Force11 White Paper:
Improving The Future of Research Communications and e-Scholarship

February 19, 2012

Editors: Philip E. Bourne[a], Tim Clark[b], Robert Dale[c], Anita de Waard[d], Ivan Herman[e], Eduard H. Hovy[f], and David Shotton[g]

Contributors: Bradley P. Allen[d], Aliaksandr Birukou[h], Judith A. Blake[i], Philip E. Bourne[a], Simon Buckingham Shum[j], Gully A.P.C. Burns[f], Leslie Chan[k], Olga Chiarcos[l], Paolo Ciccarese[b], Tim Clark[b], Laura Czerniewicz[m], Robert Dale[c], Anna De Liddo[j], David De Roure[g], Anita De Waard[d], Stefan Decker[n], Alex Garcia Castro[o], Carole Goble[p], Eve Gray[m], Paul Groth[q], Udo Hahn[r], Ivan Herman[e], Eduard H. Hovy[f], Michael J. Kurtz[s], Fiona Murphy[t], Cameron Neylon[u], Steve Pettifer[p], Mike W. Rogers[v], David S. H. Rosenthal[w], David Shotton[g], Jarkko Siren[v], Herbert van de Sompel[x], Peter van den Besselaar[q] and Todd Vision[y]

Affiliations: (a) University of California at San Diego; (b) Harvard Medical School; (c) Macquarie University; (d) Elsevier Laboratories; (e) Centrum voor Wiskunde en Informatica, Amsterdam; (f) University of Southern California; (g) University of Oxford; (h) CREATE-NET; (i) The Jackson Laboratory; (j) The Open University; (k) University of Toronto; (l) Springer-Verlag; (m) University of Cape Town; (n) National University of Ireland, Galway; (o) Universität Bremen; (p) University of Manchester; (q) Vrije Universiteit Amsterdam; (r) Universität Jena;(s) Harvard-Smithsonian Center for Astrophysics; (t) Wiley-Blackwell; (u) Rutherford Appleton Laboratory; (v) European Commission Brussels; (w) Stanford University; (x) Los Alamos National Laboratory; (y) University of North Carolina at Chapel Hill

Abstract

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. 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 highlights the findings of the Force11 workshop on the Future of Research Communication and e-Scholarship held at Schloss Dagstuhl, Germany, in August 2011: it summarizes a number of key problems facing scholarly publishing today, and presents a vision that addresses these problems, proposing concrete steps that key stakeholders can take to improve the state of scholarly publishing. More about Force11 can be found at `http://www.force11.org`. This White Paper is a collaborative effort that reflects the input of all Force11 attendees at the Dagstuhl Workshop [1], and is very much a living document [2] . We see it as a starting point that will grow and be updated and augmented by individual and collective efforts by the participants and others. We invite you to join and contribute to this enterprise.

---

[1]`https://sites.google.com/site/futureofresearchcommunications/home/attendees`

[2]Citation: Bourne P, Clark T, Dale R, de Waard A, Herman I, Hovy E and Shotton D (eds.), on behalf of the Force11 community (2011). Force11 White Paper: Improving the Future of Research Communication and e-Scholarship. 27 October 2011. Available from `http://force11.org/` Copyright: 2011 The authors. License: This is an open-access article distributed under the terms of the Creative Commons Attribution License (v3.0, unported: `http://creativecommons.org/licenses/by/3.0/`), which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

About This Document: This document contains five sections. Section 1 presents our vision of the future of scholarly publishing. In Section 2, we outline six key problems that prevent scholarly communication from achieving its full potential. Section 3 contains six specific recommendations for actions to address these problems. Section 4 offers a dynamic list of pointers to relevant research reports and related projects. Finally, in Section 5 we describe what we are doing to implement these recommendations.

The problems and recommendations we perceive can be grouped into two groups, each containing three principal themes:

- Themes 1–3 concern the *format and technologies* of scholarly publication: how scholarly data, information, and knowledge are (or could be) represented; how readers, users, authors, editors and computers can interact with these representations; and how different knowledge representations could be combined, queried, stored and otherwise treated.

- Themes 4–6 concern the enterprise of scholarly publishing, including business models and the attribution of credit. In these sections we discuss how scholarship is evaluated, accredited and monetized; current and new models and modes of assigning copyright and intellectual property rights; the financial aspects of scholarly publishing; and the mechanisms for assessing the quality and value of researchers and their research outputs, and of attributing credit and worth to them.

The problems relating to these six themes are described in Section 2, while our recommendations for their solutions are described in Section 3. These problems and recommendations are summarized in the following table.

Problems	Recommendations
Formats and Technologies	
2.1 Existing formats needlessly limit, inhibit and undermine effective knowledge transfer	3.1 Rethink the unit and form of the scholarly publication
2.2 Improved knowledge dissemination mechanisms produce information overload	3.2 Develop tools and technologies that better support the scholarly lifecycle
2.3 Claims are hard to verify and results are hard to reuse	3.3 Add data, software, and workflows into the publication as first-class research objects
Business Models and Attribution of Credit	
2.4 There is a tension between commercial publishing and the provision of unfettered access to scholarly information	3.4 Derive new financially sustainable models of open access
2.5 Traditional business models of publishing are being threatened	3.5 Derive new business models for science publishers and libraries
2.6 Current academic assessment models don't adequately measure the merit of scholars and their work over the full breadth of their research outputs	3.6 Derive new methods and metrics for evaluating quality and impact that extend beyond traditional print outputs to embrace the new technologies

1   Our Vision

A dispassionate observer, perhaps visiting from another planet, would surely be dumbfounded by how, in an age of multimedia, smartphones, 3D television and 24/7 social network connectivity, scholars and researchers continue to communicate their thoughts and research results primarily by means of the selective distribution of ink on paper, or at best via electronic facsimiles of the same.

Modern technologies enable vastly improved knowledge transfer and far wider impact. Freed from the restrictions of paper, numerous advantages appear. Communication becomes instantaneous across geographic boundaries. Terms in electronic documents may be automatically disambiguated and semantically defined by linking to standard terminology repositories, allowing more accurate retrieval in

searches; complex entities mentioned in documents may be automatically expanded to show diagrams or pictures that facilitate understanding; citations to other documents may be enhanced by summaries generated automatically from the cited documents. Documents may be automatically clustered with others that are similar, showing their relationship to others within their scholarly context, and their place in the ongoing evolution of ideas. Ancillary material that augments the text of the scholarly work may be linked to or distributed with the work; this may include numerical data (from experiments), images and videos (showing procedures or scenarios), sound recordings, presentational materials, and other elements in forms of media still on the horizon. Extracts and discussions of scholarly work on social media such as blogs, online discussion groups and Twitter may greatly broaden the visibility of a work and enable it to be better evaluated and cross-linked to other information sources. A broad range of recent technological advances provide increasingly diverse and powerful opportunities for more effective scholarly communication; we need to grasp the opportunities and make these possibilities realities.

We see a future in which scientific information and scholarly communication more generally become part of a global, universal and explicit network of knowledge; where every claim, hypothesis, argument— every significant element of the discourse—can be explicitly represented, along with supporting data, software, workflows, multimedia, external commentary, and information about provenance. In this world of networked knowledge objects, it would be clear how the entities and discourse components are related to each other, including relationships to previous scholarship; learning about a new topic means absorbing networks of information, not individually reading thousands of documents. Adding new elements of scholarly knowledge is achieved by adding nodes and relationships to this network. People could contribute to the network from a variety of perspectives; each contribution would be immediately accessible globally by others. Reviewing procedures, as well as reputation management mechanisms, would provide ways to evaluate and filter information. This vision moves away from the Gutenberg paper-centric model of the scholarly literature, towards a more distributed network-centric model; it is a model far better suited for making knowledge-level claims and supporting digital services, including more effective tracking and interrogation of what is known, not known, or contested.

To enable this vision, we need to create and use new forms of scholarly publication that work with reusable scholarly artifacts. Two principal aspects can be distinguished. First, we need to revise the artifacts of communication. As a starting point, our vision entails creating a new, enriched form of scholarly publication that enables the creation and management of relationships between knowledge, claims and data. It also means the creation of a knowledge infrastructure that allows the sharing of computationally executable components, such as workflows, computer code and statistical calculations, as scientifically valid content components; and an infrastructure that allows these components to be made accessible, reviewed, referenced and attributed. To do this, we have to develop best practices for depositing research datasets in repositories that enable linking to relevant documents, and that have high compliance levels driven by appropriate incentives, resources and policies. In addition, for scientific domains, the new forms of publication must facilitate reproducibility of results, which means, at least for *in silico* research, the ability to preserve and re-perform executable workflows or services. This will require the ability to re-construct the context in which these objects were executed, which may well contain or reference other executable objects as well as data objects that may evolve through time. In this way, the content of communications about research will follow the same evolutionary path that we have seen for general web content: a move from the static to the increasingly dynamic.

With all this, we do recognise the importance of the peer-reviewed journal article as a primary dissemination channel and public record of new research results, since it uniquely provides a dated version of record of the authors' views at the time of publication, and as such becomes an immutable part of the scientific record. But even here, with this the most traditional of scholarly communication media, we can with existing technologies provide immediated improvements: semantic enhancements to the text; greater interactivity with tables and figures; access to the data within articles in actionable form; data fusions (mashups) with data from other sources, for example Google Maps, where appropriate; direct citation of and links to underlying datasets stored in databases and data repositories; and the open publication in machine-readable form of both the full bibliographic record for the article and also the citation information contained within the article's reference list, encoded using appropriate ontologies, so that these basic facts can enter the web of linked open data `http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData` [Shotton, 2011].

The second component of our vision requires changes to the complex socio-technical scholarly and commercial ecosystem. In particular, to obtain the benefits that networked knowledge promises, we have to put in place reward systems that encourage scholars and researchers to participate and contribute. We need to acknowledge the fact that notions such as journal impact factor are poor surrogates for

measuring the true impact of scholarship, and are increasingly irrelevant in a world of disaggregated knowledge units of vastly varying granularity; and we need to derive new mechanisms that allow us more accurately to measure true contributions to the ongoing enterprise of augmenting the world's store of knowledge. The business models that are currently driving scholarly publishing, which rest mainly on libraries buying access rights to digital journals from publishers, are clearly no longer adequate to support the rich, variegated, integrated and disparate knowledge offerings that new technologies enable, and that new scholarship requires. In a collaboration involving scholars, publishers, libraries, funding agencies, and academic institutions, we need to develop models that can enable this exciting future to develop, while offering sustainable forms of existence for the constituent parties, although perhaps not in their present states.

If we get this right, the potential is immense. The changes we envisage pave the way for a revolution in the manner in which research is carried out and communicated, leading to significant improvements in scholarly productivity and quality, and enhanced transparency that can only increase the public's trust in the value of science. Similar benefits apply to scholarship in the arts and humanities.

These developments bring advantages for many parties:

- **For scholars** (also in their roles as authors, editors and reviewers) the benefits are better communication of knowledge: easier transmission of information from its creators or discovers (the producers), in more forms using richer media, permitting easier, faster and deeper interpretation of the information by the consumers (other scholars, students and their teachers, government and non-governmental agencies, industry, the media, and society at large). At the same time, these new and enhanced forms of communication will enable more accurate evaluations of the quality and the impact of scholars' work, facilitating better promotion evaluations and proposal assessments.

- Similarly, **for decision makers and managers**, the new communicative forms mean that the impacts and effects of scholarly communications, and hence of their authors, can more easily be tracked and evaluated.

- For **research funders**, enhanced communications will enable more accurate overviews of the size, direction and importance of each stream of research, and permit quicker determination of the quality of the work cited in grant proposals. But these advances mean that established practice will need to change.

- For **librarians and archivists**, while online accessibility will mean that traditional library holdings become less important, the archiving, updating and maintenance of digital data and software will increase in importance. Adapting to these changes will bring about new modes of service to users.

- Similarly, **for publishers**, the traditional functions of manuscript compilation and distribution will change radically, while quality control, access facilitation, new modes of aggregation, and the standardization, maintenance, and support of knowledge access technologies become more important. Providing these services will allow publishers successfully to face the challenges of free access to published research that is being ushered in by the open access movement.

2 Problem: The Growing Problems of Outdated Communication

We are a long way from achieving this vision today. As noted above, the impediments exist primarily in two dimensions: we have to change the *nature of the formats and technologies of communication*, that underpins the world of scholarly publishing, and we have to change the *social ecosystem of communication* that has grown up around the existing technologies. We review the key issues in these two areas in turn.

Problems with Current Formats and Technologies

2.1 Existing Formats Are Not Tailored for Knowledge Transfer

Scholarly communications are, at this mid-point in the digital revolution, in an ill-defined transitional state—a 'horseless carriage' state—that lies somewhere between the world of print and paper and the world of the web and computers, with the former still exercising significantly more influence than the latter. However, the recent development of new media and communicative possibilities using information technology, and the need to communicate and comprehend increasing amounts of additional information such as numerical and multimedia data, make the traditional forms inadequate. Continued reliance

on paper documents and their electronic shadows make it very difficult or impossible to incorporate massive amounts of data, moving images or software; there is simply no natural way to associate such ancillary information 'into' the traditional publication. Additionally, any software-based text mining or information extraction procedures require that paper-based information first be converted into machine-tractable form and made freely available for such mining.

2.2 The Ever-Increasing Problem of Information Overload

Scholars have experienced information overload for more than a century [Vickery, 1999] and the problem is just getting worse. Online access provides much better knowledge discovery and aggregation tools, but these tools struggle with the fragmentation of research communication caused by the rapid proliferation of increasingly specialized and overlapping journals, some with decreasing quality of reviewing [Schultz, 2011].

2.3 Verifying Claims and Re-using Results

Most types of scholarship involve claims, and all sciences and many other fields require that these claims be independently testable. Good results are often re-used, sometimes thousands of times. But actually obtaining the necessary materials, data or software for such re-use is far harder than it should be. Even in the rare cases where the data are part of the research communication, these are typically relegated to the status of 'supplementary material', whose format [Murray-Rust, 2007] and preservation [Rosenthal and Reich, 2010] are inadequate. Sometimes the data are archived in separate data repositories that offer a more secure long-term future. But in such circumstances efforts need to be made to ensure that their links to the relevant textual research communications are explicit, robust and persistent. At present it is difficult for a scholar easily and sustainably to record the data on which the work is based in a form that others can absorb and use, and to maintain links to the associated textual publication.

Problems With Business and Assessment Models

2.4 Next-generation Tools Require Unfettered Resource Access

Currently, a large and active movement of professionals and students, including data curators, are providing services intended to improve the effectiveness of scholarly communication, and thereby the productivity of researchers; these entail digging facts out of textual publications and presenting them in machine-readable actionable form. The need for much of this expensive manual effort would be reduced if authors were to provide the relevant metadata at the time of publication. These extraction processes are increasingly being performed by automated text mining and classification software. However, because the source material is usually copyrighted, and these rights are distributed across a large number of publishers, the service providers are forced to negotiate individual contracts with each publisher, which is extremely wasteful of time and resources. To reduce this burden, some research funders are increasingly mandating that research results of all types be made openly available. However, this results in a confusing world where some publications are immediately and freely available and others on the same topic are not.

   A related problem is the effect of the web as the medium for scholarly communication, since it is ending the role of local library collections. Libraries and archives have been forced to switch from purchasing copies of the research communications of interest to their readers, to leasing web access to the publishers' copies, with no assurance of long-term accessibility to current content if future subscriptions lapse. Bereft of almost all their original value to scholars, libraries are being encouraged to both compete in the electronic publishing market and to take on the task of running 'institutional repositories', in effect publishing their scholars' data and research communications. Though both tasks are important, neither has an attractive business model. Re-publishing an open access version of their scholars' output where research is published in subscription-access journals may seem redundant, but it is essential if the artificial barriers that intellectual property restrictions have erected to data-mining and other forms of automated processing are to be overcome [Hargreaves, 2011].

2.5   Traditional Publishing Models Are Under Attack

Academic publishers have been slower to encounter, but are not immune from, the disruption that the internet has wrought on other content industries[The Economist, 2009]. The academic publishers' major customers, academic libraries, are facing massive budget cuts [Kniffel and Bailey, 2009], and so are unlikely to be a major source of continued revenue. The internet has greatly reduced the costs of publishing, new players (such Google and other software companies) have appeared in the market, and legislative and funding bodies are actively addressing issues of free access to data and text [Hargreaves, 2011]. The advent of the internet has greatly reduced the monetary value that can be extracted from paper-based academic content, and science publishers, who have traditionally depended on extracting this value, face a crisis, since their old business models are suffering disruption. Conversely, the internet permits the creation of new added-value services relating to search, semantics and integration that present exciting new commercial opportunities. Clearly the scholarly publishing industry needs to engage in discussions with different partners within the value chain, if it is to be included in the development of the new standards, services, business models, metrics/analysis, legislation, knowledge ecosystems and evaluation frameworks that the internet now makes possible, rather than being supplanted by new agile startups that have the ability to adapt more swiftly.

The software developers who build the current research informatics infrastructure are also very aware of the shortfalls and hindrances generated by today's fragmented development efforts. The problems here can be attributed to a number of elements. First, heterogeneous technologies and designs, and the lack (or sometimes the superfluity!) of standards, cause unnecessary technical difficulties and directly affect integration costs. Second, a complex landscape of intellectual property rights and licensing for software add legal concerns to developers' requirements. Third, research software developers typically work in a competitive environment, either academic or commercial, where innovation is rewarded much more highly than evolutionary and collaborative software reuse. This is especially true in a funding environment driven by the need for intensive innovation, where reusing other peoples' code is a likely source of criticism. Finally, even under optimal technical conditions, it is still challenging for software programmers to understand what components are the most appropriate for a given challenge, to make contact with the correct people to facilitate the construction of tools, and to work within distributed teams across groups to build high-quality interoperable software. The impact of these tools is, far too often, solely based on how immediately useful they will be to researchers themselves, with no thought for the wider community.

Thus changing roles and business models form an immense challenge for libraries, publishers and software developers. The only fruitful way forward, we firmly believe, will be for all parties collaborating to build new tools that optimally support scholarship in a distributed open environment. Only by creating a demonstrably better research environment will we convince the entire system of scholarly communication and merit assessment to adopt new forms and models.

2.6   Current Assessment Models Don't Measure Merit

Not only are the products of research activity still firmly rooted in the past, so too are our means of assessing the impact of those products and of the scholars who produce them. For five decades, the impact of a scholarly work—an entity that is already narrowly defined, in the sciences as a journal article, and in the humanities as a monograph—has been judged by counting the number of citations it receives from other scholarly works, or, worse, by attributing worth to an individual's work based solely on the overall impact factor of the journal in which it happens to be published. We now live in an age in which other methods of evaluation, including article-level usage metrics, blog comments, discussion on mail lists, press quotes, and other forms of media, are becoming increasingly important reflections of scholarly and public impact. Failure to take these aspects into account means not only that the impact and/or quality of a publication is not adequately measured, but also that the current incentivization and evaluation system for scholars does not relate well to the actual impact of their activities.

3   Strategies for Change

Mirroring our identification of the six impediments to our vision that lie in the two dimensions of technology and society, we here make specific recommendations for change in these two dimensions.

New Publication Formats and Tools

3.1   Rethink the unit and form of the scholarly publication: the Research Object

At the foundation of any change is the infrastructure to support that change. One must no longer think of the journal article or research paper as the standard unit of currency by which knowledge is exchanged. Now it is but one among many forms. In the most generic sense, the new form of knowledge exchange centers on the *research object* [De Roure and Goble, 2009, Bechhofer et al., 2010], a container for a number of related digital objects—for example a paper with associated datasets, workflows, software packages, etc., that are all the products of a research investigation and that together encapsulate some new understanding. Publishing of research objects is not necessarily publishing as we know it today, achieved by the same mechanisms as used for traditional scholarly articles. It consists of providing free and open access to the component parts of the research object, that may or may not have been individually reviewed by others either pre- or post-publication.

Arriving at a suitable definition of research objects requires work on standards and provenance, and conformance to general principles, some of which are suggested here:

- Support for multiple media types—text, images, podcasts, videos, etc.

- Recognition that raw and derived data, data processing procedures, computational models, experimental protocols and workflows all need to be preserved as part of the research object, and shared publicly.

- Support for access to content at varying granularities of detail.

- Support for the automatic extraction of information from research objects at these varying granularities, and its integration with third-party information.

- Support for uniquely identifying all elements of the research object.

- Support for both human and machine access, including access by disabled humans.

- Support for existing and emerging web and semantic web standards surrounding data representation and linking.

- Inclusion of social media as legitimate components within the world of the scientific discourse.

The research object per se does not necessarily capture the processes by which research leads to new knowledge. There is a temporal aspect to research and the scholarly lifecycle that also needs to be recorded, either within research objects or between research objects, and that should also be capable of being reproduced.

Developing the tools to support these changes, if undertaken from scratch, would be an immense undertaking. Thus, where possible, existing tools should be adapted and integrated within the new open infrastructure. Several classes of tools that exist and could be considered as components for this infrastructure are detailed in the "Tools" section of the Force11 web site (`http://force11.org`).

What is happening now?

The following are examples of technological changes associated with new forms of scholarship.[3]

- Hypothesis/claim-based representation of the rhetorical structure of a scientific paper [de Waard, 2009].

- Modular formats for science publishing [de Waard, 2010].

- Developments of metadata standards and ontologies for describing publishing activities and publications, for characterizing citations between them, for identifying their structural and rhetorical components, and for describing discourse elements within the text.

- Semantic publishing initiatives and other enriched forms of publication.

---

[3]Readers should also consult our online collection of links to related activities and examples at `https://sites.google.com/site/futureofresearchcommunications/links/links`.

What are the next steps?

Change is likely to occur gradually through a series of incremental steps, most of which will not be driven by the technology. Rather, the technology should respond to the recognized requirements of scientists for improved dissemination, reproducibility, recognition, etc. These requirements need to be assessed and formalized. The very existence of Force11 is an acknowledgement of the need for changes, but these changes need to be quantified and specifications drawn up for their solution.

3.2 Develop tools and technologies that better support the scholarly lifecycle

What is happening now?

As scholarship in all fields increasingly becomes undertaken online, new tools and technologies are required to support the whole scholarly lifecycle from initial hypothesis to results publication. We are already seeing:

- the emergence of workflow systems;

- the emergence of data repositories within which datasets have globally unique identifiers and explicit links to journal articles, which by necessity provide some form of attribution and provenance information;

- the emergence of citation ontologies and corpora of open citation data;

- the emergence of software repositories with good versioning support; and

- the increasing use of online services for collaborative work: file exchange services such as Dropbox, collaborative note-taking environments such as EtherPad, and collaborative authoring environments such as Google Docs.

Nevertheless, these systems are acknowledged to be inadequate and cumbersome in their use. We require:

- better systems to permit collaborative work by geographically distributed colleagues;

- better systems to permit collaborative writing, with fail-safe versioning;

- better tools for richer interactive data and metadata visualization, enabling dynamic exploration; and

- easier data publication mechanisms, including better integration with data acquisition instrumentation, so that the process becomes automated.

What are the next steps?

To begin with, we want the scholarly community to be concerned with modes of archiving and sharing papers, data, workflows, models and software, and with the creation of research objects as part of their daily research routines. Other questions to explore include:

- What are the features of the research lifecycle and how do they impact the contents of and relationships between the artefacts that constitute digital research objects?

- How can existing tools be adapted to fit the specific workflow requirements of different scholarly domains?

- How can these tools be optimally integrated with environments to read, write and edit publications, and to create and evaluate research data?

3.3 Integration of datasets, software, mathematical models and workflows into publications as first-class research objects

Clearly, data in 21st Century science are almost always subjected to transformation by software, that undertakes either individual transformation processes, or links these into processing workflows. A full record of the research undertaken requires preservation of these processing steps and software tools employed, in addition to the datasets upon which they acted.

What is happening now?

Exemplars of repositories for research datasets, software and workflows include Dataverse [King, 2007, Crosas, 2011], the Dryad Data Repository [White et al., 2008, Greenberg, 2009a], and myExperiment, a social network relating to workflows [Goble and De Roure, 2007, De Roure et al., 2009].

What are the next steps?

Efforts at archiving, retrieving and citing digital research objects in standardized ways should be closely linked with open data and open-source software publication approaches, and should converge on common standards and practices. Citations to datasets and other digital research objects within publications should be treated on a par with the current treatment of bibliographic citations. Citations to these in the text should be made with a standard reference mark (in-text reference pointer) and the full reference should be given in the reference list of the publication, using a resolvable globally unique identifier (URL, DOI, HDL). Additionally, a formal semantic representation in OWL/RDF of the metadata describing these research objects, their provenance, their relationships to and citations of one another, etc., would be very useful and is now achievable. However, improved tools are required to reduce the labour of creating such metadata.

Openness and What it Implies

3.4 Derive new financially sustainable models of access

The emergence of the open access (OA) publishing model for the traditional scientific product, the journal article, has been a major driver in the emergence of Force11̇OA provides the gateway to new modes of scholarly communication, and is the cornerstone that must be promoted and extended if significant change to the scholarly publishing ecosystem is to take place. But OA per se is not enough. It must be shown to be sustainable through new business models, and must be weaved into the academic funding and reward system; neither will be easy. Here is what Force11 advocates to achieve the necessary change to this ecosystem through OA:

- Advocacy for OA through interactions with all the stakeholders mentioned in this document.

- Encouragement of conformance to OA licenses.

- Commitment to make all one's own scholarship as open as possible under the most liberal of those licenses.

- Education of others concerning the features and nuances of OA-based scholarship

- Development of new technologies that assume OA.

- Recognition that OA applies not just to research articles, but also to data, software, bibliographic and citation metadata, books and other components of the scientific process, and the whole scholarly enterprise.

- Recognition that OA applies just as appropriately to emergent research objects.

- Recognition that OA requires sustainable business models, and commitment to work towards achieving those new business models, that are likely to focus less on the content itself and more on the provision of revenue-generating services that facilitate discovery and reuse of that content in ways that advance scholarship.

What is happening now?

The following exemplify that change in the scholarly publishing world is already taking place and is likely to accelerate over time. It is the mandate of force to facilitate that acceleration:

- The increasing number of OA journals, including some that are regarded as comparable with the most highly regarded subscription access publications.

- The emergence of ORCID[4] as a system for creating unique personal identifiers, and hence for author disambiguation and better tagging of all aspects of scholarship.

- The creation of new tools that leverage content e.g. SciVerse[5] and Utopia[6] albeit neither yet in the open access/open source space.

- The development of new article-level metrics and other tools for assessing scholarship.

- The greater sense of awareness to be found within promotion committees concerning the value of alternative forms of scholarship.

What are the next steps?

Force11 members are stakeholders in all aspects of the scholarly enterprise and can influence it in different ways, but all start from the vision outlined above. Some specific steps we now need to take are:

- Start open enterprises that foster change: e.g., new data and software journals, institutional repositories that enable straightforward content exchange.

- Develop tools that highlight non-traditional forms of scholarly output such as database annotations created, blog posts written, and software developed.

- Develop means to assess and highlight the quality of OA content and other non-traditional forms of scholarly output.

3.5 Derive new business models for science publishers and libraries

Current business models for scholarly publication face significant disruption due to many factors: the growth in open access, the advent of alternative publication platforms that exploit new technologies for inexpensive communication and information exchange over the internet, a widening view of what constitutes a publishable research object (e.g. data, workflows), and the challenges of curating, linking and preserving the wider world of digital research objects. Furthermore, it is anticipated that the overall funds dedicated to scholarly communication may well become more restricted in future, at least on a per researcher basis. Both the major customers (research libraries) and brokers (currently, publishers) have an interest in being an active party in shaping the transition to new, sustainable business models, to ensure that the transition is a smooth one.

What is happening now?

The overall market for scholarly communications is on the order of $10 billion per year. The market is not a monolithic one, and disruptions are likely to be somewhat different in different disciplines. For example, there is an important distinction between those disciplines where publications are primarily in the form of books rather than journal articles. Also, researchers are growing accustomed to relying on an increasing number of free services. These pose both sustainability risks and opportunities. While freemium services typically manage to recruit only a few percent of users, some of these services can be sustained by a wider marketplace.

Some of these functions face significant challenges. For example, archiving and preservation of research objects, despite its high potential cost, is unique in not directly contributing to reward for producers. For this reason, it will likely be the most difficult to sustainably fund, and may require higher public investment.

What are the next steps?

To be financially viable, new communication modes will need to demonstrate tangible value to both producers and consumers. To be sustainable, the cost recovery streams will need to be aligned to perceived value. An additional factor that should be taken into account is that there are at least three different market sectors to which new products and services may be targeted: tools for producers (aka researchers),

---

[4]http://orcid.org/
[5]http://www.hub.sciverse.com/
[6]http://getutopia.com/documents/

enhanced products for consumers (researchers again), and reputation management (for individuals, institutions, and funding bodies).

In Dagstuhl, the Force11 group started to work on a more detailed business model, based on the Business Model Generation methodology [Osterwalder and Pigneur, 2010]. The results of this work will be made available on the Force11 web site `http://force11.org`

3.6 Derive new methods and metrics for evaluating quality and impact that exploit the technology

Scholarly practices and the way that science is undertaken is changing, as are the possibilities and associated activities of scholarly communication. Yet measures of assessment and impact have not caught up with these changes. Impact is a measure of change. Since these changes can be arbitrarily removed from the immediate outcome, one cannot always easily attribute the changes solely to the action performed. Measuring impact is complex because it depends on context, on purpose, on audience. It can have different effects for different individuals. Similarly, a communication can have different degrees and even polarities of effect. For example, a research paper might be simplified and published by newspapers to make headline news with great societal impact, but be roundly criticized or even ignored by academic colleagues.

What is happening now?

Presently, online versions of 'scholarly outputs' have tended to replicate print forms rather than exploit the affordances and functionalities of the digital terrain. The historical limits of print space are one reason, amongst others, that traditional journal articles tend to represent truncated versions of findings. The assumption is that technology will enable more effective enhanced papers. In addition, scholarly outputs will broaden to include, for example, software tools and social media channels. Work being undertaken under the Alt-metrics[7] umbrella pertains here and is to be supported. This has implications for policy. The challenge will then be how to get these metrics accepted by universities, funders and national decision makers.

What are the next steps?

It is accepted that metrics are still needed; however better mechanisms of measurement need to be put in place, that allow for different types of impact and influence.. A multi-dimensional measurement instrument would be useful. It needs to be customisable for specific situations and individual and it must be easy to use both for the individual academic and for the reviewer or decision-maker. What is being measured could include:

- Quality (exploiting new forms of measurement mechanisms)

- Influence (using new forms of alternative metrics)

- Social impact (measured, for example, through development goals)

- Economic impact

- Contribution to education (use in lectures, reading lists etc)

- Openness, making scholarly resources shareable, accessible, and re-usable

Mechanisms for measuring need to be reviewed in an age where traditional forms of peer review are also under critical scrutiny.

Although work has been undertaken to formalise these alternative notions of impact, none are directly applicable today. On the Force11 website, we make some concrete proposals for describing and utilising such new metrics.

---

[7]http://altmetrics.org/manifesto/

4 Related Efforts

The Force11 members have compiled, and will continue to update, a list of others ongoing efforts to improve digital scholarship.[8] You are invited to add to this living document, because we are sure that many other efforts exist, unknown to us. The catalog provides pointers to important papers, relevant blogs government and private sector reports, funding opportunities, policies, domain specific considerations, upcoming and past activities, and organizations.

Relevant papers and books are listed on the Force11 web site `http://force11.org`. These relate to various aspects of digital scholarship including, but not limited to the reward system, annotation, tools, repositories, text mining, citation of data, textual content in digital form other than research articles (e.g. of eBooks and technical reports), ontologies, metadata standards, semantics, provenance, features of research objects, and workflows.

5 Fulfilling this Vision

Force11 has identified the following actions that will contribute towards fulfilling the vision. Some actions apply to all stakeholders, others only to specific groups.

- Improved collaborative practice, which implies:
  - Increased social media presence
  - Maximizing informal contacts through conferences, workshops, meetings, calls, webcasts
  - Joint grant-funded activites leading to the creation of new tools and their description in publications
  - Other group technology development projects
- Coordinated standard and technology development, which implies:
  - Wholehearted adoption of W3C web standards and core ontologies
  - Open source development in response to user specifications from relevant stakeholders
  - Emphasis on reusability and extensibility
  - Creation of exemplars which act as drivers for future coordinated efforts, thereby insuring creativity and innovation is part of the development effort; such examples might be:
    * Novel tools that facilitate the use of digital objects
    * Development of novel metrics to measure non-traditional scholarship
    * Models for creating useful discipline specific digital repositories
    * New publishing paradigms
- Advocacy, which implies:
  - Promoting improved digital scholarship through traditional publication and non-traditional means
  - Participating in appropriate committees and other organizational bodies that can precipitate change
  - Fundraising for specific activities in support of change in digital scholarship

6 Acknowledgements

- Force11 web site (under ongoing development): `http://force11.org/`
- Full list of Force11 Participants: `https://sites.google.com/site/futureofresearchcommunications/home/attendees`
- Dagstuhl page on Force11: `http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=11331`

---

[8]See `https://sites.google.com/site/futureofresearchcommunications/links/links`.

References

[Altman and King, 2006] Altman, M. and King, G. (2006). A proposed standard for the scholarly citation of quantitative data. *DLib Magazine*, 13(3/4).

[Altman et al., 2008] Altman, R. B., Bergman, C. M., Blake, J., Blaschke, C., Cohen, A., Gannon, F., and Valencia, A. (2008). Text mining for biology—the way forward: opinions from leading scientists. *Genome Biology*, 9 Suppl 2(S7). doi: 10.1186/gb-2008-9-s2-s7.

[Attwood et al., 2009] Attwood, T. K., Kell, D. B., McDermott, P., Marsh, J., Pettifer, S. R., and Thorne, D. (2009). Calling international rescue: knowledge lost in literature and data landslide! *Biochemical Journal*, 424(3):317–333.

[Bechhofer et al., ress] Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., and Goble, C. (in press). Why linked data is not enough for scientists. *Future Generation Computer Systems*.

[Bechhofer et al., 2010] Bechhofer, S., De Roure, D., Gamble, M., Goble, C., and Buchan, I. (2010). Research objects: Towards exchange and reuse of digital knowledge. Paper presented at The Future of the Web for Collaborative Science (FWCS 2010), April 2010, Raleigh, NC, US. `http://eprints.ecs.soton.ac.uk/18555/`.

[Bourne, 2010] Bourne, P. E. (2010). What do i want from the publisher of the future? *PLoS Computational Biology*, 6(5). e1000787.

[Brase, 2009] Brase, J. (2009). Datacite: A global registration agency for research data. Paper presented at COINFO '09: The Fourth International Conference on the Cooperation and Promotion of Information Resources in Science and Technology.

[Chan et al., 2011] Chan, L., Kirsop, B., and Arunachalam, S. (2011). Towards open and equitable access to research and knowledge for development. *PLoS Med*, 8(3). e1001016.

[Ciccarese et al., 2011] Ciccarese, P., Ocana, M., Garcia-Castro, L. J., Das, S., and Clark, T. (2011). An open annotation ontology for science on web 3.0. *BMC Bioinformatics*, 2 Suppl 2(S4).

[Crosas, 2011] Crosas, M. (2011). The Dataverse Network: An open-source application for sharing, discovering and preserving data. *D-Lib Magazine*, 17:1–2. `http://www.dlib.org/dlib/january11/crosas/01crosas.html`.

[De Roure and Goble, 2009] De Roure, D. and Goble, C. (2009). Lessons from myexperiment: Research objects for data intensive research. Paper presented at the eScience Workshop 2009, October 15-17, 2009, Pittsburgh, US. `http://eprints.ecs.soton.ac.uk/17744/`.

[De Roure et al., 2009] De Roure, D., Goble, C., and Stevens, R. (2009). The design and realisation of the myexperiment virtual research environment for social sharing of workflows. *Future Generation Computer Systems*, 25:561–567. doi: 10.1016/j.future.2008.06.010.

[de Waard, 2009] de Waard, A. (2009). Hypotheses, evidence and relationships: The hyper approach for representing scientific knowledge claims. Paper presented at the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), co-located with the 8th International Semantic Web Conference (ISWC-2009), Washington DC, USA.

[de Waard, 2010] de Waard, A. (2010). From proteins to fairytales: Directions in semantic publishing. *IEEE Intelligent Systems*, 25(2):83–88. doi: 10.1109/MIS.2010.49.

[Engestrom, 1999] Engestrom, Y. (1999). Communication, discourse and activity. *The Communication Review*, 3(1):165–185.

[Goble and De Roure, 2007] Goble, C. A. and De Roure, D. C. (2007). myexperiment: social networking for workflow-using e-scientists. Paper presented at the Proceedings of the 2nd workshop on Workflows in support of large-scale science, Monterey, California, USA.

[Greenberg, 2009a] Greenberg, J. (2009a). Theoretical considerations of lifecycle modeling: An analysis of the dryad repository demonstrating automatic metadata propagation, inheritance, and value system adoption. *Cataloging and Classification Quarterly*, 47(3–4):380–402. doi: 10.1080/01639370902737547.

[Greenberg, 2009b] Greenberg, S. A. (2009b). How citation distortions create unfounded authority: analysis of a citation network. *BMJ*, 339(b2680). doi: 10.1136/bmj.b2680.

[Hargreaves, 2011] Hargreaves, I. (2011). Digital opportunity: A review of intellectual property and growth. Retrieved from http://www.ipo.gov.uk/ipreview-finalreport.pdf.

[King, 2007] King, G. (2007). An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods Research*, 36.

[Kniffel and Bailey, 2009] Kniffel, L. and Bailey, C. W. (2009). Cuts, freezes widespread in academic libraries. Technical report, American Libraries. http://www.ala.org/ala/alonline/currentnews/newsarchive/2009/may2009/academiclibrarywoes051309.cfm.

[Murray-Rust, 2007] Murray-Rust, P. (2007). Data-driven science: a scientist's view. Paper presented at the NSF/JISC Repositories Workshop, Phoenix AZ, April 10, 2007. http://www.sis.pitt.edu/~repwkshop/papers/murray.html.

[Neylon, 2011a] Neylon, C. (2011a). Open research computation: An ordinary journal with extraordinary aims. Science in the Open, http://cameronneylon.net/blog/open-research-computation-an-ordinary-journal-with-extraordinary-aims/.

[Neylon, 2011b] Neylon, C. (2011b). Time for total scientific openness. *New Scientist*, 2828.

[Osterwalder and Pigneur, 2010] Osterwalder, A. and Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley and Sons.

[Rosenthal and Reich, 2010] Rosenthal, D. S. H. and Reich, V. (2010). Archiving supplemental materials. *Information Standards Quarterly*, 22(3). doi: 10.3789/isqv22n3.2010.04.

[Sanderson and de Sompel, 2011] Sanderson, R. and de Sompel, H. V. (2011). Open annotation: Beta data model guide. Retrieved 9 September 2011, 2011, from http://www.openannotation.org/spec/beta/.

[Schultz, 2011] Schultz, D. M. (2011). The proliferation of scientific literature. Eloquent Science: http://eloquentscience.com/2011/06/the-proliferation-of-scientific-literature/.

[Shotton, 2009] Shotton, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, 22(2):85–94. doi: 10.1087/2009202.

[Shotton, 2011] Shotton, D. (2011). The five stars of online journal articles: an article evaluation framework. *Nature Preceedings*, 17.

[Smith et al., 2009] Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., and Rambaut, A. (2009). Origins and evolutionary genomics of the 2009 swine-origin h1n1 influenza a epidemic. *Nature*, 459(7250):1122–1125. doi: 10.1038/nature08182.

[Star and Griesemer, 1989] Star, S. L. and Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907–39. *Social Studies of Science*, 19(3):387–420.

[The Economist, 2009] The Economist (2009). A world of hits. *The Economist*. http://www.economist.com/node/14959982.

[Vickery, 1999] Vickery, B. (1999). A century of scientific and technical information. *Journal of Documentation*, 55:476–527.

[Wenger, 2000] Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, 7(2):225–246. doi: 10.1177/135050840072002.

[White et al., 2008] White, H., Carrier, S., Thompson, A., Greenberg, J., and Scherle, R. (2008). The Dryad data repository: A Singapore framework metadata architecture in a DSpace environment. In Greenberg, J. and Klas, W., editors, *Proceedings of the International Conference on Dublin Core and Metadata Applications,* pages 157–162.