Report from Dagstuhl Perspectives Workshop 11331

The Future of Research Communication

Edited by

Tim Clark[1], Anita De Waard[2], Ivan Herman[3], and Eduard Hovy[4]

1 **Harvard Medical School & Massachusetts General Hospital, US,** twclark@nmr.mgh.harvard.edu
2 **Elsevier Labs – Jericho, US,** a.dewaard@elsevier.com
3 **W3C/CWI – Amsterdam, NL,** ivan@w3.org
4 **University of Southern California – Marina del Rey, US,** hovy@isi.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 11331 "The Future of Research Communication". The purpose of the workshop was to bring together researchers from these different disciplines, whose core research goal is changing the formats, standards, and means by which we communicate science.

Seminar 15.–18. August, 2011 – www.dagstuhl.de/11331
1998 ACM Subject Classification H.2.8 Database Applications, H.3 Information Storage and Retrieval, H.5 Information and Interfaces and Presentation, I.7.4 Electronic Publishing, J.7 Computers in other Systems, K.4 Computers and Society
Keywords and phrases science publishing, online communities, science policy, new forms of publishing, bioinformatics, digital repositories, semantic publishing, citation analysis, data publication, information access and integration, reporting standards
Digital Object Identifier 10.4230/DagRep.1.8.29
Edited in cooperation with Aliaksandr Birukou

1 Executive Summary

Philip E. Bourne
Tim Clark
Robert Dale
Anita de Waard
Ivan Herman
Eduard Hovy
David Shotton

Research and scholarship lead to the generation of new knowledge. The dissemination of this knowledge has a fundamental impact on the ways in which society develops and progresses, and at the same time it feeds back to improve subsequent research and scholarship. Here, as in so many other areas of human activity, the internet is changing the way things work: it opens up opportunities for new processes that can accelerate the growth of knowledge, including the creation of new means of communicating that knowledge among researchers and within the wider community. Two decades of emergent and increasingly pervasive information technology have demonstrated the potential for far more effective scholarly communication.

The Future of Research Communication, *Dagstuhl Reports*, Vol. 1, Issue 8, pp. 29–52
Editors: Tim Clark, Anita De Waard, Ivan Herman, and Eduard Hovy
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

However, the use of this technology remains limited; research processes and the dissemination of research results have yet to fully assimilate the capabilities of the web and other digital media. Producers and consumers remain wedded to formats developed in the era of print publication, and the reward systems for researchers remain tied to those delivery mechanisms.

Force11 (the Future of Research Communication and e-Scholarship) is a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. Individually and collectively, we aim to bring about a change in scholarly communication through the effective use of information technology. Force11 has grown from a small group of like-minded individuals into an open movement with clearly identified stakeholders associated with emerging technologies, policies, funding mechanisms and business models. While not disputing the expressive power of the written word to communicate complex ideas, our foundational assumption is that scholarly communication by means of semantically-enhanced media-rich digital publishing is likely to have a greater impact than communication in traditional print media or electronic facsimiles of printed works. However, to date, online versions of 'scholarly outputs' have tended to replicate print forms, rather than exploit the additional functionalities afforded by the digital terrain. We believe that digital publishing of enhanced papers will enable more effective scholarly communication, which will also broaden to include, for example, better links to data, the publication of software tools, mathematical models, protocols and workflows, and research communication by means of social media channels.

This document reports on the presentations and working groups that took place during the Force11 workshop on the Future of Research Communication and e-Scholarship held at Schloss Dagstuhl, Germany, in August 2011. More about Force11 can be found at http://www.force11.org. This document is structured as follows. Sections 3-5 report on the presentations of the participants. The presentations discuss, respectively, the past (Section 3), present (Section 4) and future (Section 5) of research communication. Section 6 presents the notes from the working groups. The notes are presented with only minor modifications, to capture the spirit of what was happening "in situ". Section 7 lists the websites and other documents related to the workshop. Section 8 contains the timetable of the workshop. Finally, we list the participants of the workshop.

2 Table of Contents

Executive Summary
Philip E. Bourne, Tim Clark, Robert Dale, Anita de Waard, Ivan Herman, Eduard Hovy, and David Shotton . . . 29

Overview of Talks. The Past

The Future of Research Communications: The Past
Anita de Waard . . . 33

Net-Centric Scholarly Discourse?
Simon Buckingham Shum . . . 33

A Brief History of E-Journal Preservation
David S. H. Rosenthal . . . 34

Overview of Talks. The Present

Open Citations
David Shotton . . . 34

Scholarly Communication in the Present
Paul Groth . . . 34

What is holding us back? A short exploration of current impediments to integrated publishing of data and primary research
Fiona Murphy . . . 35

Making "Beyond the PDF" Current Practice
Philip E. Bourne . . . 36

A (very) short history of the ADS
Michael J. Kurtz . . . 36

How to communicate the data described in publications? The Dryad model
Todd Vision . . . 36

More than just data!
Cameron Neylon . . . 37

Overview of Talks. The Future

The Future. Or: What I would Like from Publications of the Future
Eduard H. Hovy . . . 37

Introduction to the Future of Research Communication
Tim Clark . . . 38

Networked Knowledge
Stefan Decker . . . 38

The Execution of Dave 2.0
David De Roure . . . 38

"Towards Horizons 2020" — The Framework Programme for Research and Innovation 2014 to 2020 and Role of scientific data
Mike W. Rogers . . . 39

Working Groups

Data 40

Tools and Technologies 41

Business models for the research communications in the future 46

Assessment and Impact 49

Relevant links 51

Agenda 51

Participants 52

3 Overview of Talks. The Past

3.1 The Future of Research Communications: The Past

Anita de Waard (Elsevier Labs – Jericho, US)

URL http://slidesha.re/pkspZZ

To see where we need to go in the future, it can be useful to look at the past with a critical eye. For instance, the concept of hypertext: selecting portions of text and linking them to other portions of text, has been around, conceptually, since Vannevar Bush, and practically, since Engelbart's 1968 seminal work. Yet apart from the web, which is a low-hanging fruit realisation of this idea – with only simple links that bring you to another page; not the conceptual networks that were originally conceived – the idea has never really come about, although it is reinvented with startling regularity. Why is this the case? We define four factors that contribute to a technology being accepted:

- Commercial support (e.g. Microsoft Word)
- Community uptake (e.g. LaTeX)
- Ease of use (e.g. the web)
- Academic Credit (e.g. grant proposals)

and discuss how these played a role for the topics discussed at Force11: New Formats, Research Data, Tools and Standards, Business Models, and Attribution and Credit.

3.2 Net-Centric Scholarly Discourse?

Simon Buckingham Shum (The Open University – Milton Keynes, GB) – twitter `@sbskmi`

URL http://slidesha.re/qvoqoU

To make science and scholarship into a more agile sensemaking and problem solving system, better able to respond to the demands of a rapidly shifting environment, we need tools designed for an infrastructure unimaginable in the 17th Century when the first scholarly journals were born. However, the paradigm these established still dominates how we continue to disseminate knowledge. The founding fathers of hypertext, Vannevar Bush (1945) and Doug Engelbart (1963), clearly had the future of scholarly communication in mind when they presented use cases for their pioneering intellectual technologies. In this talk I will trace the core ideas which research since has sought to bring to reality. The essence of the idea is that scholarly communication is the crafting and contesting of networks of ideas, such as claims, concepts, evidence, arguments, and that linear prose is only one way in which to express knowledge. I will give a few examples of how new contributions to the long term reflective conversation of scholarly communication can now be made using the social, semantic web operating across many kinds of device.

3.3 A Brief History of E-Journal Preservation

David S. H. Rosenthal (Stanford University Libraries, US)

URL http://blog.dshr.org/2011/08/brief-history-of-e-journal-preservation.html

Overview of the evolution of e-journal preservation from the initial Mellon Foundation projects to the present. How well did the various business models and technologies work? Where do the costs come from? What are the implications for the future?

4 Overview of Talks. The Present

4.1 Open Citations

David Shotton (University of Oxford, GB)

URL http://bit.ly/vnRNEQ

The Open Citations Corpus (http://opencitations.net/) contains references to 3.4 million biomedical papers, representing 20% of all PubMed Central papers published between 1950 and 2010, and including all the highly cited papers in every biomedical field. The Open Citations web site provides access to the entire corpus with various search and browse options. The entire dataset is downloadable in various formats, including RDF and BibJSON, for reuse. Incoming and outgoing citation networks of selected references can be displayed in different ways and downloaded in various formats. The citation contexts of in-text citation pointers can be used to text mine the cited article and pull back sentences of relevance, to assist the reader in evaluating the quality of the citation and the cited article.

4.2 Scholarly Communication in the Present

Paul Groth (VU University Amsterdam, NL)

URL http://bit.ly/uHWmE9

Current scholarly communication practices can be broadly classified into four main categories: papers, professional meetings, databases, and informal communication. We briefly describe these categories to provide a picture of communication practice in the year 2011.

Papers are the predominate category of scholarly communication and still follow roughly the same form as for the past 200 years. Books and monographs take the role of papers in some disciplines. The Internet has changed the manner in which papers are distributed and managed. Digital libraries and search engines are the primary means to find papers in many disciplines. Social media is playing an increasing role in surfacing particular papers. Interestingly, papers are now often referred to, not by a citation, but using a URL of the paper on the Web. Papers are managed by specific reference management software. Publication of

papers is still largely journal oriented and mediated through peer review and other editorial processes. Open access journals have become more common.

Professional meetings such as conferences, symposiums, and workshops play an important role as they provide forums for scientists to meet and discuss their latests findings and approaches without the lag of publication. This is particular important as research is often international in nature and thus requires face-to-face meetings. Increasingly, conferences leave traces on the Web through the posting of slides and other material as well as live conversations in social media.

Databases have become a primary mechanism for communicating results across scientific disciplines. Many journals in the life sciences, for example, require the deposition of data within on-line databases before a paper can be published.

Informal communication is an important part of the scholarly communication life cycle. The internet and in particular social media (blogs, microblogging, email forums) have become increasingly prevalent. However, the primary means of informal communication is email. Indeed, it can be safely said that email is the main means for scholarly communication today.

Finally, it is important to note that scholarly communication acts as one of a central proxy by which scientific performance is measured. Indeed, the publication of papers in journals is the single proxy often used and is often the basis for career advancement decisions.

While the Internet has changed the way scholarly communication is done. The journal paper still dominates as the primary trackable product of this communication.

4.3 What is holding us back? A short exploration of current impediments to integrated publishing of data and primary research

Fiona Murphy (Wiley-Blackwell, UK)

Others have also highlighted these points — towards promoting discussion. The issues/stakeholders are: Technology/systems, Funding bodies/mandates, Researcher behaviour, Publishers, Other.

Tech/systems: People collect data ad hoc on laptops. Often not collected with the final deposit/site in mind so incurring expense and difficulty, Interfaces may be unhelpful (BADC), Formats issue — danger of outdated media.

Funders: Historically unhelpful. Remote, not communicating or incentivising. Demanding compliance but not following through. In the process of changing gradually.

Researchers: Suspicious of sharing IP/politics (Climategate), Anecdotally data underground/siege mentality, No time, Do not see benefits. There is a missing member of the team. Not trained.

Publishers: Not facilitating — hesitant to invest do not see the benefits either, Used to dealing with libraries rather than end users, Locked into traditional mind-sets — incunabular, Not yet built expertise to required level, Partnerships unknown.

Other: Confusion about where data should sit: who is responsible?

4.4 Making "Beyond the PDF" Current Practice

Philip E. Bourne (UC San Diego, US)

URL http://www.slideshare.net/pebourne/dagstuhl

I report on my perspective as a computational biologists on what I consider major developments in scholarly communication that have happened in the past 7 months since the beyond the PDF workshop. Notable is the announcement of SciVerse from Elsevier which in my opinion has the potential to change the model for how we interact with scholarly content. I also describe my experiences and approach to the established notion of a data journal and how I propose to contribute. Finally, I describe recent experiences with workflows and my perceived impact that they might have on the reproducibility of science.

4.5 A (very) short history of the ADS

Michael J. Kurtz (Harvard-Smithsonian Center for Astrophysics, US)

Main reference Michael J. Kurtz, "The Emerging Scholarly Brain," in Future Professional Communication in Astronomy II, Astrophysics and Space Science Proceedings, 2011, Volume 1, pp. 23–35.
URL http://dx.doi.org/10.1007/978-1-4419-8369-5_3

The Smithsonian/NASA Astrophysics Data System is a sophisticated digital library/ information system; it is used at least daily by nearly every astronomer. It was conceived in 1987, and came on-line in 1992. It is a central engine of astronomy's large and complex information environment, linking together literature and data.

The ADS is in the process of a massive re-engineering. The prototype for the new system can be found at: http://adslabs.org/ui

4.6 How to communicate the data described in publications? The Dryad model

Todd Vision (University of North Carolina – Chapel Hill, US)

Of the tens of millions of research articles that have been published, the underlying data for validation and reuse are available for only small fraction. This compromises the quality and credibility of science. To realise a world in which the publication of research data is customary, it will be necessary to adopt a multifaceted strategy. This includes technological innovations in data repositories, alterations to the landscape of researcher incentives, experimenting with new models of sustainability, and exploring new roles for publishers, learned societies, and funders. Leveraging the close relationship between research data and scholarly publication lessens these challenges, and we are experimenting with such a model in Dryad, a repository for data associated with articles in the basic and applied biosciences.

4.7 More than just data!

Cameron Neylon (Rutherford Appleton Lab. – Didcot, GB)

Much of the discussion of enhanced research communication turns on the availability of digital assets, mostly data, but with an increasing emphasis on software and workflows as well, and the exploitation of these assets to provide a rich media experience, enhanced functionality and discoverability or other benefits of online interactions. Less explored are the issues of how the data was collected, what the relevant physical artefacts are, and how best to capture the information on this in a useful way. As is also the case for effective data and digital process publication this requires systems that help the user to think about publication earlier than is traditional but there are unique challenges to capturing the record of physical processes and in particular the physical world provenance trail that leads to the first relevant digital artefact. This means that effective laboratory recording systems that enhance communication as opposed to just record keeping need to be built and configured in a way that makes those recording processes easy, automatically captures records of physical and digital artefacts via data model that can deliver immediate benefits to the user, but also renders the ultimate aggregation and collation of records into a useful form for communication easy as well. These challenges are not yet sufficiently addressed by the tooling that supports the capture and communication of digital research artefacts and processes.

5 Overview of Talks. The Future

5.1 The Future. Or: What I would Like from Publications of the Future

Eduard H. Hovy (University of Southern California – Marina del Rey/ISI, US)

URL http://bit.ly/t6N2NI

This talks presents the overall vision of the enterprise, which it defines as "To improve the communication of knowledge between scholars using new informatics technology", and lists the general kinds of communicative services that a Publication of the Future (PoF) should provide. These include:

1. Better knowledge access
   - Using terminology standards
   - Automating access
2. Better knowledge communication
   - Reflecting the foundational theory and methodology
   - Contextualising the work in relation to current world
   - Using the best media at hand
3. Better knowledge verification/extension
   - Exposing the reasoning
   - Providing non-text info and tools

The talk illustrates each point with examples, taken from both the sciences and the humanities. It ends with a draft outline of the eventual report.

5.2 Introduction to the Future of Research Communication

Tim Clark (Harvard Medical School & Massachusetts General Hospital, US)

URL http://www.slideshare.net/twclark/dagstuhl-future-sesssion-intro-slides

Research Communication exists in a complex web of technology, information, people and activities. It is currently in a transitional state between print media and Web media. A number of problems are posed for its future development. These include research reproducibility and data provenance, interoperability, dealing with masses of data on previously unknown scales, algorithmic assistance to readers, and in general dealing with the issue of volume of publications, which is intractably large for even highly specialized disciplines.

Technological solutions alone will not be sufficient. The most productive solutions to these and other problems will adopt the "ecosystemic" perspective. They will emphasize the interaction of technology, information, and social formations in mutually beneficial ecosystems, or more correctly, "activity systems", in which value chains are built and sustained for participants from multiple interacting disciplines and communities.

5.3 Networked Knowledge

Stefan Decker (National University of Ireland – Galway, IE)

A new publishing paradigm as a social-technical system. A first approach to the necessary infrastructure for Networked Knowledge – initial ranking, abstractions and access mechanisms.

5.4 The Execution of Dave 2.0

David De Roure (University of Oxford, GB)

URL http://www.myexperiment.org/packs/206.html

What happens when there are millions and millions of executable papers, sitting there and executing away...? "Executable journals" are a step towards this vision – a world of inter-related executable papers, in an altered ecosystem of scholarly publishing with new intermediaries like observatories and a new role for existing intermediaries like libraries and publishers. What will that world be like? It will help us do science-on-demand ("press this button to re-run your thesis"), and equally the papers can process new data autonomously, generating new results which in turn get processed by other papers. You'll receive an email notification when the paper you wrote five years ago is re-run with new inputs from other

people's papers, and so will the people who used yours. Automated execution assists curation and indeed validation and quality checking – and whatever replaces peer review as we know it. Is this crazy or inevitable? The co-evolutionary design of the myExperiment website (http://www.myexperiment.org) for sharing computational workflows gives us as glimpse into this world of executable "Research Objects", which is being further developed under the Wf4Ever project.

5.5 "Towards Horizons 2020" — The Framework Programme for Research and Innovation 2014 to 2020 and Role of scientific data

Mike W. Rogers (European Commission Brussels, BE)

Europe's aim to be the leading knowledge based economy will be supported by the new Framework Programme. The development of the specific programme will be an outcome of intensive public debate and stakeholder participation, based on a number of guiding principles which are rapidly emerging after the first wave of consultations:

- Strong support for bringing research and innovation together in an integrated funding programme.
- Simplification is a key priority for all stakeholders.
- All stages in the innovation chain should be supported, with more attention for close to the market activities (e.g. demonstration, piloting).
- Continuity for the successful elements of current programmes, e.g. European Research Council, Marie Curie, collaborative research.
- EU funding should be tied closely to societal challenges and EU policy objectives (climate change, ageing, energy security, ...).
- More openness and flexibility is needed, less prescriptive calls, better use of bottom-up instruments (also in programme parts guided by clear policy objectives).

The presentation developed the rationales and scope of the various consultations in order to enable participants to better understand the future roadmaps for European Research Models where the connectivity from research to Innovation will be addressed systemically. More specifically, the current consultation on the future of Scientific Data was presented and a number of themes highlighted which the Workshop could develop as a core to its response to the European Commission, both as individuals, as representatives of organisations and as a body of expertise in its own right.

6 Working Groups

The seminar participants formed several working groups that tried to focus on various issues related to the future of research communication and e-scholarship. This section presents the notes of these groups. Note that since the working groups took place in parallel, there was no single terminology: for instance, digital artefact, research object, publication of the future are likely to have the same meaning. The seminar participants agreed to publish a white paper based on these notes (see Force11 white paper in Section 7).

6.1 Data

This group aimed at brainstorming on the main issues related to the creation and publishing of data. Below we include the list of main questions raised during the discussion. More notes can be seen at http://bit.ly/usiQOE.

- How much does the research domain matter when thinking about new publication forms?
- How do we effectively collaborate?
- How do we relate the kind of tool we have and integrate them so that the scientist can play with them?
  - We need to formulate use cases, we could generate a vision on what happens if we put all this together
- What we are already doing in this community to improve scientific publishing and what could we do next?
- How do we get researchers to buy into new ways of publishing?
  - Should we aim to be contagious? People can register and share their things
- Information and data curation
  - How can we maximise the input and first pass of information curation?
  - What is the role of curators to validate NLP results?
- Issues of time
  - How rapid should be the science production loop, from data to publication to science communication? There is a pre- and post- publication aspect, the quicker a publication can be devolved the faster is the impact on people citing that data.
- What would researchers need to know from others?
- What are the bottlenecks of open science?
  - Sharing is a bottleneck: scientists are not available to share they are fine to collaborate and publish but not sharing, because there is no recognition and there are potentially negative effects. There is no policing and penalty.
  - We need funding bodies and journals to penalise those who do not share
  - We can invent mechanism to detect who does not share, i.e if a pub derives form a work it is not collaborative, probably it does not connect to the other research artefacts out there
- We do not have a good value proposition: reservation for self use
- Knowing what I have in my lab
- What do scientists need?
  - Recovery and archive of data, plus access control to data
  - Productivity, I want to be helped into publishing more
  - How do we make the literature more effectively used
  - People do not like paper summarisation, they do not trust the conceptual model presented they think it is limited
  - There is no accreditation for doing annotation, knowledge curation or any king of paper summary
  - What are the incentives we propose for doing this activities?
- Information complexity: I want to be helped to read the right papers in the right (also interdisciplinary papers)
  - Would be good knowing what the most relevant paper are

6.2 Tools and Technologies

This group aimed at organising and predicting the requirements of tools and technologies for Scholarly Communications. This group was made up of: Carole Goble (chair), David De Roure, Anna De Liddo (notes), Phil Bourne, Paolo Ciccarese, David Shotton, Herbert Van de Sompel, Tim Clark, Gully Burns, Udo Hahns. Below we include the list of main discussion items. The list of tools and notes are available at

https://sites.google.com/site/futureofresearchcommunications/force11-tools-framework

and the participants hope to put there a systematic profiling of the tools later.

What are the communication artefact we use in science?

What are the communication functionalities and their integration?

What is the lifecycle of Digital Artefact?

The lifecycle of a Digital Artefact includes the following stages:

1. Registration
2. Certification
3. Archiving
4. Rewarding
5. Enactment of the Digital Artefact: presentations, videos
6. Writing
7. Discoursing
8. Reuse/reproduceable
9. Formal/informal
10. Granularity of publications

How would you alter these tools so that they may become more valuable for the publication lifecycle?

What are the main categories of tools for supporting the entire lifecycle of scientific publication?

Tools that deal with the Digital Artefact and that are used formally and informally to support the lifecycle. There is also another dimension that is the speed of production of Digital Artefact and their development, including the issue of granularity of Digital Artefact.

What are the main tools in place now to support the entire lifecycle of scientific communication?

1. Literature programming
2. Scientific publication
3. Spreadsheets
4. Reference management system
5. Web pages + Web sites
6. Powerpoint
7. Word/LaTeX Google docs
8. Supercomputing
9. Gmail
10. Digital library

11. Poster
12. Analysis workflow + R scripts, codes
13. Amazon for papers/books
14. Catalogues: s/w library, N/F, Yellow pages
15. YouTube
16. Recordings/broadcast/webinars of talks and presentations
17. Dropbox/SlideShare/Flickr/Twitter
18. Terminologies, thesauri, mapping, ontology
19. Search services
20. Analytical tools to survey the landscape, understand the science landscape, i.e mapping and research literature mapping (Compendium, Cohere, knowledge mapping tools)
21. Technologies thesauri
22. Hubs for communication: centres of communities (automated versions of it)
23. What are the more formal tools
24. EasyChair: Conference reviewing tools
25. Journals
26. Grant repositories/applications: generating documents
27. Database schema, data repositories
28. Google+, Facebook, social networks
29. Learned society
30. Conference call (Skype)
31. Publishers
32. Chat
33. Directories of WhoIs/yellow pages

How can these tools be categorised?

- Social Technologies
- Info tech-tool

Off the shelf: What is different in how those tools are used in scholarly communication, compared to other forms of informal communication?

What needs to be added to make of this tools recognised scientific tools: i.e., so that tweets on the last paper you published would be considered by your boss

- Twitter
  - Self-promotion
  - Conference reporting
  - Community intelligence
  - Data observations, cities sensors
  - Reluctant to negative critics
- Dropbox
  - View data in real-time
  - Easy data maintenance
  - Writing

What are the properties necessary to move a tool from formal to informal tool for scientific communication/publishing?

1. Citeability
2. Preservation
3. Highly shareable
4. Known provenance
5. Trustable
6. Accessibility
7. Stability
8. Granularity
9. Cost (or lack of it)
10. Speed
11. IP restrictions
12. Inherent rewardability
13. Annotatabilility
14. Protectability
15. Staking claims
16. Portability
17. Palpability
18. Easiness
19. Capacity
20. Multimediality

What are the categories of tools that are emerging?

1. Communication Instant Discourse
2. Training tools
3. Document composition, editing, authoring
4. Sensemaking
5. Scientific publication/research sharing
6. Preservation/storage
7. Presentation
8. Search tools
9. Digital artefact/ file sharing
10. Terminology services
11. Curation: metadata/indexing/ managing tools
12. Social
13. Certification tools and commenting
14. Execution tools

What are the Media Types of Digital Artefacts?

1. Image
2. Video
3. Text
4. Code
5. To be continued...

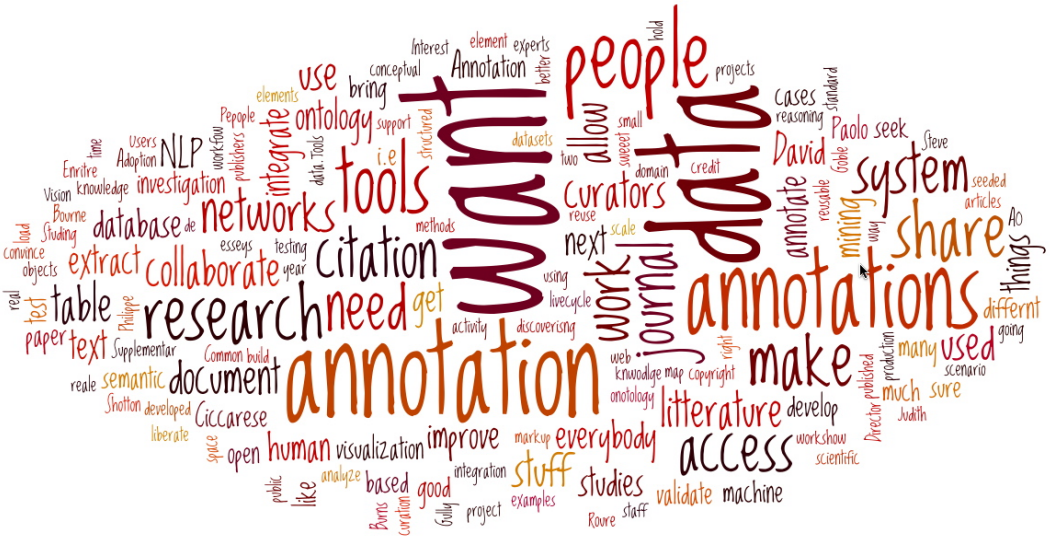

Figure 1 A Wordle map from the notes of day 1 represents the most frequent keywords mentioned in the discussion

Back office

1. Citation
2. Identifiers
3. Interoperability and best practices
4. Capability matrix: a map of the skills and tools we have in the group, for understanding, when I need something, what are the interoperability conventions between the tools
5. Machine actionability
6. Economic sustainability and community involvement (SWAT analysis)
7. Problem of VERSIONING at all layers (FLUX) – What is the CONTRACT OF INTEROPERABILITY SERVICES? What is the change protocol/standards of a tool? What is the contractual cleanliness and coherence of the tool? (Within the group)
8. Making use of cutting edge computer science technologies

6.2.1 Actionable recommendations to

6.2.1.1 Funders, policy makers and reviewers of projects

- Reward the funding of tweaks recomposition and interoperability of already existing software. Reward REUSE and REPURPOSE
- Proactively identify missing components and services that proposals should be focusing on. Assess risk. Do not leave all to market!
- Source Material:
  - Software sustainability
  - Putting all the recommendations in an e-infrastructure policy document
- Specific fund archives (AHRC counter example)
- Citation effort
- Best strategies models for assessing the success of scholarly communication
- Identify the obstacles to scholarly communication

6.2.1.2 Scholars

- Put your data in a open repository and cite it and include it in your CV
- Promote tools and propagate practice to scholars
- Get your colleagues to do the same
- Complain and engage in the battle (senior scholars to advocate and promote sharing and defend young scholars that do that by rewarding them for doing that)
- Enlightenment

6.2.1.3 Publishers

- No walled gardens
- Metadata/splash pages should be open including references
- Allow open annotation schemes and name entities access
- Enable citeability of components
- Provide APIs and encourage developers to build applications
- Provide a unified standards
- Exposing content for text mining
- Embrace linked data
- Expose item level download data

Following these recommendations will drive better bigger and access to scholarly contents.

6.2.1.4 Technology developers

- Place your software in the Force11 roadmap and framework at at https://sites.google.com/site/futureofresearchcommunications/force11-tools-framework
- Reuse of existing components and standards
- Collaborate to develop new components that do not exist yet
- Place your software in the value chain of improving research and science communication
- Encourage "enlightened self-interest" in your users

6.2.2 Vision: Making scholarship useful and usable

An interoperable serviced based ecosystem of sustainable core components as the basis for a healthy, innovative and vibrant market of interoperable and usable tools fit for scholarship in the 21st century. These tools and technologies will exploit the full potential of information and communication technologies to serve and not hinder scholarship, thus improving the quality and productivity and dissemination of research. This ecosystem will provide a basis for more rapid and cost-effective innovation of software for scholarship.

6.2.3 Questions raised during the discussion

- How much does the research domain matter when thinking about new publication forms?
- How do we collaborate effectively?
- How we relate the kind of tool we have and integrate them so that scientists can play with them?
- We need to formulate a use cases, we could generate a vision on what happen if we put all these things together
- What we are already doing in this community to improve scientific publishing and what we could do next?
- How do we maximise the input and first pass of information curation?

- What is the role of curators to validate NLP results?
- How do we have researchers buying in new ways of publishing?
- We should aim to be contagious? People can register and share their things
- Issues of time
- How rapid should the science production loop be, from data to publication to science communication?
- There are pre- and post- publication aspects, the quicker a publication can be devolved the faster is the impact on people citing that data.
- What do researchers would need to know from others?
- What are the bottlenecks of open science
- Sharing is a bottleneck: scientists are not available to share, they are fine to collaborate and publish but not share, because there is no recognition and there are potentially negative effects. There is no policy and penalty
- We need funding body and journals that penalise those who do not share
- We can invent mechanism to detect who does not share, i.e if a publication derives from a work which is not collaborative, probably, it does not connect to the other research artefacts out there
- We do not have a good value proposition: reservation for self use, knowing what I have in my lab
- What do scientists need?
- Recovery and archive of data, plus access control to data
- Productivity, I want to be helped into publishing more
- How do we make the literature more effectively used, and how we make people understandable and useful to them.
- People do not like paper summarisation, they do not trust the conceptual model presented they think it is limited
- There are no credits for doing annotation, knowledge curation or any kind of paper summary
- What are the incentive we propose for doing this activities?
- Information complexity I want to be helped to read the right papers in the right (also interdisciplinary thing)
- Would be good knowing what the most relevant paper are.

6.3 Business models for the research communications in the future

This group aimed at brainstorming on possible business models for the research communication, taking into account the changes happening in scholarly publishing nowadays. This group included the following participants: Bradley P. Allen (notes), Aliaksandr Birukou, Philip E. Bourne, Leslie Chan, Olga Chiarcos, Robert Dale, Eve Gray, Paul Groth, Ivan Herman, Eduard H. Hovy, Fiona Murphy, David S. H. Rosenthal (chair), Jarkko Siren. Below we reproduce the summary of the discussion (also found at http://bit.ly/tyaWcL). More notes can be found at http://bit.ly/usiQOE.

Building a sustainable approach to research communications of the future will require the exploration of the space of potential business models. By business model, we mean a conceptual description of how an organisation provides value to customers — and gets paid for doing so. This last consideration of describing how the money flows is key to

understanding and resolving the sustainability and access issues that dog the ecosystem of researchers, institutions, publishers and funding agencies today.

In keeping with the ideas discussed at the Workshop about the future of research communications, the group focused on the notion of the research object as the contained of information being communicated. A research object is composed of one or more of the following types of sub-objects:

- Documents (textual, multimedia, pictures, etc)
- Experimental data
- Methods and procedures
- Relationships among constituents
- Context metadata
- Asset metadata
- Relational metadata
- Provenance

The group used the Business Model Generation [2] methodology to describe the space of potential business models for research communication in the future. Specifically we developed a set of optional choices for elements of the nine components of a Business Model Canvas [1]. These are:

- Value Propositions: What value is being delivering to the customers?
- Customer Segments: Who pays for that value?
- Channels: How is this value delivered to the paying customers?
- Customer Relationships: How is the relationship with the customers managed, and by whom?
- Revenue Streams: In what ways do customers pay us for this value, and optionally, how much?
- Key Resources: Who and/or what is required to build and operate the systems and organisations need to deliver the value?
- Key Activities: What tasks need to performed to deliver the value to the customers?
- Key Partners: Who are the key partners needed for the organisation to be able to deliver value?
- Cost Structure: What costs does the organisation incur to operate and deliver the value?

Table 1 illustrates the sketch of the business model designed during the working group.

Table 1 A sketch of the business model in Business Model Canvas format.

Business Model Component	**Possible Values**
Value Proposition	Seamless Management of Research Objects • Discovery • Preservation • Version control • Exploration & Integration • Metrics • Review, Evaluation, Annotation Seamless Management of Researchers • Reputation • ID

	• Profile • Aggregation, Syndication • Personalisation
Customer Segments	One of: • Creators • Funding agencies • Consumers (Researchers, Public, Industry) • Evaluators (Reviews, Tenure, Regulators) • Advertisers • Sponsorship

Customer Relationships
(who is accountable to the customer)
	One of:
	• Institutional support organisation (universities, research organisations,...)
	• Independent non-profit support organisation
	  • learned societies
	  • foundations
	  • self-organising communities
	• For profit publishers

Channels	One of:
	• Software-as-a-service platform
	• Direct software distributor
	  • shrink-wrap
	  • open source

Revenue Streams	Metered access to objects • "All you can eat" (one time, recurring, ...) • per object • per use Payment for service & support • subscription (one time, recurring, ...) • pay per call Software purchase
Key Activities	All of: • Build and maintain platform • Run platform • Develop and support communities
Key Resources	One or more of: • Platform/software developers • Operations staff • Content experts • Community managers • Marketing, PR • Business development
Key Partners	One or more of: • Learned societies • Funding agencies • Institutions (universities, research institutes, ...)

	• Subject matter experts • General public (crowdsourcing, citizen science, professional/amateur collaboration) • Government
Cost Structure	All of: • Hardware • Software • Network • Power • Staff costs • Market communications
Business Model Component	**Possible Values**

6.4 Assessment and Impact

This group aimed at identifying the critical issues pertaining to the research assessment and impact. The following people took part in this group: Eve Gray, Laura Czerniewicz, Ivan Herman (chair), Herbert van den Sompel (notes), Michael Kurtz, Jarkko Siren, Peter van den Besselaar, Anita de Waard. Below we include the list of main issues discussed. More notes can be seen at https://sites.google.com/site/futureofresearchcommunications/contributions-1/contributions.

6.4.1 Opening statements

- Current assessment mechanism is counter productive to scholarly communication. Need to make policy makers realise and accept that. Only formal citations count. Not other impact.
- What is impact assessment? Assess based on what? Do we need assessment of individuals?
- Impact factor doesn't work for across disciplines. Metrics on people or on artefacts?
- Perspective should be about value and how value relates to business and impact. Measure value! But how?
- Scholarly communication system is skewed by impact assessment as it is.
- The system is counter productive. But measures are essential because of assessing individuals, setting funding policy. Question how to come up with other metrics that can be generated in an open and scaleable way. Question how to get those metrics accepted.

6.4.2 Questions raised

- Do institutions, funding agencies base decisions on impact factor? Not element in decision whether a project gets funded, but does it play role in setting funding policies?
- Do we need to also talk about e.g. service to community as part of assessment? Is that science communication?
- What are metrics to assess research communication system, rather than to assess individuals?

6.4.3 Issues

- Need a multidimensional metrics model to count various things. If possible, the model should apply across disciplines. Simplicity of metric is important.
- What are those new dimensions? Is Altmetrics an answer?
- Why current system is broken?
  - Africa can not publish in ranking journals even if paper is about millions of people dying of some disease
  - The way we conducts science has changed so fundamentally that a metrics mechanism that ignores this change is totally passe
  - Real impact is manifested in different ways now (e.g., we know who the core players are in a scientific community and that is not based on "objective" metrics)
- The stellar researchers are known by their community. All the others not necessarily. Metrics can help.
- Innovations systems thinking. Research => Patent => Commercial. Need to change that thinking.
- Accessibility of metrics (or data from which to derive metrics) across systems is big issue, e.g. download data not consistently available; API to obtain metric only allows limited amount of calls a day.
- Author disambiguation – ORCID?
- Reputation management
- Need to define output types and metrics for output types

6.4.4 Possible dimensions

- How do we measure how research contributes to society (e.g., development goal in Africa)
- Netherlands: "evaluate research in context" effort. Quality of communication between research and community at large determines societal impact.
- local versus global impact
- economical impact
- quality of communication to general public
- measures depend on goals. In many cases citations are good. But, for example, in nursing, readership becomes important.
- need to be able to get at metrics otherwise you have done nothing
- download counts (better to measure social impact). Can be gamed. Can use under right conditions.
- crowd sourcing evaluation (e.g. Faculty of 1000)
- used for teaching (knowledge with of being transferred)
- used in lectures
- general level of reuse
- openness

The problem we see with Impact Factor and other simple metrics are individually taken with grain of salt. But if we would use multiple dimensions we might get a more just system. Decisions makers may choose which dimensions to use.

References

1 Wikipedia: http://en.wikipedia.org/wiki/Business_Model_Canvas, last checked 2011-11-04.
2 Alexander Osterwalder and Yves Pigneur, *Business Model Generation*, John Wiley and Sons, 2010.

7 Relevant links

- A huge collection of relevant links is maintained at http://bit.ly/tFQnkL.
- Future of Research Communication and e-Scholarship (FORCE11) website: http://force11.org/
- Force11 Manifesto *"Improving Future Research Communication and e-Scholarship"*: http://force11.org/node/1688

8 Agenda

time	August 15	August 16	August 17	August 18
8:00 – 9:00		Breakfast	Breakfast	Breakfast
9:00 – 10:30		The Past: Herman (Chair) De Waard, Buckingham Shum, Rosenthal	The Future: Hovy, Clark, Decker, De Roure, Rogers	Presentations Working groups 1, 2, 3
10:30 – 10:45		Break	Break	Break
10:45 – 12:15		The Present: Shotton, Groth, Murphy, Bourne, Kurtz	Working groups II	Presentations Working groups 4,5
12:15 – 13:00		Lunch/Demo's	Lunch/Demo's	Lunch
13:00 – 14:00	Arrive, register, settle in	Network, email, leisure; demo's	Network, email, leisure; demo's	Summary of items from working group, Action items/Calendar for next steps
14:00 – 15:30		The Present: Vision, Neylon; Planning working groups	The Future: Hovy, Clark	
16:00 – 18:00	Welcome/Introductions	Working groups I	Working groups III	Departure
18:00 – 19:30	Dinner	Dinner	Dinner	
19:30 – 20:30	Discuss goals for the week, divide into Working groups	Recap; touch base Working groups, settle questions	Recap day; plan calendar/tasks after Dagstuhl	
20:30	Wine and cheese (and music)	Wine and cheese (and music)	Presentation of working groups – prequel	

Working groups:

number	planned name	final name
1	Research data & code	Data
2	Assessment and impact	Assessment and impact
3	New forms and tools	Tools and technologies
4	Business models	Business models for the research communications in the future
5	Social platform & continuity	*this group was merged with other groups*

Participants

- Bradley P. Allen
Elsevier – Manhattan Beach, US
- Aliaksandr Birukou
CREATE-NET – Povo, Trento, IT
- Judith A. Blake
The Jackson Laboratory – Bar Harbor, US
- Philip E. Bourne
UC San Diego, US
- Simon Buckingham Shum
The Open University – Milton Keynes, GB
- Gully Burns
Univ. of Southern California – Marina del Rey, US
- Leslie Chan
University of Toronto, CA
- Olga Chiarcos
Springer-Verlag – Heidelberg, DE
- Paolo Ciccarese
Harvard University, US
- Timothy W. Clark
Harvard Medical School & Massachusetts General Hospital, US
- Laura Czerniewicz
University of Cape Town, ZA
- Robert Dale
Macquarie University, AU
- Anna De Liddo
The Open University – Milton Keynes, GB
- David De Roure
University of Oxford, GB
- Anita De Waard
Elsevier Labs – Jericho, US
- Stefan Decker
National University of Ireland – Galway, IE
- Alex Garcia Castro
Universität Bremen, DE
- Carole Goble
University of Manchester, GB
- Eve Gray
University of Cape Town, ZA
- Paul Groth
VU University Amsterdam, NL
- Udo Hahn
Universität Jena, DE
- Ivan Herman
W3C/CWI – Amsterdam, NL
- Eduard H. Hovy
Univ. of Southern California – Marina del Rey/ISI, US
- Michael J. Kurtz
Harvard-Smithsonian Center for Astrophysics, US
- Fiona Murphy
Wiley-Blackwell, UK
- Cameron Neylon
Rutherford Appleton Lab. – Didcot, GB
- Steve Pettifer
University of Manchester, GB
- Mike W. Rogers
European Commission Brussels, BE
- David S. H. Rosenthal
Stanford University Libraries, US
- David Shotton
University of Oxford, GB
- Jarkko Siren
European Commission Brussels, BE
- Herbert van de Sompel
Los Alamos National Lab., US
- Peter van den Besselaar
Free Univ. – Amsterdam, NL
- Todd Vision
University of North Carolina – Chapel Hill, US