed online verification
- Minimum assumption verification, combined with a risk model to assess the risk involved in the assumptions, combined with runtime verification of the assumption to get the minimum risk system
- Systematic combination of partial verification and testing, verify those parts that can be verified, and then systematically test the rest
- Verify models that are used by for example solvers
- Understand the limitations of what is verifiable
- How to build systems that can be verified? What architectures enables verification?

- Verifying learning systems
- Verifying systems that interact with people
- Continuous (online) verification of learning and interacting systems

5.2.2 General Direction of Development

- From static deterministic simple environments to dynamic non-deterministic complex adversarial environments
- From one-off large-scale efforts for verifying particular components to systematic methods for verifying components to tools that automate the verification of components
- From simple components to complex components to simple static systems to complex dynamic systems of components to open, dynamic and learning systems-of-systems

5.2.3 Roadmap

- 5-years
  - Develop verified plan verifiers that can verify plan instances
  - Verified solvers, such as planners, which are guaranteed to generate verified solutions
  - Verified skills under (potentially strong) assumptions about sensors and external behaviors
  - Principled combination of testing, off-line verification and on-line verification of static systems
  - Early involvement with regulators to jointly agree on what to verify, also related to translating high-level abstract properties into things that can be quantified and (probabilistically) verified
- 10-years
  - A formal understanding of what can be verified
  - Methods for formally verifying solvers
  - Methods for verifying skills (and other robot behaviors) and reducing the assumptions under which these are guaranteed to work
  - Verified (simple) cognitive robots using (verified) solvers and (verified) skills to achieve non-trivial goals
  - Principled (off-line/on-line) verification of (simple) cognitive robots that improve their behavior over time (learning)

5.3 HRI and Robot Ethics

Human-robot interaction (HRI) with its many facets and interdisciplinary nature is of key importance for cognitive robotics, with ethical concerns playing an important role as well. In this working group challenges for HRI and robot ethics were discussed and collected along several dimensions: humans modeling robots and vice versa, norms, communication and information flow, and proactive behavior. In the following, we summarize our main findings.

5.3.1 Humans modeling robots

- How can we build systems where it is possible for people to have an accurate model of the robot's capabilities and internal state?
- Possibly we will always interpret its behavior or lack thereof as if it were a human.
  - Does it matter if it is human-shaped?

  - We will update our mental model over time to more accurately capture the robot's state.
  - How can we design a robot such that people's model of it is more accurate?
- Maybe a principle of robot design should be to work with the model of the robot that people have, rather than trying to affect that model.
- Transparency – where does the data come from for learning?

5.3.2 Robots modeling humans

- Understanding/demonstrating social behavior:
  - Depends in part on reasoning about plans, beliefs, goals
  - Timing, dialog, cutting in, ...
  - What can we learn from cognitive science interests, e.g., human-human interaction studies?
- Some characteristics can be learned from data, but not all.
  - What are the features that such a representation would need to learn that model humans?
  - It is different if you are learning ethical principles.
- Need to consider roles and role-switching to handle such learning.
  - Speeds up planning in a collaborative setting to have an understanding of acceptable behavior/social norms.

5.3.3 Norms: representing, learning, following them

- What is the best formalism to express norms/ethical principles?
- Need dialog/some capability of learning from being "told."
- Need more general reasoning and more commonsense/general knowledge
  - It depends on how expensive plan changes are, how long the planning horizon is, etc.
- How to learn norms? From observations, instructions, ...?
  - Norms vary in importance, consequences.
  - We learn norms from a variety of mechanisms:
    important things are written down, less important things are told, some things are just learned from demonstrations.
- Important for norm learners: must be able to learn norms online.
  - Can't do a single model and then be done with it.
  - Online learning and online adaptation.
- Do we have to learn norms in context?
  - General vs. specific vs. culturally-modulated norms.
- Challenge: doing online learning, but not trying stupid things that violate social requirements.
  - But children push boundaries to improve understanding.
  - Learn in simulation?
- Concept of risk, balancing information gain with possible seriousness of a transgression.
  - Four choices: be extremely conservative to try to minimize norm violations; watch and see; ask; or try it and see what happens
- What can we do in simulation?
- Norms can be complex/contradictory/overlapping.
  - Learning sufficiently to act appropriately is difficult.
- Need to watch for signals and adjust norms over time.

    - How do you know how to adjust behaviors?
    - How do you recognize signals that you have violated a norm?
- There is a gradation from benign to strong social norms (mild vs. serious), long-term vs short-term – is there a spatio-temporal hierarchy?
    - People violate norms all the time.
- Challenge: what do we start with?
- Challenge: a lot has to happen in parallel; there is a control problem of making the layers of the robot architecture work together with timing.

5.3.4 Communication and information flow

- For HRI, humans and robots need to communicate.
    - Many modalities of interagent communication.
    - Language, legible behavior, ...
- Need some model of information flow that is deliberate on the part of at least one actor.
    - There exists work on recognition and activity/plan recognition–what else is there that robots can learn from passive observation?
- How can we communicate by inferring from behavior?
    - If you act to make your model clear via inference, you are communicating.
    - Some things are also best conveyed via being told, e.g., driving regulations.
- Just conveying information is not enough.

5.3.5 Proactive behavior

- Desirable for robots to be not purely reactive, but
- More of a problem for the robot to get things wrong when assisting than to do nothing.
    - Do people have a charitable view of a robot if the robot meant well but messed up?
    - Apologizing helps.
    - Depends partly on horizon – for how long will it be bad at something before it becomes good/helpful?

5.3.6 Grand and small challenges

- Supermarket:
    - Sub-problem: socially aware spill detector;
    - Sub-problem: getting something from the shelf for someone.
- Polite restaurant server:
    - When to interrupt, how long to leave the table alone, ...
- Shared manipulation/physical HRI:
    - Joint manipulation (putting all the dishes on the trolley);
    - Joint cooking.
- Seeing-eye Spot robot
    - Intelligent disobedience;
    - Epistemic reasoning about human’s beliefs, intentions.

5.4 Planning and Learning

Planning and learning have traditionally been two separate research tracks within cognitive robots. Lately, several research groups have started to study the combination and integration of planning and learning. For example learning symbols or primitives from observations. To achieve this, it is important to use the right inductive biases in learning to ground the AI system in the world. The key to complex behavior is being able to compose these into more complex plans or composite behaviors, thus planning based on these learned primitives clearly adds a significant value. This section provides a roadmap to achieve this in the form of three short-term challenges and four longer term challenges.

5.4.1 Roadmap

- Short-term Challenge 1: Bootstrap the knowledge – preliminary information, partial models
  - Examples of success: Given a partial PDDL for any planning model learning approach, and observing an improved performance in acquiring the rest of the model. Easy for aspects like SAT/ASP-based learning, but not so clear for deep learning methods that try to acquire things in an end-to-end way.
  - ETA 4 years
- Short-term Challenge 2: Life-long learning – models will drift, change, no longer be valid, etc
  - Example of success: Reliably able to detect when the model is no longer valid and how it has changed. Detect when new object types or new instances of an object are introduced. Being able to modify existing knowledge/model so that action which failed after world change now succeeds. Taking advantage of past experiences to quickly adapt to new environments.
  - ETA: 4 years
- Short-term Challenge 3: Leveraging our model specifications / formal languages to help traditional learning, e.g. interpretability/explainability: "why did my model do this?"; robustness: ensuring a DL system performs as expected; fairness: detecting biases, establishing and verifying fairness criteria
  - Example of success: Reasoning-based approach to verify / validate the concepts learned by traditional DL systems (e.g., interrogating LLM's for consistent reasoning)
  - ETA: 4 years
- Long-term Challenge 1: Integration of learned dynamics and hand-crafted models. Understanding the aspects of the models learned by agents – aligning / grounding the symbols specified & learned (including grounding language).
  - This challenge includes generating high level plans to deal with all kinds of complex environment which could include those with non-rigid objects (e.g. bed sheets) but also cluttered environments, uncertain environments, environments with other agents who change the environment dynamically.
  - Example of success: Creating plans that involve learning dynamics e.g., folding a bed sheet.
  - ETA: 10+ years
- Long-term challenge 2: HRI-style Model Acquisition. How to ground symbols interactively (with human users) to iteratively build a planning model (including objects, fluents, actions, etc). Aligning agent's internal language to the one used by the human. Extension – ability to align to multiple humans, using different concepts and languages/phrasing.

  - Example of success: Robot being capable of interactively receiving instructions (while clarifying) and performing the task given by human operators. ALFRED may be an initial starting point (language is all pre-known, as are the goals).
  - ETA: 10+ years
- Long-term challenge 3: Exploration based learning – i..e the robot actively exploring the world and trying to perform experiments to learn more about the world, and its capabilities and how actions affect the world.
  - Example of success: Simulated environment to place an egocentric agent in – success measured in properly acquiring a correct (or correct enough) planning model
  - ETA: 10+ years
- Long-term challenge 4: Multi-agent/human collaboration – learning how to collaborate with another agent to perform a task
  - Example of success: learning how to hand over an object, or jointly moving some large object, or collaborating to build some object (one agent holding the work-piece to resist forces such as sawing or drilling being applied by a second agent).
  - ETA: 5+ years

Participants

- Michael Beetz
Universität Bremen, DE
- Mohamed Behery
RWTH Aachen, DE
- Jens Claßen
Roskilde University, DK
- Anthony Cohn
University of Leeds, GB
- Frank Dignum
University of Umeå, SE
- Alexander Ferrein
Fachhochschule Aachen, DE
- Michael Fisher
University of Manchester, GB
- Hector Geffner
UPF – Barcelona, ES
- Jasmin Grosinger
University of Örebro, SE
- Nick Hawes
University of Oxford, GB
- Fredrik Heintz
Linköping University, SE
- Till Hofmann
RWTH Aachen, DE
- Mikhail Khodak
Carnegie Mellon University – Pittsburgh, US
- Sven Koenig
USC – Los Angeles, US
- Gerhard Lakemeyer
RWTH Aachen, DE
- Yves Lesperance
York University – Toronto, CA
- Setareh Maghsudi
Universität Tübingen, DE
- Cynthia Matuszek
University of Maryland, Baltimore County, US
- Sheila McIlraith
University of Toronto, CA
- Christian Muise
Queen's University – Kingston, CA
- Bernhard Nebel
Universität Freiburg, DE
- Tim Niemueller
Intrinsic Innovation – München, DE
- Ron Petrick
Heriot-Watt University – Edinburgh, GB
- Sebastian Sardiña
RMIT University – Melbourne, AU
- Matthias Scheutz
Tufts University – Medford, US
- Stefan Schiffer
RWTH Aachen University, DE
- Maayan Shvo
University of Toronto, CA
- Gerald Steinbauer
TU Graz, AT
- Brian C. Williams
MIT – Cambridge, US

Report from Dagstuhl Seminar 22392

Transparent Quantitative Research as a User Interface Problem

Chat Wacharamanotham*[1], Yvonne Jansen*[2], Amelia A. McNamara*[3], Kasper Hornbæk*[4], Judy Robertson*[5], and Lahari Goswami†[6]

1 **Swansea University, GB & University of Zurich, CH.** `chat@acm.org`
2 **CNRS – Talence, FR.** `yvonne.jansen@cnrs.fr`
3 **University of St. Thomas – St. Paul, US.** `amelia.mcnamara@stthomas.edu`
4 **University of Copenhagen, DK.** `kash@di.ku.dk`
5 **University of Edinburgh, GB.** `judy.robertson@ed.ac.uk`
6 **University of Lausanne, CH.** `Lahari.Goswami@unil.ch`

Abstract

The replication crises in many scientific fields galvanize movements toward Open Science. Within this movement is a push for increasing research transparency. Although researchers in the areas of Human–Computer Interaction (HCI) and Visualization (VIS) face these challenges, they have methodological expertise to study, design, and evaluate innovations that could help improve research transparency. This Dagstuhl Seminar gathers HCI and VIS researchers and those from adjacent fields such as statistics and psychology to discuss challenges in promoting and adopting research transparency, create prototypes of potential solutions, and receive feedback from policy influencers in the research community. This seminar fostered seeds for future initiatives and collaboration toward improving research transparency in HCI, VIS, and other scientific fields.

Seminar September 25–30, 2022 – `http://www.dagstuhl.de/22392`
2012 ACM Subject Classification Human-centered computing → Empirical studies in HCI; Human-centered computing → Empirical studies in visualization
Keywords and phrases Open Science, Research Transparency, Replicability, Reproducibility, Dagstuhl Seminar
Digital Object Identifier 10.4230/DagRep.12.9.220

1 Executive Summary

Chat Wacharamanotham (Swansea University, GB and University of Zurich, CH, `chat@acm.org`*)*
Yvonne Jansen (CNRS – Talence, FR, `yvonne.jansen@cnrs.fr`*)*
Amelia A. McNamara (University of St. Thomas – St. Paul, US, `amelia.mcnamara@stthomas.edu`*)*

Introduction

Many scientific fields face a **replication crisis**: A sizable portion of quantitative research studies could not be replicated. When these studies were re-run with higher statistical power (i.e., more participants), their results yielded effects substantially weaker or even opposite of

* Editor / Organizer
† Editorial Assistant / Collector

that in the original studies. This lack of replicability threatens the credibility of research claims and undermines the general public's trust in science. The replication crisis motivated the **Open Science** movement that **promotes transparency throughout the scientific process**: research funding, research design, data collection and analysis, peer reviewing, and knowledge dissemination. These phenomena attracted the interest of researchers in the fields of **Human–Computer Interaction (HCI)** and **Visualization (VIS)** for two reasons. Like other fields, HCI and VIS researchers face challenges in promoting transparency among their peers, effectively implementing and educating transparent practices, and incorporating transparency in the research evaluation processes. However, HCI and VIS researchers have the methods and skills to empirically study these phenomena and design potential solutions. The fields of HCI and VIS also provide a challenging testbed for these inventions.

This Dagstuhl seminar initiated and advanced works on these issues by bringing together 23 researchers from HCI, VIS, statistics, psychology, data science, and philosophy. They were from Australia, Austria, Canada, Denmark, Finland, France, Germany, the Netherlands, Sweden, Switzerland, the UK, and the USA. Three participants joined online due to the COVID situation and travel difficulties.

Program

We worked in groups to identify problem areas and prototype potential solutions in a Hackathon. We solicited feedback on these prototypes from conference and journal editors and community leaders. The seminar unfolded as follows:

Day 1: After a brief introduction to the purpose of the seminar and the overall plan, participants discussed in small groups to identify problems and challenges to work on in the Hackathon. These discussions were intentionally designed to be free-form to avoid prematurely limiting the areas of interest. To stimulate discussions and spark ideas, we provided the participants access to free-text responses to a survey on the perception of research transparency that we collected from HCI researchers in the weeks before the seminar. Four rounds of discussion were interleaved with three-minute presentations of intermediate results in the plenary to facilitate convergence and consolidation.
In each plenary round, we also asked a few participants to interview each other in front of the room to acquaint everyone with their background and research interest. Day 1 concluded with four clusters of topics to be worked on: (1) Educating researchers, (2) Clarifying the threats from the lack of transparency, (3) Clarifying the "transparency" concept, and (4) Working on how to influence policy and procedures in the publication process.

Day 2: Participants joined the problem cluster according to their interests and started the Hackathon. We provided each group with collaborative workspaces on Google Docs and Miro (an online whiteboard platform). After two Hackathon sessions in the morning, we further stimulated their work with an input lecture from Tim Errington, the Senior Director of Research at the Center for Open Science (see below for an abstract). This lecture highlighted challenges in promoting research transparency and provided a framework for changing research culture at multiple levels: from top-down research funding policy and bottom-up to ease the implementation of transparent practices by providing infrastructure and incentives. After the lecture, the Hackathon continued. We wrapped up the day with a 3-minute presentation from each group and a plenary discussion.

Day 3: The Hackathon continued in the morning. We gave the participants prompts to encourage them to hone in on a concrete idea and realize a prototype that demonstrates the idea's essence. The afternoon is free time for the participants to self-organize group activities to promote trust and informal interactions. We did not organize an excursion because the transportation companies were unavailable.

Day 4: The Hackathon continued in the morning. In the afternoon, the participants presented their preliminary results to four panelists who joined online. The panelists hold influential positions in the research publication process in HCI and VIS: the SIGCHI President, the TOCHI editor-in-chief, the TVCG Associate Editor in Chief and Eurographics Publication Board, CG&A Associate Editor-in-chief, TVCG Associate Editor, and the vice-chair of the IEEE VIS Steering Committee. A discussion on feedback from policy-making perspectives followed each presentation. The conversation with the panelists broadened participants' views about stakeholders and potential concerns. After the discussion, there was a plenary discussion to process the input from the panel collectively. We identified four areas to work on in the manifesto: definition, benefits, subfield-suitability, and progressive transparency.

Day 5: Participants worked in groups to draft a manifesto on research transparency. The seminar concluded with a plenary session where we identified possible future projects, their follow-up actions, and coordinators.

Results

The tangible results of the seminar comprise four prototypes from the Hackathon and a draft manifesto:

1. To influence research funders' policies, we drafted a list of policy suggestions for incentivizing research transparency and Open Science.
2. To inspire researchers and students, we prototyped how we could collect, analyze, and showcase papers in the visualization field that are exemplary in their transparent practices.
3. To improve infrastructure, we identified low-hanging fruits in improving the user interfaces of the ethical review and publication processes to encourage transparent practices.
4. To ease the adoption of transparency practices, we prototyped a cheat sheet that provides reminders for considering transparent practices at each of the research stages. The cheat sheet also provides pointers to relevant guides and resources.

The draft manifesto clarifies the definition of research transparency, describes its benefit, calls for each subfield to identify its suitable set of transparent practices, and argues how transparent practices could be viewed as a progression instead of demanding everyone to be perfect at their first try. These results provide a starting point for future follow-up research and educational activities that will advance the understanding and adoption of research transparency in HCI, VIS, and beyond.

2 Table of Contents

3 Input Lecture

3.1 How science knows what it knows: Challenges in research transparency

Timothy M. Errington (Center for Open Science – Charlottesville, US, tim@cos.io*)*

In large-scale replication studies such as in Psychology and Cancer Biology, many replication studies yielded a smaller effect size than the original study [1, 2]. Low replicability challenges the credibility of science. Replicability is associated with several research best practices, such as preregistration, using large samples, and sharing research materials [3]. However, this knowledge seems inadequate to change scientists' behaviors widely. To change the research culture, we need to address both the lack of know-how and the lack of motivation. Both top-down and bottom-up efforts are necessary as shown in Figure 1: Funders should design incentives and policies that will change the norm of research practices. Conversely, research communities must establish infrastructures and invent technologies to facilitate these practices.

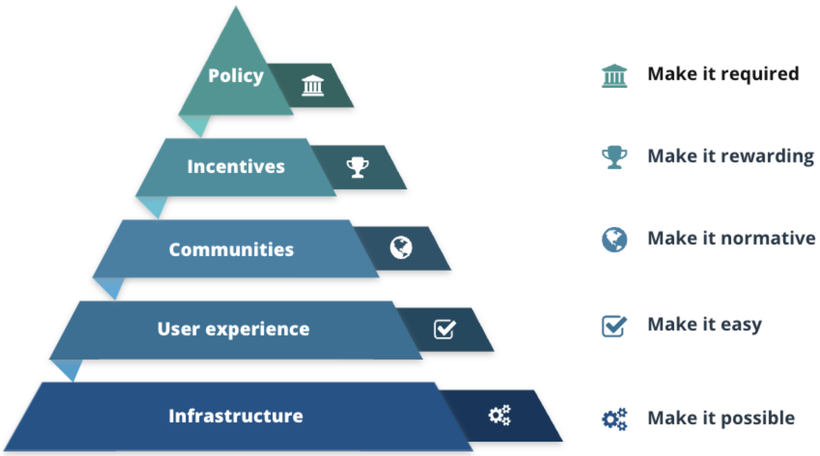


Figure 1 Center for Open Science strategy for scale sustainable adoption of open behaviors by researchers. Source: Center for Open Science www.cos.io/blog/continuing-acceleration-new-strategic-plan (Creative Commons BY 4.0 International license).

References

1 Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. http://dx.doi.org/10.1126/science.aac4716
2 Timothy M. Errington, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, and Brian A. Nosek (2021). Reproducibility in cancer biology: challenges for assessing replicability in preclinical cancer biology. Elife, 10, e67995. https://doi.org/10.7554/eLife.67995
3 John Protzko, Jon Krosnick, Leif D. Nelson, Brian A. Nosek, Jordan Axt, Matthew Berent, Nick Buttrick, Matthew DeBell, Charles R. Ebersole, Sebastian Lundmark, Bo MacInnis, Michael O'Donnell, Hannah Perfecto, James E Pustejovsky, Scott S. Roeder, Jan Walleczek, and Jonathan Schooler (2020). High replicability of newly-discovered social-behavioral findings is achievable. https://doi.org/10.31234/osf.io/n2a9x

3.2 Summary of Q&A and discussion

Lahari Goswami (University of Lausanne, CH, `lahari.goswami@unil.ch`*)*
Chat Wacharamanotham (Swansea University, GB & University of Zurich, CH, `chat@acm.org`*)*

Temporary changes. Trying out proof-of-concept solutions to transparency may require a shift in the incentives involved in the research process. Such a shift could be agreed upon to be temporary and to be adapted as needed.

Scaling up. Proof-of-concept solutions that are implemented in smaller conferences/journals require reflecting on the actual changes and their results. The transfer is more likely to happen by abstracting the lessons learned.

Methodological diversity. For fields that use more than quantitative research methodology – including HCI and VIS – preregistration could be a common starting point because it is widely applicable. Dialogues are needed with people from those methodologies to develop suitable infrastructure. For example, the preregistration template for qualitative research was developed by the lead from qualitative researchers.

Synergy across practices. To maximize the effectiveness of various novel transparent practices introduced into the scientific process, educating researchers on how these practices are connected to each other as an ecosystem is necessary. Such consideration could also help reduce researchers' effort, e.g., by aligning the information required for ethical review with those in preregistration.

4 Working Groups

After the first day of the seminar, we formed four working groups to address different aspects of research transparency. Below is a summary of the results from each working group ordered top-down: from the policy level to a concrete checklist for individual researchers.

4.1 Influencing research funders' policies

Contributors in alphabetical order:
Chat Wacharamanotham (Swansea University, GB & University of Zurich, CH, `chat@acm.org`*)*
Duong Nhu (Monash University – Clayton, AU, `duong.nhu1@monash.edu`*)*
Lynda Hardman (CWI – Amsterdam, NL & Utrecht University, NL, `Lynda.Hardman@cwi.nl`*)*
Sophia Crüwell (University of Cambridge, GB, `slbc2@cam.ac.uk`*)*

We reviewed the existing policies from funding agencies: Australia ARC and NHMRC, European ERC, the Netherlands' NWO, and UK's UKRI. Although these policies encourage open-access publication and sharing of research data, they lack other transparency practices.

We also found the policies focus on goals without adequately providing resources and infrastructure. Therefore, we drafted seven policy recommendations. For each recommendation, we discussed the status quo, its benefits, possible barriers, and possible improvements. Below is a summary of each recommendation:

R1: Mandate Open Science as the default in all funding schemes, with open-access publication as a required minimum. Several research funders already required open access to research publications. Where ethically feasible, researchers should make all data and analysis methods available on FAIR-compatible repositories and annotate them with appropriate domain-specific metadata.

R2: Provide funding for infrastructure and personnel to implement Open Science practices. Institutions, departments, or individual researchers may need more resources or knowledge to meet all requirements of conducting Open Science research. Funding should be made available for (inter)national infrastructures and repositories for secure data storage and controlled access to data.

R3: Include Open Science practices as an independent dimension in evaluating proposals, individual researchers, and departments. Open Science practices take additional resources (e.g., time, money, personnel). Instead of spending resources on Open Science practices, researchers channel these resources to write more papers. However, more papers that are not transparent increase noise in the body of knowledge. Therefore, when evaluating research proposals, researchers, and department performance, Open Science practices should be incorporated as an independent dimension that could be considered in the context of other metrics such as grant income or publication count.

R4: Positive discrimination of researchers who use or have used Open Science. Adoption of Open Science practices may lead to a disadvantage in the current system, which is focused on quantity and novelty over quality. If engagement with Open Science is only optional, we could punish Open Science pioneers, thus disincentivizing engagement with Open Science. Funders can address this problem by giving special consideration to researchers whose track records demonstrate Open Science practices.

R5: Develop recommendations and rules for sharing research artifacts. Funders should mandate or recommend types of licenses and – where available – specialized repositories for sharing research artifacts. These rules or recommendations will facilitate data dissemination and reuse.

R6: Encourage Open Science Best practices. Best practices in Open Science evolves as innovation are developed and tested in various subfields. Funders can accelerate these processes by publishing and periodically updating lists of recognized best practices. Additionally, funders should invest in developing Open Science innovations and maintaining essential Open Science infrastructures.

R7: Provide Open Science training for all personnel involved in research. Researchers at different career stages have different training needs: Junior researchers might need awareness and practical skills in implementing Open Science practices. Established researchers will need to be convinced of the relevance of Open Science and why they are relevant in a new norm of scientific practices. Funding agencies should open calls for research projects to develop innovative educational materials for Open Science and to evaluate their effectiveness. Additionally, funding agencies could incentivize institutions to allocate resources and provide competence in training Open Science knowledge and skills.

After presenting this draft to the panelists, we received feedback that buy-in from the research community is essential to make changes, especially for the fields of HCI and VIS, where there's a wide range of research methods and contribution types. Any policy changes mustn't marginalize any research methodologies or domains. The follow-up action for this work is to refine the recommendations further to address the methodological diversity concern and engage in conversation with funding organizations.

4.2 Positive Examples of Research Transparency

Contributors in alphabetical order:
Eunice Jun (University of Washington – Seattle, US, `emjun@cs.washington.edu`*)*
Lonni Besançon (Linköping University, SE, `lonni.besancon@gmail.com`*)*
Michael Sedlmair (Universität Stuttgart, DE,
`michael.sedlmair@visus.uni-stuttgart.de`*)*
Pierre Dragicevic (INRIA – Bordeaux, FR, `pierre.dragicevic@inria.fr`*)*
Theophanis Tsandilas (Université Paris-Saclay, Orsay, FR & Inria – Orsay, FR,
`fanis@lri.fr`*)*
Wesley Willett (University of Calgary, CA, `wesley.willett@ucalgary.ca`*)*

We saw an opportunity to encourage transparent research practices by curating a set of examples from visualization research. These examples will serve as a starting point for teaching transparent research, motivating visualization researchers to practice research transparency, and motivating the community to invent new creative transparent methods. Toward these goals, we identified forms of transparency: (1) research process and methods, (2) artifacts, (3) data, and (4) claims and limitations. Combining these forms with a taxonomy of visualization contribution types (Munzner, 2008) resulted in a matrix for the examples. As a prototype for this seminar, we brainstormed and discussed how some papers we knew fit into this matrix.

From the panelists' feedback, we realized the importance of clarifying that these examples show the possibilities of transparency. Some contribution types may lend themselves to some forms of transparent practices easier than others. The paper should be framed as a recommendation instead of a checklist to avoid alienating some research contributions.

For the next step, we plan to collect examples that span this matrix by conducting a survey targeting visualization experts, and we drafted survey questions. We plan to publish this work in IEEE Computer Graphics and Applications (CG&A) as a viewpoint paper with extensive supplementary materials on OSF.

References

1 Tamara Munzner (2008). Process and pitfalls in writing information visualization research papers. In Information visualization (pp. 134-153). Springer, Berlin, Heidelberg. `doi.org/10.1007/978-3-540-70956-5_6`

4.3 Tweaks to the Ethics Review & Publication Pipeline that Encourages Transparency

Contributors in alphabetical order:
Julien Gori (Sorbonne University – Paris, FR, gori@isir.upmc.fr*)*
Kavous Salehzadeh Niksirat (University of Lausanne, CH, kavous.salehzadehniksirat@unil.ch*)*
Olga Iarygina (University of Copenhagen, DK, olia@di.ku.dk*)*
Ulrik Lyngs (University of Oxford, GB, ulrik.lyngs@cs.ox.ac.uk*)*
Yvonne Jansen (CNRS – Talence, FR, yvonne.jansen@cnrs.fr*)*

We looked at the steps of the research pipeline, from ethics application to published articles, as shown in Figure 2. We identified several opportunities to encourage open and transparent research practices along this pipeline. Instead of looking to impose the "sticks" – punishments or requirements – we focus on "carrots", emphasizing selfish benefits to incentivize the practices. This analysis results in minor changes to ethics review templates, transparency statements on the CHI website, PCS submission interface, and acceptance notifications. These all emphasize reasons for authors to strive for transparency in their research.

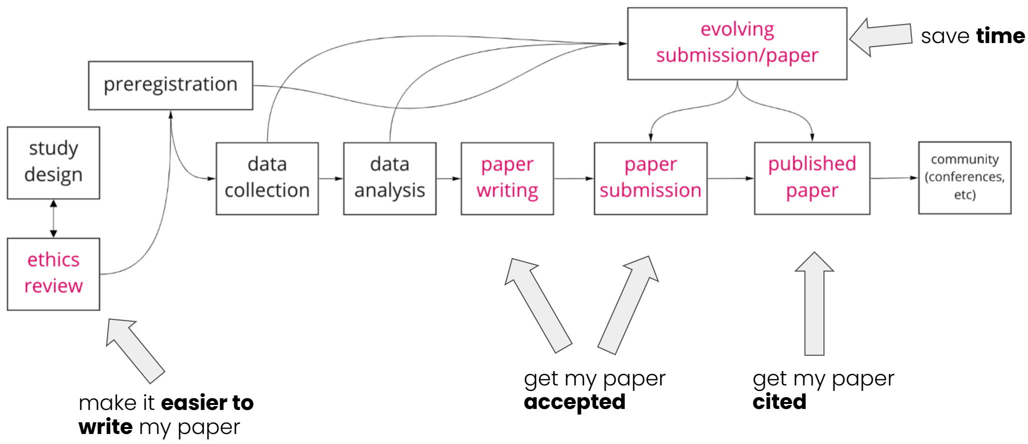


Figure 2 Research process and selfish benefits that could be used to incentivize research transparency.

For the ethics review template, we prototyped additions of instructions that could remind researchers to consider preregistration, documenting their data analysis process or making them reproducible, and repositories to share their data.

Specifically to the ACM CHI Conference on Human Factors in Computing Systems (CHI) – a major HCI conference, we drafted an update to the Guides for Authors to provide concrete suggestions and be more concise. We also created a mockup of the paper submission form (Figure 3) and acceptance email message that will encourage researchers to think about different types of research artifacts and data to share.

We also discussed an idea of repurposing the comments fields from the paper websites, e.g., the ACM Digital Library, for the authors to point to research materials that become available post-publication. Reference manager software could alert researchers of these updates – similar to Retraction Watch alert plug-in for Zotero.

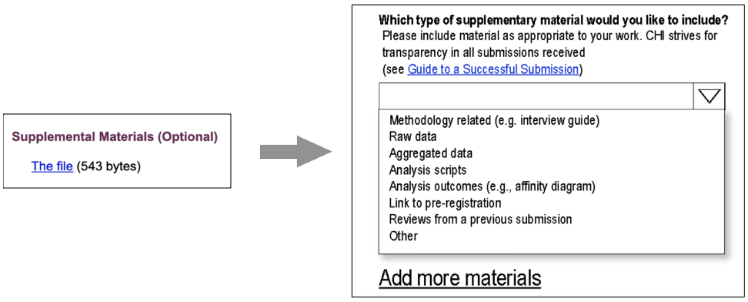


Figure 3 A mock-up of a change in the user interface of the paper submission form to promote awareness and sharing of research artifacts.

The panelists are positive about a minor modification to the submission form and suggested that this idea could be tested in small conference venues by convincing their paper chairs. As for the evolving paper idea, a panelist warned that a side effect of this feature might discourage authors from submitting materials on the paper deadline. Lastly, although placing reminders in the ethics review form is helpful, not all research activities go through the ethics approval process. Besides, several countries may still need to implement an ethics approval process.

The following steps in this direction are to refine the draft items for ethics submission templates, solicit feedback from local ethics committees, and publish a guiding resource for local ethics boards that wish to include open research practices in ethics review.

4.4 Infrastructure for evolving research and mega studies

Contributors in alphabetical order:
Florian Echtler (Aalborg University, DK, `floech@cs.aau.dk`*)*
Pierre Dragicevic (INRIA – Bordeaux, FR, `pierre.dragicevic@inria.fr`*)*
Ilya Musabirov (University of Toronto, CA, `ilya@musabirov.info`*)*

We challenged the idea that research ends with publications. Instead, research should be considered evolving scholarly records where scholarly discussions are added and refined over time. Although this idea has been previously discussed [1], a technical infrastructure must be developed. We discussed the requirements of such technical infrastructure and its necessity to support the research lifecycle from pre-publication, such as preregistrations and internal reviews. We discuss the pros and cons of repurposing version control systems (e.g., GitHub) or tools like hypoths.is that overlay upon existing infrastructure. We also discussed the emerging trend of mega studies where multiple labs collaborate on one study to enable larger sample sizes. These studies lend themselves to transparent practices because much information must be digitally shared among collaborators. However, there are also challenges in tracing rationale in study design and data analyses across multiple actors.

The panelists pointed out challenges. Requiring the reviews to be public may discourage potential reviewers, such as junior researchers, who risk their careers if their review disagrees with senior paper authors. Signed public reviewing could face copyright issues, while anonymous public reviewing might lead to spam.

References

1 Brian Lavoie, Eric Childress, Ricky Erway, Ixchel Faniel, Constance Malpas, Jennifer Schaffner, and Titia Van der Werf (2014). The Evolving Scholarly Record. OCLC Online Computer Library Center, Inc. 6565 Kilgour Place, Dublin, OH 43017.

4.5 A Cheatsheet for a Transparent CHI Paper

Contributors in alphabetical order:
Amelia A. McNamara (University of St. Thomas – St. Paul, US, `amelia.mcnamara@stthomas.edu`*)*
Erich Weichselgartner (Universität Graz, AT, `erich.weichselgartner@uni-graz.at`*)*
Jan B. Vornhagen (Aalto University, FI, `jan.vornhagen@aalto.fi`*)*
Lahari Goswami (University of Lausanne, CH, `lahari.goswami@unil.ch`*)*
Viktorija Paneva (Universität Bayreuth, DE, `Viktorija.Paneva@uni-bayreuth.de`*)*

In recent years there has been an upsurge of high-quality and accessible learning materials for Open Science practices (e.g., FORRT, FOSTER[1]). However, these have yet to see widespread use in the CHI community. Therefore, we offer CHI authors a tailored cheat sheet as an easy-to-use reference and guide to existing resources, allowing for convenient integration of these practices into their current workflow. We identified typical research phases in HCI studies and brainstormed relevant guides and resources for each stage. These pieces of information are assembled on two-page cheat sheets that briefly explain the rationale for each practice, provide concrete and concise action items, and point to resources to learn more (Figure 4). The current version of the cheat sheet is available on OSF[2]. Further contributions are welcome on the GitHub repository[3].

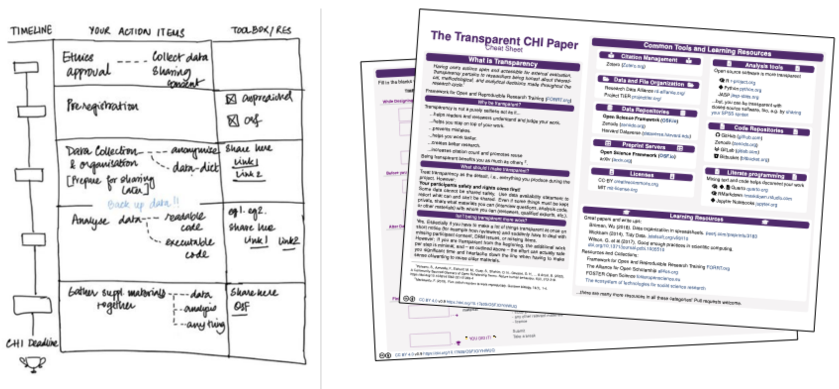


Figure 4 Left: An early draft of the design of the cheat sheet. Right: The realized cheatsheet as of December 2022.

[1] `https://forrt.org/`, `https://www.fosteropenscience.eu/`
[2] `https://doi.org/10.17605/OSF.IO/YHWUQ`
[3] `https://github.com/jvornhagen/ACheatSheetForTheTransparentCHIPaper/`

5 A Manifesto for Transparent Quantitative Research

We distilled lessons learned from the seminar into the following manifesto.

5.1 Definition

Contributors in alphabetical order:
Chat Wacharamanotham (Swansea University, GB & University of Zurich, CH, `chat@acm.org`*)*
Jan B. Vornhagen (Aalto University, FI, `jan.vornhagen@aalto.fi`*)*
Julien Gori (Sorbonne University – Paris, FR, `gori@isir.upmc.fr`*)*
Pierre Dragicevic (INRIA – Bordeaux, FR, `pierre.dragicevic@inria.fr`*)*
Yvonne Jansen (CNRS – Talence, FR, `yvonne.jansen@cnrs.fr`*)*

"**Transparency in research** refers to honesty and clarity in all communications about the research processes and outcomes – to the extent possible."

We unpack each facet of this definition as follows:

Honesty and clarity are both necessary. A paper that is very clear but makes misleading claims is not transparent. But a paper written by authors who are fully honest but unskilled at clear communication is not fully transparent either.

Communications include communications between the researchers and their colleagues, peers, institutions, the press, and the general public.

The research process includes all the known, crucial decisions made to achieve the reported outcome.

The research outcomes include research materials (including data and software), findings, and communication artifacts (research papers, videos). Being transparent about research outcomes includes sharing material (e.g., code and data) but also being clear about the limitations of the research.

To the extent possible. Transparent research practices operate within ethical, resource, legal, and other constraints. These include ethics constraints (such as participant rights and protections), legal constraints (such as data protection laws), and resource constraints (such as access to data repositories). We also acknowledge that there is information that is not accessible by the researchers; we can only ask them to communicate the information they can reasonably know.

5.2 Benefits

Contributors in alphabetical order:
Erich Weichselgartner (Universität Graz, AT, erich.weichselgartner@uni-graz.at*)*
Eunice Jun (University of Washington – Seattle, US, emjun@cs.washington.edu*)*
Lonni Besançon (Linköping University, SE, lonni.besancon@gmail.com*)*
Olga Iarygina (University of Copenhagen, DK, olia@di.ku.dk*)*
Sophia Crüwell (University of Cambridge, GB, slbc2@cam.ac.uk*)*
Wesley Willett (University of Calgary, CA, wesley.willett@ucalgary.ca*)*

Transparent research practices **support the refinement, reuse, and verifiability** of published research as well as its **extension** through follow-on research and meta-analyses. Transparent practices also support the use of research materials and findings for **instruction** and can increase the visibility, perceived credibility, and citation of both individual research findings and broader research areas. Transparent sharing of materials may also allow for more error detection and correction, thus fostering a larger error-correction culture. Furthermore, making research materials available to other researchers to use in their contexts will reduce duplication of effort while also helping grow and diversify research communities.

5.3 Transparency needs tailoring

Contributors in alphabetical order:
Amelia A. McNamara (University of St. Thomas – St. Paul, US, amelia.mcnamara@stthomas.edu*)*
Kavous Salehzadeh Niksirat (University of Lausanne, CH, kavous.salehzadehniksirat@unil.ch*)*
Sophia Crüwell (University of Cambridge, GB, slbc2@cam.ac.uk*)*
Ulrik Lyngs (University of Oxford, GB, ulrik.lyngs@cs.ox.ac.uk*)*

Transparent research is not about rigid box-checking. **Not all transparent research practices apply to all kinds of research, and the specific implementation of any given transparent practice depends on the subfield and research type.** The practices that are useful for, for example, in-depth interview studies differ from those that are useful for iterative design & user testing studies or for large-scale online experimental work. Therefore, we do not want to make specific, one-size-fits-all recommendations. Instead, we suggest questions researchers can ask themselves and their field. Thinking about the typical research process and outputs in your field, consider the following:
1. What elements would help build credibility for your work for reviewers and readers?
2. If another researcher wanted to build on your work, what elements would be helpful?

These questions could be used to build transparency guidelines for a particular subfield or guide an individual researcher in implementing transparent practices in their own research.

5.4 Transparency could be progressive

Contributors in alphabetical order:
Duong Nhu (Monash University – Clayton, AU, duong.nhu1@monash.edu*)*
Lahari Goswami (University of Lausanne, CH, lahari.goswami@unil.ch*)*
Theophanis Tsandilas (Université Paris-Saclay, Orsay, FR & Inria, Orsay, FR, fanis@lri.fr*)*
Viktorija Paneva (Universität Bayreuth, DE, Viktorija.Paneva@uni-bayreuth.de*)*

Becoming transparent requires our research community to give up old research practices and adopt new research methods. Transparency is a progressive process that starts with small steps before it becomes a long-term, well-established practice. It can seem intimidating at first, but here are some steps that can guide you on your journey to making your research practice more transparent:

Become aware. Understanding and adopting transparent practices for the first time can seem difficult and complicated. There is no one-size-fits-all mechanism for adopting transparent practices. So it is essential not to set unrealistic expectations for ourselves when starting. Let's start with the easy step, which is documenting our research process – initially for ourselves. An example would be maintaining a research journal. Understanding that transparent practices depend on the context of our research is crucial – as a result, one should self-regulate, reflect and learn from our process. The documentation should be refined for others to comprehend.

Learn. Learn from others. Other practitioners have spent numerous hours learning about transparent research practices and incorporating those into their research. Use this valuable resource. Learn by example – look for good examples of papers related to your area of research that showcase how this particular type of research can be made more transparent. Also, you are not alone in this. Discuss with the people in your immediate environment, e.g., peers, supervisors, and other colleagues, how you can collaborate to make your work more transparent and learn from each other.

Adopt. Transparency can be adopted step by step. Researchers should consider what parts of the project can be shared and plan from the beginning of the project, from grant proposal to ethics application. Transparency is not only about sharing but also about self-reflection and learning from your progress. Documenting the project's progress is an easy task and an excellent start to integrating transparency into your research. It will help better understand the research problem, detect mistakes early, and make it easier to review and evaluate later. This documentation will eventually become a foundation for writing comprehensive documentation for sharing with the community.

Educate. Your practice of transparent research methods can serve as an example for your students and your community. A minimal step is to share your open research knowledge with your peers and, if you are a teacher, to incorporate it into your teaching materials.

Influence. We can take advantage of our roles within our institution and research community to advance transparent research practices. We can encourage and reward openness and transparency as reviewers and program committee members. Criteria of transparency may vary across subcommunities in our field, so we need to work with our peers in these communities to understand better how transparency applies to their research methods.

Participants

- Lonni Besancon
Linköping University, SE
- Sophia Crüwell
University of Cambridge, GB
- Pierre Dragicevic
INRIA – Bordeaux, FR
- Julien Gori
Sorbonne University – Paris, FR
- Lahari Goswami
University of Lausanne, CH
- Lynda Hardman
CWI – Amsterdam, NL &
Utrecht University, NL
- Olga Iarygina
University of Copenhagen, DK
- Yvonne Jansen
CNRS – Talence, FR
- Eunice Jun
University of Washington –
Seattle, US
- Ulrik Lyngs
University of Oxford, GB
- Amelia A. McNamara
University of St. Thomas –
St. Paul, US
- Duong Nhu
Monash University –
Clayton, AU
- Viktorija Paneva
Universität Bayreuth, DE
- Michael Sedlmair
Universität Stuttgart, DE
- Kavous Selahzadeh Niksirat
University of Lausanne, CH
- Theophanis Tsandilas
Universitë Paris-Saclay, Orsay,
FR & Inria, Orsay, FR
- Jan Benjamin Vornhagen
Aalto University, FI
- Chat Wacharamanotham
Swansea University, GB
- Erich Weichselgartner
Universität Graz, AT
- Wesley J. Willett
University of Calgary, CA

Remote Participants

- Florian Echtler
Aalborg University, DK
- Ilya Musabirov
University of Toronto, CA

Panelists

- Neha Kumar
SIGCHI President
Georgia Institute of Technology –
Atlanta, US
- Jean-Daniel Fekete
TVCG Associate Editor in Chief
and Eurographics Publication
Board
Inria – Orsay, FR
- Kristina Höök
TOCHI editor-in-chief
KTH Royal Institute of
Technology – Stockholm, SE
- Petra Isenberg
CG&A Associate Editor-in-chief,
TVCG Associate Editor, and the
vice-chair of the IEEE VIS
Steering Committee
Inria, Orsay, FR