Sharing knowledge: EC-funded projects on scientific information in the digital age

Conclusions of a strategic workshop - Brussels, 14-15 February 2011

EUROPEAN COMMISSION

Directorate-General for Research and Innovation
Directorate B – European Research Area
Unit B.6 – Ethics and gender

Contact: Francesco Fusaro

Office SDME 03/17
B-1049 Brussels

Tel. (32-2) 29-87458
Fax (32-2) 29-84694
E-mail: francesco.fusaro@ec.europa.eu
RTD-OPEN-ACCESS@ec.europa.eu

EUROPEAN COMMISSION

SHARING KNOWLEDGE: EC-FUNDED PROJECTS ON SCIENTIFIC INFORMATION IN THE DIGITAL AGE

Conclusions of a strategic workshop

Brussels, 14-15 February 2011

REPORT
by
Caroline Sutton
(Rapporteur)

Prepared by:
Caroline Sutton,
Co-Action Publishing
caroline.sutton@co-action.net

EUROPE DIRECT is a service to help you find answers to your questions about the European Union

Freephone number(*):

00 800 6 7 8 9 10 11

(*)Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2011

ISBN 978-92-79-20451-7
doi:10.2777/63780

Contents

EXECUTIVE SUMMARY

Since 2000, the European Commission has paid increased attention to access, dissemination, and preservation of scientific information. While the Treaty of Lisbon (2000) provides a legal basis for the Commission's work in this area, practical policy has developed through communications and council conclusions related to scientific communications, the European Research Area (ERA), the digital agenda and the innovation union.

The report presents and analyses discussions that took place at a workshop held and organized by the European Commission, with the title "Workshop on EC-funded projects on scientific information in the digital age". The workshop was primarily organized by the *European Commission Directorate-General for Research and Innovation; Science, Economy and Society,* with the participation of the Directorate-General for Information Society and Media. Leaders from eighteen projects supported under different European Commission programmes, including ICT Policy Support[1] (formerly eContentPlus[2]), Science in Society[3], and Infrastructures (under the FP7 Capacities programme) [4] attended the workshop. The overarching purpose of the workshop was to discern learning from the projects and to reflect upon how this learning can be translated into policy recommendations and concrete actions. The workshop also provided an opportunity for networking and information exchange.

This event follows the workshop held on 25 and 26 November 2010 by the European Commission with national experts on open access and preservation in the ERA. A Communication and Recommendation on scientific information in the digital age is planned in this area for late 2011, and it is the intention of the organizers that outputs of the workshop will contribute to identifying points to be addressed in the forthcoming policy documents.

RESULTING POLICY RECOMMENDATIONS

The workshop was organized as a series of highly participatory exercises, using the methods of *Participatory Leadership* (see e.g. www.artofhosting.org). In general, exercises moved participants through a process that began by connecting individual participants to the aims of the workshop, and then to harvesting knowledge produced collectively from the different projects, to then using this knowledge to identify key questions and issues and finally for formulating policy recommendations. The resulting policy recommendations are list here.

[1] http://ec.europa.eu/information_society/activities/ict_psp/index_en.htm

[2] http://ec.europa.eu/information_society/activities/econtentplus/closedcalls/econtentplus/index_en.htm

[3] http://cordis.europa.eu/fp7/sis/home_en.html

[4] http://cordis.europa.eu/fp7/capacities/research-infrastructures_en.html

On Scientific Data:

1) Publicly funded research data should be made publicly available as a general rule. Exceptions should be possible (e.g. in relation to privacy considerations in the area of medical research). Embargos for privileged usage may apply (and should be discipline-dependent and less than 3 years).
2) The EU should fund preparatory phases to allow the scientific community to assemble and identify common grounds (i.e. with respect to data policies, interoperability standards, data models, data formats, exchange protocols, types of protocols, etc.).
3) Each project proposing to generate a significant amount of data should include a data management plan (DMP) and devote a fraction of its budget to its execution. The DMP should define the logistics of data management and dissemination.
4) Projects should deliver their scientific data to certified/accredited data repositories. To comply with this, repositories would need to be certified/accredited and the group therefore further recommends that a limited number of certification schemes should be defined (different levels adapted to different situations).

On Infrastructures:

1) Scientific information is a patrimony/heritage that we cannot afford to lose or to close. What is needed to protect against loss is:
   - Lowering of the barriers to sharing (costs, technology, legal, cultural, linguistic, disciplinary)
   - Provide incentives to share
2) EC to instruct Member States to require funded projects to address the issue of lowering the barriers.
3) EC and Member States to coordinate capacity-building investments and their sustainability.
4) EC and Member States to promote physical and virtual organizational mobility.

On Preservation:

1) For sustainable/trustworthy preservation:
   - Set up a digital preservation bank to evaluate and fund start-up preservation "companies".
   - Set up evaluation systems (e.g. certification) for 1) repositories and (2) techniques (tools, infrastructures), e.g. the EC in FP8 requires data to be deposited in certified repositories, e.g. require publicly funded repositories to be certified.
2) EU Directive/ "Scientific Davos"/ PSI Directive
   - Address a directive to remove obstacles to preservation (e.g. by the copyright laws, account for privacy issues) on a European level, and inaugurate a "scientific Davos" to guide and monitor the implementation.
   - Evaluate and (potentially) widen the PSI (Public Sector Information) directive to include preservation.
3) Incentives for researchers to deposit
   - Set up a citability system for data
   - Make data a first-class publication

- Ensure that the academic merit system takes data publication into account
- Set up career structures for data scientist
- Training for data scientists as "data re-users"
- Provide support to make publication of data easier (or possible in the first place) and select what data deserves to be preserved and for how long (e.g. what is patrimonial data).

4) Preservation layer
   - Ensure the funding of a "preservation layer" on top of the developing e-Infrastructure.
   - Promote the social cross-support between repositories (e.g. setting up SLAs between repositories).
   - Work to extend to a global infrastructure (e.g. persistent identifiers)

On Publications:
IF the EC wishes to make publications which stem from the Innovation Union freely available, then there should be:

1) Bipartisan critical assessment of the existing system
2) An investigation of infrastructural barriers and future researcher needs and innovations to deliver them.
3) Definitions and enabling of the infrastructure to deliver them.
4) Funding available to enable the transition.
5) Long term funding structures which support the new structures within reasonable budgetary constraints.
6) The issue of commercial use should also be considered in this area.

On Awareness:

1) Promote and explore new ways of measurement using the new and innovative tools in the digital age.
2) Improve measurement systems by including discipline-specific criteria.
3) Make knowledge-sharing and open access an evaluation criterion in project proposals (with an opt-out for certain types of data).
4) Create qualified collections of publications harvested from repositories
5) Universities should have non exclusive licenses to research results for their institutional use.
6) Provide and make public EU project results in certain formats in a centralized public platform.

ANALYSIS OF THE DISCUSSIONS AND RECOMMENDATIONS

The workshop on EC-funded projects on scientific information in the digital age brought together a diverse group of individuals working on diverse projects who might not otherwise have had an opportunity to share and discuss their projects and the future of scientific information together. The discourse that emerged during the two days gave an opportunity to identify points of convergence, the relevant importance of different issues

for different areas of activity, and to contrast different scholarly communities in relation to transformations taking place in the context of digital technology.

An important theme that emerged is that scientific information is entering a phase of consolidation, in which a greater degree of coordination and efficiency can be detected. Six of the eighteen projects referred to defragmentation or used terms that similarly referred to efforts to gain a meta perspective upon developments within a sector or field (e.g. preservation, scientific data deposition, repositories, etc.) and the need to work horizontally and coordinate efforts. From such a viewpoint, it was pointed out that one can then identify gaps and opportunities for better coordination, efficiency, and overall benefits. Specifically, the project descriptions contained such language as: "Defragment scientific information" (APARSEN), "Coordinating resources" (CLARIN), "Coordination of Grid" (EGI), "Recommendations, coordination" (E-IRGSP2-3), "Interoperability of different data sources" (GRDI2020), "Provide picture of FP7 research outcomes" (OPENAIRE). A few projects specified defragmentation as one of the main aims or primary impacts. Defragmentation was also named in several of the key issues that were identified for the different areas as well as in some of the policy recommendations.

The process of defragmentation is progressing at an uneven pace across areas of activity within scientific information, between scholarly communities and across countries. Some scientific communities are able to explicitly envision and state future needs and drive change in the systems surrounding scientific information to meet these needs. In contrast, other research communities are being transformed by technological advancements, and must be convinced that solutions indeed meet needs they might not be aware they had. An important question in this context is how we can learn from those communities who are driving change. This suggests that the European Commission can continue to play an important role in supporting developments and coordinating efforts both across member states, but also across stakeholder groups and areas of activity.

Finally, key issues surfaced in the discourses of the workshop. These were Technological, Sociological/ Behavioural, Financial Sustainability, Legal and Overall Vision. Like the other themes, these issues are unevenly distributed in terms of their relevant importance to the areas of activity. However, all of these issues do surface in some way in all areas of activity. Financial sustainability, in particular, was of great importance to all projects and areas of activity. Similarly, the Sociological/ Behavioural aspects were important for all groups who referred to inciting researchers to embrace and integrate the new tools and services being developed through these projects into their own personal workflows and everyday research activities. Associated with behavioural aspects were references to the need to re-evaluate the rewards and assessment systems that are currently in place across European institutions and beyond. It was pointed out in several work groups and in several project descriptions that this system is a critical barrier to success. The European Commission can play a continued role in working with stakeholders to address these issues through dialogue, policy and other means.

When juxtaposed against the fields of scientific data, preservation and infrastructures, it is striking that a vision is still lacking for one of the key outputs of research activity: publications.

The policy direction and vision for the ERA, and with it the Digital Agenda and Innovation Union, suggest that it will be important for the European Commission to formulate a vision for publications/scientific communications in the forthcoming communication on scientific information in the digital age or to provide a framework for creating this common vision. A vision for research outputs must be coordinated with other policy areas related to the realization of the ERA and broader economic development goals of the European Union. This type of coordination can only be facilitated at a higher (European) level, with input from necessary stakeholder groups.

BACKGROUND

Since 2000, the European Commission has paid increased attention to access, dissemination, and preservation of scientific information. Science and innovation have been identified as key components of economic growth in the European Union. In turn, the free flow of information and research outputs between researchers and between researchers, innovators and society at large, within a single European Research market, has come to be regarded in policy as an important component of achieving this growth. Moreover, as the new millennium has unfolded, there has been recognition that new technologies offer new possibilities and new challenges as science transitions as a result of the digital revolution, and that these issues need to be addressed to assure a globally competitive research community.

The legal foundations for the European Commission's involvement in access, dissemination and preservation in the Treaty of Lisbon are:

> **Article 179**, states "The Union shall have the objective of strengthening its scientific and technological bases by achieving a European Research Area in which researchers, scientific knowledge and technology circulate freely [...]. "
>
> **Article 180c** – "[...] the Union shall carry out the following activities: [...] - dissemination and optimisation of the results of activities in Union research, technological development and demonstration".
>
> **Article 183** "For the implementation of the multiannual framework programme the Union shall: - lay down the rules governing the dissemination of research results".

In 2000, the Lisbon Agenda which defined a goal of at least 3% GDP to be invested in research and development by EU Member States by 2010 in order to become "the most dynamic and competitive knowledge-based economy in the world".[5] In the same year the European Research Area was created to achieve an "internal market" for research, to restructure European research by coordinating national research activities and policies, and to develop a European Research Policy.[6] At the heart of the ERA has been an ambition

[5] European Commission (2004) Facing the Challenge. The Lisbon strategy for growth and employment. Report from the High Level Group chaired by Wim Kok, November.

[6] COM (2000) 6 - 18.01.2000

to create a Europe-wide space or "single market" for research and innovation, also known as the "fifth freedom". Coupled with this ambition is the notion that there should be a free movement of knowledge, implying access to, dissemination of and exploitation of publicly-funded research. The most recent reports from the European Research Area Board in 2009 and more recently in 2010 have called for a "New Renaissance" and set clear principles on the management of intellectual property resulting from publicly funded research, as well as policies on access to, and dissemination of publications and research data resulting from publicly funded research. These principles concern not only the flow of information between scientists but also between scientists and other communities.

The creation of the ERA has involved the development of scientific infrastructures to support research and innovation. Among other areas, efforts have been directed to developing libraries through the Digital Libraries Initiative, launched by the European Commission in 2006. Attached to the initiative was a high level group (HLG), which submitted a final report in 2009 and three sub-groups on: Intellectual Property Rights, Public-Private Partnerships and Scientific Information. The latter sub-group contributed to the launch of the PEER project, which will be further presented below. The most tangible result of the digital libraries initiative, as noted by the HLG, is the creation of Europeana, a digital library positioned at the European level. The final report from the Digital Libraries HLG, noted that "New paradigms of scientific information" constitutes one of the challenges for the future.

Scientific information has also been directly addressed as an area of European Commission activity. In 2004, then DG-Research commissioned a study to assess the evolution of the market for scientific publishing and to "[…]discuss the potential desirability of European level measures to govern access to and the exchange, dissemination and archiving of scientific publications." A final report was presented in January 2006 and a public consultation on the recommendations revealed that access issues were the most controversial.[7] Proponents of open access regarded the benefits to be related to improving the process and impact of science, and emphasized the role of science as a public good to achieve economic and social impacts. Subscription publishers and publishers' associations reacted to the report by emphasizing that greater access was already being achieved through current policies, and some expressly opposed open access. In 2007 the European Commission presented its communication on Scientific information in the digital age: access, dissemination and preservation[8] and thereafter the European Council put forth its conclusions. The Open Access Pilot in the Seventh Framework Programme (FP7) was launched in response and is "intended to provide researchers and other interested members of the public with improved online access to EU-funded research results."[9] The pilot requires researchers to deposit articles resulting from FP7 projects in seven defined areas into an institutional or subject based repository; and to make best efforts to ensure open access to these articles within six months or twelve months, depending

7 Summary of the Responses to the Public Consultation on the "Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe" Commissioned by the Research Directorate-General.

8 COM (2007) 56 final.

9 Ibid.

upon the research discipline. This pilot mandate places pressure upon the overall system to provide access to scientific information, and at the same time ensure preservation and appropriate infrastructures (e.g. OpenAIRE) and support. The OA pilot also involves providing reimbursement for open access publication charges.

European policy links access, dissemination and preservation with innovation and is at the centre of the Europe 2020 Strategy, which argues that "...future economic growth and jobs will increasingly have to come from innovation in products, services and business models."[10] The strategy places science at the forefront of job creation and economic development and highlights the potential impact that investments in research can lead to. A 2010 Communication from the Commission on the Innovation Union[11] specifically states that "In 2012, the Commission will propose a European Research Area Framework and supporting measures [...]. They will notably seek to ensure through a common approach [...] dissemination, transfer and use of research results, including through open access to publications and data from publicly funded research". The communication further states that the Commission "[...]will promote open access to the results of publicly funded research. It will aim to make open access to publications the general principle for projects funded by the EU research Framework Programmes [...]".

One of seven flagship initiatives under the Europe 2020 Strategy is the adoption of a Digital Agenda for Europe[12], also with a number of consequences for the scholarly information arena. In this area the Commission's Communication on A Digital Agenda for Europe includes an aim to drive ICT innovation by exploiting the single market, which entails "[...publicly funded research should be widely disseminated through Open Access publication of scientific data and papers" and "[...] the Commission will appropriately extend current Open Access publication requirements [...]".

[10] Ibid.

[11] COM (2010) 546.

[12] COM(2010) 245 final du 19.5.2010

REPORT ON THE WORKSHOP ON EC-FUNDED PROJECTS ON SCIENTIFIC INFORMATION IN THE DIGITAL AGE

This document reports on discussions that took place at a workshop held and organized by the European Commission, with the title "Workshop on EC-funded projects on scientific information in the digital age". The workshop was primarily organized by the *European Commission Directorate-General for Research and Innovation; Science, Economy and Society,* with the participation of the Directorate-General for Information Society and Media.

The workshop brought together leaders from eighteen projects supported under different European Commission programmes, including ICT Policy Support[13] (formerly eContentPlus[14]), Science in Society[15], and Infrastructures (under the FP7 Capacities programme)[16]. The projects related to access, dissemination and preservation as well as to infrastructures.

PURPOSE

The purpose of the workshop, according to the invitation, was to

- *cluster and create synergies and networking opportunities among EC-funded projects in related areas,*
- *take stock of project findings and discuss future strategies on access, dissemination and preservation,*
- *foster a constructive debate and exchange information among stakeholders, and*
- *create an opportunity and a space for common reflection on how the results of the projects can be translated into policy recommendations and concrete actions.*

This workshop follows the workshop held on 25 and 26 November 2010 by the European Commission with national experts on open access and preservation in the European Research Area (ERA). Both events aimed to encourage networking among different stakeholders and to provide information and inspiration for the Commission's future work towards developing policy on access and preservation of scientific information in the digital age. A Communication and Recommendation on scientific information in the digital age is planned in this area for 2011, and it is the intention of the organizers that outputs of the workshop will contribute to identifying points to be addressed in the forthcoming policy documents.

FORMAT

The workshop was organized as a series of highly participatory exercises. This format is known as *Participatory Leadership* (see e.g. www.artofhosting.org). In general, exercises moved participants through a process that began by connecting individual participants to

[13] http://ec.europa.eu/information_society/activities/ict_psp/index_en.htm

[14] http://ec.europa.eu/information_society/activities/econtentplus/closedcalls/econtentplus/index_en.htm

[15] http://cordis.europa.eu/fp7/sis/home_en.html

[16] http://cordis.europa.eu/fp7/capacities/research-infrastructures_en.html

the aims of the workshop, and then to harvesting knowledge produced collectively from the different projects, to then using this knowledge to identify key questions and issues and finally for formulating policy recommendations.

PARTICIPANTS

Participants for the workshop were drawn from 18 projects that are receiving support from the European Commission under the Programmes noted above, or have recently completed a project supported by the Commission under one of these.

In total, 44 individuals participated in the workshop, including: 10 representatives of the European Commission, 33 Project Leaders and 1 Rapporteur. See also the list of participants contained in Annex 1.

The participant group was selected in line with the workshop aims (i.e. to encourage networking across related projects and to provide a space for reflection on how the results can be translated into policy). Participants were members of various stakeholder groups that are generally recognized within the scholarly communications arena, but they did not necessarily represent these in an official capacity. Further, because participants were drawn from project groups, some stakeholder groups were not present, e.g. funding bodies other than the EU and scholarly societies.

The participants came from both public sector and private sector organizations, with the largest number coming from the former. While it was observed during the workshop that few active researchers were present, in fact, seven of the 33 project participants held positions at university research departments or positions that required close engagement with researchers at their institutions (e.g. as Assistant University Director or as Research Director).

THE PROJECTS, THEIR IMPACTS AND BENEFITS

Before moving into the actual workshop, the first day began with some introductory remarks by Gilles Laroche, Head of the Gender and Ethics Unit in DG Research and Innovation, and Octavio Quintana Trias, Director in charge of the European Research Area in DG Research and Innovation. Matthieu Kleinschmager, in-house consultant in the EC Learning & Development Unit thereafter introduced participants to the format of the workshop and the flow of activities and exercises for the two days.

The first session began with each project being presented briefly by its project leader. For the next session, participants divided themselves into groups of four, with two projects represented in each resulting group. The aim of this session was to dive deeply into two projects during a two-hour period. Thus, participants gained deep insight into two projects focusing the discussion on the following topics: 1) Description of the project, 2) What problems did the project address?, 3) What are/were the project impacts/benefits/results? Through the discussion, the group participants both listened intently, asked questions and

then reflected back to the project leaders what they felt were key insights. Notes were recorded for each project discussion.

For easy reference, a table summarizing the different projects, key contacts and links to the project homepages, is presented in Appendix 2. This section provides short complimentary descriptions of the projects as they were presented at the workshop, including the impacts and benefits that were identified and discussed.

Describing the projects

APARSEN - *Metadata for preservation, curation and interoperability*
www.aparsen.eu

Over the last decade digital preservation efforts have intensified and the field of preservation has been characterized by many projects with a wide array of aims. There is a continued need for metadata for preservation, curation and interoperability. APARSEN recognizes the necessity of gaining an overview of the whole and of developing an integrated vision for data preservation. Rather than conducting additional research, APARSEN aims to create a common vision for research into digital preservation across Europe that encompasses all types of digital objects. In doing so, it brings together different stakeholders such as universities, research institutions, national and international coalitions, vendors, national libraries and archives, bid science and industry. This vision should be based on evidence that is produced through the project and through testing of preservation techniques. The project is closely associated with the Alliance for Permanent Access (APA) and will continue at the end of the project period as a Virtual Centre of Excellence built upon the APA.

The main impact of APARSEN is to *defragment* scientific information while providing digital preservation. The benefits of this are a spreading of excellence through *training courses, certification and sustainability paths.*

CESSDA - *Council of European Social Science Data Archives*
www.cessda.org

CESSDA is a long standing entity, having existed since 1976 as an umbrella organisation for social science data archives across Europe. The Council's work is predicated on the belief that "sharing of research data is of great value to the research community at large as well as to the individual researcher". The European Commission is currently funding a process to create a European Research Infrastructure Consortium (ERIC) as a legal entity. The resulting CESSDA-ERIC is intended to enhance the existing CESSDA through collective activity, developing and maintaining procedures, practices, protocols and tools, beyond the scope or resources of any single organization or country.

The primary impact derived from CESSDA is the introduction of *structure* and *standards* in conjunction with deposition of social science data. Benefits include supporting higher quality research and the introduction of new tools and solutions.

CLARIN - *Common language resources and technology infrastructure*
www.clarin.eu/external/

CLARIN is working to create a federation of existing archives that contain data with a language component. Further, CLARIN seeks to make these repositories available for all Social Science and Humanities scholars in Europe. Beyond making documents available, CLARIN is also developing advanced tools to allow researchers to conduct more specialized searches on content and to manipulate documents in new ways. CLARIN has cooperated to some extent with DARIAH (see below), but is focused on a specific area of research. The CLARIN project is currently at a preparatory phase which will end in June 2011. In preparation for the next phase, the project members noted the need to identify a funding model. In describing their project, representatives explained the need for a federation in response to the existence of language data "all over Europe" and to tackle the challenge of language databases that are very different from one another.

The main impact of the project is the European Coordination of language resources and the benefit is the added value that this brings to national collections.

COMMUNIA - *The European Thematic Network on the Digital Public Domain*
www.communia-project.eu

COMMUNIA is a thematic network that is working to become a European point of reference for theoretical analysis and strategic policy discussion of existing and emerging issues concerning the public domain in the digital environment. COMMUNIA's work also extends to consideration of such topics as alternative licensing for creative material, open access to scientific publications and research results and orphan works. In addition, COMMUNIA is involved in community building through funding travel and organizing workshops, conferences, etc. Through this community, a team of partners is emerging that are committed to the topic of public domain knowledge and has made it possible to produce a public domain manifesto. A range of stakeholders are involved that include librarians, consumer organizations, scientific data organizations, NGOs (IPR-related), and activists. Members of the project possess expertise in law, economics, the social sciences and computer science. This group is politically active, having produced two policy recommendations to date with aspirations to provide others in future.

The main impact of COMMUNIA is *building awareness* and *sustainability*. The primary benefit is the public domain manifesto that was launched and ultimately access to more objects in the public domain.

DARIAH - *Digital Research Infrastructure for the Arts & Humanities*
www.dariah.eu

As the name of the project suggests, DARIAH aims to create infrastructures for research in the Arts & Humanities. More specifically, DARIAH aims to: build a durable data infrastructure to preserve, reuse and analyse research data; establish the necessary organizational structures to evolve and expand infrastructures to the needs of the communities working within the Arts & Humanities; and create a living environment in which data, services, methods and research results are shared and discussed, within

individual l research projects as well as across communities. Unlike other infrastructures that are based on heavy technology, DARIAH uses technology to provide procedures, standards, and capacity building in addition to technological tools to support these. The project leaders report that an in-depth understanding of what individual groups are doing and from this an understanding of their needs has been critical. This is because familiarity with digital technology varies greatly across the social sciences and humanities. Project work has led project members to observe that open access is important but one size does not fit all communities. What constitutes an incentive varies by community.

Similar to CESSDA (see above) the primary benefit of DARIAH is higher quality research in the social sciences and humanities, while the main impact is that DARIAH provides a *structure* for conducting research and *standards*. During the discussions it was also noted that DARIAH contributes to *defragment* this arena by addressing the diversity of the humanities.

EGI - *European Desktop Grid Initiative;* *www.egi.eu*

The European Grid Initiative has a ten year history of support by the European Commission. The initial project sprang forth out of the computing needs of high energy physics researchers at CERN who required more powerful computing technology to manage analysis and simulation. The EGI is founded on the National Grid Initiatives (NGI), which operate the grid structures in each country. Today the EGI is available not only to high energy physicists, but to all researchers across Europe. Most recently, the project focused on coordinating efforts and identifying processes and mechanisms for establishing EGI, to define the structure of a corresponding body, and ultimately to initiate the construction of the EGI organization. At the time of the workshop the EGI had become established as a foundation, with 20 staff members, and as such had moved into a new phase of existence as a permanent entity. This new phase presents new challenges, not least of which is to identify a steady revenue stream to support continued work. The EGI represents a mature example of horizontal activity and integration of efforts across national borders and for the needs of researchers across disciplines.

The main impact of the project is the coordination of the grid, while the benefits include a more *economically efficient system*, and a best use of available resources.

E-IRGSP2 & 3 - *e-Infrastructures Reflection Group* *www.e-irg.eu*

The E-IRGSP2 & 3 supports a high level group of representatives from the different Member State ministries and e-infrastructure specialists who have been nominated by national ministries. The goal of the high level group is to produce recommendations for the EC, Member States and major EU initiatives. The E-IRGSP2 & 3 supports this group through producing policy support in the form of producing white papers and recommendations, as well as through dissemination of information (websites, mailing list, newsletters), management, and support to the board, delegates, etc. The e-infrastructure, as addressed by the group includes supercomputing, grid and cloud computing, networking and data infrastructure. The project representative reflected that a *common, European*

e-infrastructure can speed up developments in this area and that the E-IRGSP helps bring national e-infrastructure problems and issues to a European level. Some general observations that were put forth by the project representative are that the project's work makes clear that there are both technological as well as human/sociological issues to be addressed.

The main impact derived from the group is the generation of recommendations, coordination and best practices. Benefits include resulting policy documents and greater awareness.

EUROCANCERCOMS - *Establishing an Efficient Network for Cancer Communication in Europe*
www.eurocancercoms.eu

EUROCANCERCOMS is addressed to a specific area of research and a specific illness – namely Cancer. The aim of the two-year project is to provide proof-of-concept for integrating cancer information using cutting edge technology to provide a unique platform for providing cancer intelligence to the public, patients and cancer healthcare professionals. This project represents one of a just a few projects at the workshop that was concerned with delivering research information to the broader community. Also notable is that the project looks to bring together the vast array of cancer information available yet "fractured" across the virtual world in a more integrated way. Information is collated for a wide range of sources and packaged in meaningful ways for different audiences who require such knowledge.

The main impact of the project is improved dissemination and communication while the benefit is that different cancer communities are able to be more engaged with cancer intelligence.

EUROVO-AIDA - *Euro-VO Astronomical Infrastructure for Data Access*
www.euro-vo.org/pub/

EUROVO-AIDA is an Integrated Infrastructure Initiative funded under the framework of the FP7 e-infrastructure Scientific Research Repositories initiative. The project has been running since February 2008 and is aimed at implementing the Virtual Observatory concept on behalf of Europe. The Virtual Observatory is based on the notion that the corpus of the world's astronomical data should be available from each individual astronomer's desk, regardless of where that individual is sitting.

The primary impact of the project is to implement the VO such that scientists are united with data, service providers, technical issues, standards and education. Important benefits of the project are the lessons learned, as well as the training and education on how to use the resulting product.

GRDI2020 - *Towards a 10-year Vision for Global Research Data Infrastructures*
www.grdi2020.eu

One of the unique features of this project, in comparison with the other projects represented at the workshop, was that it is run by a private enterprise. GRDI2020 is a high level project; it exists to create a Roadmap and a 10 year vision for Data infrastructures in Europe. The

work of the group also builds on a global Scientific Data Infrastructures Community to devise common use cases, to share experiences and to plan and innovate. The primary issues addressed are open access, interoperability, and privacy issues. The project representatives noted at the workshop that there is an emphasis in the project on working bottom-up, from the end user point of view. This has revealed a wide array of challenges across different disciplines and across technical issues. New data models are necessary in order to respond to the needs of different disciplines. Moreover, it is clear that different levels and forms of education and training are necessary in order to take advantage of infrastructures that are built. The roadmap will include a definition of the infrastructure, an identification of the challenges to be overcome (technological, application, organization), a specification of main functionalities in order to support data intensive research and multi-disciplinary research.[17]

The main impact of the project is to build the next generation scientific data infrastructure. GRDI2020 benefits the community through interoperability of different data sources, and by providing recommendations to address the organizational and technological challenges.

LIQUIDPUB - *Liquid Publications: Scientific Publications meet the Web*
http://liquidpub.org/

This project rethinks how knowledge and intelligence is shared and evaluated as well as investigates and develops methods for capturing this knowledge. Scientific knowledge distribution is still largely based on the traditional notion of the "paper" publication and on peer review as a quality assessment method. LIQUIDPUB proposes a paradigm shift in the way scientific knowledge is created, disseminated, evaluated and maintained. Among other things, LIQUIDPUB recognizes less mature or less formal stages of knowledge formation, such as the thoughts and ideas that arise amongst those attending a conference, or the collaborative process of writing a "liquid book", and is working to develop tools for capturing these forms of knowledge. The project group includes both ICT developers as well as philosophers and a publisher. Among the new questions the project gives rise to is the need to examine the assumption that usage is an indicator of quality, as well as what rewards users should receive.

The main impact of the project is the sharing and dissemination of knowledge, connecting different tokens of knowledge. In addition is a better understanding of the status quo (e.g. in relation to peer review and methods of measuring the relevance of knowledge), and the development of new metrics for measuring quality and social dissemination of knowledge. The benefit of the project is the ability to share what has not been shared earlier, as well as a more efficient sharing through new tools.

OAPEN - *Open Access Publishing in European Network*
http://project.oapen.org/

OAPEN has worked towards developing and implementing a sustainable Open Access (OA) publishing model for academic books in the Humanities and Social Sciences (HSS).

[17] At the time of completing this report the roadmap was completed. GRDI2020- A Coordination Action: Towards a 10-year vision for global research data infrastructures. GRDI2020 Preliminary Roadmap Report. Global Research Data Infrastructures: The Big Data Challenges.

To achieve this, OAPEN considered the changing economic conditions of academic books publishing, business models, rewards for authors and how to measure impact. Out of the project a common electronic platform for the publication of open access books by European university presses has emerged. Reflecting on further questions that should be asked within the project, the representative noted that a further cost analysis of the technical level and of the services should be considered.

One of the major impacts of the project was to make open access publishing better known among books publishers, such that OA publishing encompasses not only journal articles but also monographs in the form of published books. An important benefit of OAPEN is that certain domains of publications have been made more visible and that OAPEN has now formed a foundation for OA book publishing.

ODE - *Opportunities for Data Exchange*
www.ode-project.eu

ODE takes its point of departure in the data deluge that is emerging. The potential to unlock the answers to the so-called "grand challenges" of our times can only be unlocked, argues the ODE website, through an interoperable data-sharing, re-use and preservation layer coupled with today's system of e-infrastructures. ODE does not seek to build this layer, but rather to identify what needs to be built and how it can best be built to meet the needs of researchers and society. The project has only recently launched and entered phase one of a three-phase process. The first phase involves collecting stories of successes and failures with respect to the preservation of valuable or particularly large data sets. These stories are being harvested from both the social sciences and hard sciences, and include new data and old data. In a second phase, the human element will be investigated by way of confronting individuals who conscientiously deposit data and those who are apathetic to understand who shares and why. This should provide valuable insight for political interventions and technical developments. Finally, the project involves groups such as STM, LIBER, national libraries to investigate how data deposition can be integrated into the scholarly publishing workflow. Like APARSEN (see above), ODE is associated with the Alliance for Permanent Access (APA).

The impact of ODE will derive from the rich basis of information the project will collate. The benefits include the analysis of this information and recommendations to funding agencies and the broader community regarding data deposition.

OPENAIRE - *Open Access Infrastructure Research for Europe*
www.openaire.eu

OPENAIRE was formed to support the implementation of Open Access in Europe. Specifically it provides the means to promote and realize the widespread adoption of the Open Access Policy, as set out by the ERC Scientific Council Guidelines for Open Access and the Open Access pilot launched by the European Commission under FP7. OpenAIRE is involved in providing useful tools to researchers to enable them to comply with the deposition requirements of the FP7 pilot programme. National help desks have been established across Europe, a web page with useful information and instructions has been created, and an orphan repository has been constructed to allow for deposition of articles

where an institutional or subject repository is lacking. In addition to this practical help, Open AIRE is also contributing to improving data management services by working with 5 disciplinary communities to further investigate this.

The main impact of OpenAIRE is the provision of standardized data that is connected to research, namely OA to FP7 publications. The benefits derived so far are a European OA orphan repository, new metrics on OA use, organized access to FP7 publications, and with time, the ability to provide a picture of FP7 research outcomes.

PEER - *Publishing and the Ecology of European Research*
www.peerproject.eu

As stated in the 2010 Annual Report for Year 2 of the project, PEER "...is investigating the potential effects of the large-scale, systematic depositing of authors' final peer-reviewed manuscripts (so called Green Open Access or stage-two research output) on reader access, author visibility, and journal viability, as well as on the broader ecology of European research." PEER brings together stakeholders from the publishing industry as well as from the library/repository sector, funders and researchers. Creation of an observatory through which the project team can study behavioural, usage and economic impacts has been established and is being populated with content. Through its work, PEER has also contributed to addressing technical issues such as file formats, metadata, etc. As well as other questions such as how to ensure that articles that are deposited with the PEER depot are sent to the appropriate repository.

The main benefit from PEER is evidence based results upon which future dialogues on article deposition can be based. An unintended benefit of PEER is the PEER depot, which is hosted by INRIA. Participants suggested that there is an interest in maintaining the depot after the project period if funding can be attained.

SISOB - *An Observatorium for Science in Society based in Social Models*
http://sisob.lcc.uma.es/

SISOB addresses the need for funding agencies and others to evaluate the impact of research, while also considering the possibilities that are now possible within the digital environment. Traditionally peer review prior to publication and bibliometric measures after publication have been used as tools for this purpose. The goal of SISOB is to develop tools to measure and predict the social appropriation of research knowledge, modelled as the product of complex interactions within and between multiple, intersecting communities of scientists, journalists, industrial, decision makers and consumers.

The eventual impact of SISOB could be a shift in the way in which the impact of projects and research are assessed based on the benefit derived from the project, namely, the ability to measure the social appropriation of knowledge.

SOAP - *Study of Open Access Publishing by Key Stakeholders*
http://project-soap.eu/

The SOAP project recently held a final workshop to present the results of the project and was scheduled to complete at the end of February. SOAP set out to describe and analyze the open access publishing landscape as well as explore the risks and opportunities of the transition to open access. Specifically, the SOAP project ran a large scale survey of attitudes of researchers on, and their experiences with, open access publishing. Approximately 40 000 responses were collected across disciplines and countries worldwide. The results demonstrated support for the idea of open access and highlighted funding (39%) and (perceived) quality issues (30%) as the primary barriers to publishing in open access journals. One of the conclusions of the project is that open access will grow automatically if funding issues are resolved and when a larger number of open access journals that are recognized for their quality are also available. The results raise other questions such as, what needs to change within the overall publishing system if we are to achieve more widespread or full open access in publishing? And, will open access lead to cheaper solutions in publishing?

The project impact is derived from the large body of information that was generated through the large scale study, while the benefit is that the project addressed both drivers and barriers to open access to attain a more balanced understanding of the landscape.

Distilling commonalities from diverse projects

Following the deep discussions of the 18 projects, participants regrouped and were divided into five groups in order to distil commonalities from across the diverse projects present. The five groups focused on one of five subjects:

1) Project descriptions
2) What problems did the projects address?
3) What are/were the project impacts/benefits/results?
4) What other questions come to us from listening to the projects?
5) From circulating between all groups, what strikes us?

The fifth group was referred to as a "floater group" as those in it wandered between the four other groups, eavesdropping to determine whether any common themes could be further distilled across the discussions taking place.

A number of key phrases emerged in the **project descriptions** and were identified by the project description groups. These can be grouped under higher order headings that include:

- Need to share, disseminate information (journals, books, data...) across Europe
- Who pays for what? Sustainability? Business Models
- Assessment of scientists
- User needs
- Data sharing
- Language

- Policy-makers
- Public
- Identity/Curation
- Registries/Portals
- Data Ownership
- Awareness of implications of licensing/copyright

The resulting **impacts/results/benefits of the projects** reflected this list of themes, and are listed below each project descriptions in the section above. This group did not distil the commonalities, but these are analyzed in the current report in a later section.

Among the **other questions** that emerged were several more overarching questions regarding principles as well as a number of very specific questions. Among the overarching questions were:

- Questions about sustainability issues for Research Infrastructures (RIs)
- Turning projects into RI: What? Why? How?
- What constitutes success?
- At what level does public (investment) in RI make sense?
- Could better coordination from a single EC department?
- How is policy turned into projects and how does evidence lead to policy?

The **floater group** overheard at least five topics/themes that were being discussed across the other four groups in some form. These were:

- Sustainability – referring largely to the need for projects to transition from a project phase to a more long term sustainability entity, as well as to longer term financing issues.
- Interests exist at different levels – interests at different levels (local, national, European, International) can be in harmony or compete.
- Quality, success and assessment – these discussions referred to the need to ensure that the products and continuing contributions of projects are measured and contribute to overall quality within scholarly activity and communications. This topic also related to how to measure the success of a project.
- Researcher/user needs – all groups discussed and emphasized researcher needs as a driving force in the work of different projects and questions to be considered.
- Lifetime of a project – the different phases of development in a project were discussed (also related to the first bullet point in this list)

KEY ISSUES AND POLICY RECOMMENDATIONS

The second day of the workshop involved a four-stage process. The morning began by each individual stating a personal key question. As these key questions were read aloud in a circle, the breadth of ideas and issues that were identified, as well as the themes across these, provided some inspiration and warmed up participants to begin thinking in terms of critical questions. Thereafter, participants chose to participate in one of the five work groups centred on key areas within scholarly information in the digital age. After some

discussion regarding the umbrella names given to the key areas, five work groups were defined as: Awareness, Infrastructures, Preservation, Scientific Data, and Dissemination & Publications. Initially the title of each group contained the word "open", but this was removed to allow for a broader consideration. There was discussion regarding where *dissemination* should be placed. Some participants argued that one cannot disengage dissemination from publications and hence these topics should be handled together. However, as becomes apparent below, the publications group did not make direct recommendations regarding dissemination, while other groups did (e.g. SCIENTIFIC DATA). For this reason, this report refers to a PUBLICATIONS work group.

The five work groups first engaged in two sessions centred on key questions. In the first session, a long list of key questions was composed, which were thereafter honed in the second session and pared down to three or four key questions. The latter part of the day centred on two sessions aimed at transforming the key questions into policy recommendations. Each group first arrived at 3-4 policy recommendations based on the key questions. Once these were formulated, all groups engaged in "ritual dissent", whereby a representative of the group presented the recommendations and reasons for them to three additional groups and then received direct critique of the recommendations, without discussion. This critique was brought back to the original group and the recommendations were honed in relation to it to create a final list of recommendations to present in a final plenum session.

Scientific Data

Three overarching themes were identified as the basis for strategic questions facing scholarly communications in the area of SCIENTIFIC DATA. These were:

1) Carrots & Sticks
   This issue relates to how each discipline/community of practice can identify appropriate standards, rules, attitudes, and 'licensing', while also ensuring that these are compatible at a global level. Further, this group posed the question, "Will new technology (social networks) change/influence this (in 10 years)?"

2) Funders: Business Models (rewards)
   Here the scientific data group pointed out that it is key to ask who shall/should pay whom such that data is published, linked, preserved, etc. in a sustainable and trustworthy manner.

3) Technology
   How do we ensure global and interdisciplinary access and reuse of data despite the plethora of infrastructures, data models, data languages, etc.? And related to this, how shall the entire data life cycle by supported? It is necessary to ask this latter question in order to ensure both usefulness and affordability.

Based on these key questions, the SCIENTIFIC DATA group put forth the following policy recommendations:

5) Publicly funded research data should be made publicly available as a general rule. Exceptions should be possible (e.g. in relation to privacy considerations in the area

of medical research). Embargos for privileged usage may apply (and should be discipline dependent and less than 3 years).
6) The EU should fund preparatory phases to allow the scientific community to assemble and identify common grounds (i.e. with respect to data policies, interoperability standards, data models, data formats, exchange protocols, types of protocols, etc.).
7) Each project proposing to generate a significant amount of data should include a data management plan (DMP) and devote a fraction of its budget to its execution. The DMP should define the logistics of data management and dissemination.
8) Projects should deliver their scientific data to certified/accredited data repositories. To comply with this, repositories would need to be certified/accredited and the group therefore further recommends that a limited number of certification schemes should be defined (different levels adapted to different situations).

Infrastructures

As emerges in the points below, the key issues identified and policy recommendations made by the INFRASTRUCTURES group overlap to some extent with those of the SCIENTIFIC DATA group. This is logical given that infrastructures are a necessary component in enabling the use of scientific data in a digital environment.

The key questions facing INFRASTRUCTURES were identified as:

1) What are the fundamental principles the infrastructure should help realize?

2) How can one trigger and support a sustainable ecosystem of infrastructures for knowledge sharing?

3) How can one provide and enable infrastructures to support a new paradigm of knowledge dissemination?

Arguing for the importance of these questions, the group noted that they are important to ask if one holds the opinion that “Science is to be shared” and that “Sharing science is part of building the European Commission”.

The policy recommendations put forth by the INFRASTRUCTURES group were:

5) Scientific information is a patrimony/heritage that we cannot afford to lose or to close. What is needed to protect against loss is:
   - Lowering of the barriers to sharing (costs, technology, legal, cultural, linguistic, disciplinary)
   - Provide incentives to share
6) EC to instruct Member States to require funded projects to address the issue of lowering the barriers.
7) EC and Member States to coordinate capacity-building investments and their sustainability.
8) EC and Member States to promote physical and virtual organizational mobility.

Preservation

The PRESERVATION work group identified the following key questions:

1) How can business models for sustainable/trustworthy preservation (both institutional and commercial) be stimulated?
2) How can legal frameworks across Europe be unified? For example, how can one facilitate preservation (e.g. copyright ownership)? And, how can one ensure that publicly funded digital data is preserved and publicly accessible to the greatest extent possible?
3) How can we institute an infrastructure to enable sharing efforts for preservation? Related to this, what incentives can be used to encourage researchers to deposit 'their' data for preservation: make the benefits of top down requirements and bottom-up benefits more transparent?

The PRESERVATION work group arrived at recommendations to answer each of these strategic questions specifically and one additional recommendation. The recommendations were:

5) For sustainable/trustworthy preservation:
   - Set up a digital preservation bank to evaluate and fund start-up preservation "companies".
   - Set up evaluation systems (e.g. certification) for 1) repositories and (2) techniques (tools, infrastructures), e.g. the EC in FP8 requires data to be deposited in certified repositories, e.g. require publicly funded repositories to be certified.
6) EU Directive/ "Scientific Davos"/ PSI Directive
   - Address a directive to remove obstacles to preservation (e.g. by the copyright laws, account for privacy issues) on a European level, and inaugurate a "scientific Davos" to guide and monitor the implementation.
   - Evaluate and (potentially) widen the PSI (Public Sector Information) directive to include preservation.
7) Incentives for researchers to deposit
   - Set up a citability system for data
   - Make data a first-class publication
   - Ensure that the academic merit system takes data publication into account
   - Set up career structures for data scientist
   - Training for data scientists as "data re-users"
   - Provide support to make publication of data easier (or possible in the first place) and select what data deserves to be preserved and for how long (e.g. what is patrimonial data).
8) Preservation layer
   - Ensure the funding of a "preservation layer" on top of the developing e-Infrastructure.
   - Promote the social cross-support between repositories (e.g. setting up SLAs between repositories).
   - Work to extend to a global infrastructure (e.g. persistent identifiers)

Publications

The PUBLICATIONS work group attracted the largest number of participants. Members of the group reported that there were strong and varying opinions and that the critical questions identified and recommendations made were the result of discussions to identify a position that all members could accept.

The resulting key questions from the PUBLICATIONS group were put forth as follows: Given the objectives of the Innovation Union and the Digital Agenda, and against the background of global competitiveness and the need for sustainability:

1) What has to be accessible, to whom and how?
2) What transformations need to happen to the current infrastructure and across all stakeholders for this to happen? (EC, Member States, publishers, libraries, research funders, researchers)?
3) How best can this be financially enabled in the transformation phase and with long term sustainability, without ignoring the importance of the legal framework and quality assurance?

Policy recommendations that were agreed upon by the PUBLICATIONS work group were:

IF the EC wishes to make publications which stem from the Innovation Union freely available, then there should be:

7) Bipartisan critical assessment of the existing system
8) An investigation of infrastructural barriers and future researcher needs and innovations to deliver them.
9) Definitions and enabling of the infrastructure to deliver them.
10) Funding available to enable the transition.
11) Long term funding structures which support the new structures within reasonable budgetary constraints.
12) The issue of commercial use should also be considered in this area.

Awareness

The main questions posed in the AWARENESS work group were:

1) How can/should quality and success of scientific information be measured and who should do it?
2) How can we make researchers and society-at-large benefit from sharing scientific information?
3) What are the other barriers to sharing of scientific information and how can they be overcome?

Policy recommendations from the AWARENESS work group were:

7) Promote and explore new ways of measurement using the new and innovative tools in the digital age.

8) Improve measurement systems by including discipline-specific criteria.
9) Make knowledge-sharing and open access an evaluation criterion in project proposals (with an opt-out for certain types of data).
10) Create qualified collections of publications harvested from repositories
11) Universities should have non exclusive licenses to research results for their institutional use.
12) Provide and make public EU project results in certain formats in a centralized public platform.

In relation to the final point, the group pointed out that while information on European Union projects is available on CORDIS this can be difficult to navigate. More use of multi-media opportunities would also be helpful.

THE STATE OF SCIENTIFIC INFORMATION IN THE DIGITAL AGE – AN ANALYSIS OF THE DISCUSSIONS & RECOMMENDATIONS

What does the discourse of the workshop reveal about the current state of transformation of scientific information in the digital age and for access, dissemination and preservation specifically? In order to make sense of and interpret the discussions that took place during the workshop and the resulting policy recommendations, it is useful to consider some of the main themes that emerged and how the different areas of activity are positioned in relation to these.[18] Below I present three exercises for organizing the outputs of the discussions that took place during the workshop as the relate to the current state of development in the area of scientific information in the digital age, specifically considering this in relation to four of the areas of activity around which Day 2 discussions were organized: publications, scientific data, preservation and infrastructures.

From fragmentation to consolidation

The project and policy discussions provide evidence that scholarly communications in the digital age is entering a new phase in its evolution. Evidence can be found in one of the main themes that emerged out of the workshop discussions, namely *defragmentation*. Six of the eighteen projects referred to defragmentation or used terms that similarly referred to efforts to gain a meta perspective upon developments within a sector or field (e.g. preservation, scientific data deposition, repositories, etc.) and the need to work horizontally and coordinate efforts. From such a viewpoint, it was pointed out that one can then identify gaps and opportunities for better coordination, efficiency, and overall benefits. Specifically, the project descriptions contained such language as: "Defragment scientific information" (APARSEN), "Coordinating resources" (CLARIN), "Coordination of

[18] These exercises are based on a consideration of the documentation from the workshop (personal statements that were put forth during the initial and final sessions on each day of the workshop, the notes from the project descriptions provided on Day 1, the key issues that were identified and the policy recommendations). In addition, I spoke with one participant from each of the 18 projects by telephone, asking them to recall briefly how that project had been presented on Day 1. This was necessary in order to confirm that I had understood correctly the notes that had been recorded by the participants themselves.

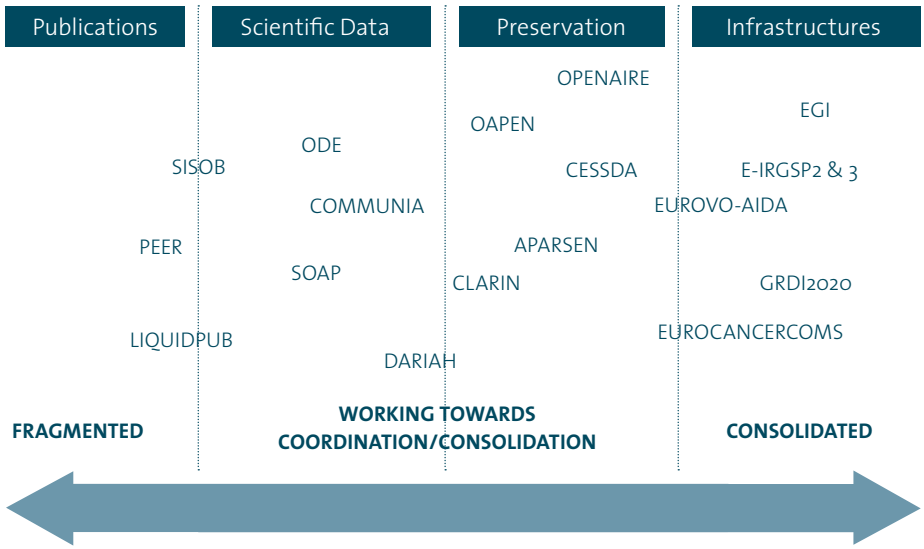
Grid" (EGI), "Recommendations, coordination" (E-IRGSP2-3), "Interoperability of different data sources" (GRDI2020), "Provide picture of FP7 research outcomes" (OPENAIRE). A few projects specified defragmentation as one of the main aims or primary impacts.

References to the need to consolidate/coordinate were also made in the context of the key questions that were discussed and policy recommendations. An example of the latter is found in recommendation (4) from the SCIENTIFIC DATA work group, which states "EU to fund preparatory phases where scientific community gets together to find common ground (data policies, interoperability standards, data models, data formats, exchange protocols, types of protocols, etc.). Notably, all three recommendations from the INFRASTRUCTURES work group include an element of consolidation.

Despite the clarity with which some participants referred to defragmentation and consolidation as necessary, some reflected on the disparity that characterized the communities with which they worked or upon the lack of coordination within the area of activity they were involved with within scientific information. Given the prominence of this theme, a useful exercise is to plot the four areas of activity as well as the projects along a continuum from fragmented to consolidating.

Looking across the four main areas of activity – publications, scientific data, preservation, and infrastructures – different degrees of fragmentation contra consolidation can be detected. In the diagram below, the different areas of scientific information in the digital age have been plotted along a continuum from fragmented to consolidating. Similarly, the project descriptions have been used to plot the projects along the same continuum. The placement below is not meant to indicate an exact placement of these areas and projects, but rather to illustrate their general relationship to the two end points of the continuum and to one another.

FIGURE 1: Areas of activity and projects in relation to degree of consolidation



Defragmentation and convergence was most evident in the projects related to infrastructures, particularly those funded under the 7th Framework Programme, for which one of the three areas of activity is defined to be "Integrating existing RIs".[19] GRDI 2020, for example, is working to create a road map on the scientific data infrastructure, which shall be used specifically by policy makers. E-IRGSP2 supports the e-IRG which is an inter-governmental policy organization and therefore by nature oriented towards consolidation and defragmentation. Likewise, the EGI is based on the National Grid Initiative (NGI) and operates to link these. EUROVO-AIDA is practically oriented towards achieving a consolidation that has earlier been defined by the Euro VO in conjunction with the global concept of the Virtual Observatory. While not an infrastructure initiative, EUROCANCERCOMS represents consolidation at a subject level as it is working to collate intelligence across professional groups in the area of cancer.

Generally speaking, most of the projects contain an element of defragmenting and consolidating efforts even if the projects as such are currently placed to the left of consolidation. Common to OpenAIRE, APARSEN, CESSDA, CLARIN, OAPEN, DARIAH is that these projects are working towards consolidation within a discipline or more broadly, but are working with users that may not recognize the benefit of a common structure or solution. Similarly, ODE, COMMUNIA and SOAP recognize the need for a common vision. However, they are not directly engaged in consolidating an area, but are focused on generating knowledge upon which coordination and consolidation could be built. PEER is also directed towards generating knowledge within a somewhat contested area of activity. SISOB and LIQUIDPUB represent challenges to the current system as they seek to introduce new forms of assessment and new units of scientific outputs, respectively. As such, these projects embrace a new paradigm but have yet to be adopted broadly within mainstream scientific activity.

One additional observation that can be made in relation to consolidation status is that those projects and areas of activity that are closer to consolidation share an overall conception of knowledge as something that is to be shared. At the core of activity and projects in the areas of INFRASTRUCTURE and PRESERVATION is a shared understanding of knowledge and information as constituting a network; one could even use the metaphor of "knowledge as an infrastructure".[20] The INFRASTRUCTURES work group went so far as to explain the strategic questions they identified by stating "Science is to be shared – Sharing science is part of building the European Union." One of the three strategic questions that were identified by the group included an important assumption, "How to provide and enable infrastructures supporting the *new paradigm* of knowledge dissemination?"(emphasis my own). Sharing information is also understood as reusing information, as is reflected in a statement from a member of the SCIENTIFIC DATA work group who commented in a plenary session, "[t]he key is to have data and information in a form which allows *reusing* it: formats, metadata, semantics, etc. disciplinary and interdisciplinary....", while another stated "How can we use unfamiliar data (time/discipline) in an automated way?"

[19] See e.g. http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=existing_infra, accessed on March 10, 2011.

[20] John Wilbanks "New Metaphors in Scientific Communication: Libraries and the Commons", presentation at IATUL 2007 Conference, Stockholm, Sweden.

Digital preservation and digital infrastructures were born out of the digital world. As such, their history is short and behaviours are less entrenched, perhaps making it easier to agree upon basic principles that are in line with the opportunities afforded by current technology. In contrast, scientific data and publications have been the key outputs of research for centuries, and as such are imbued with long standing tradition and habits that must be transformed.

While sharing of data and scholarly thinking characterized early science, with time research outputs became the target of more proprietary behaviour. John Wilbanks has argued that the underlying metaphor behind scientific information and knowledge is that of property[21], though there are some notable exceptions (e.g. high energy physics, astronomy). The results of the SOAP project suggest that a larger group of scientists are taking part in a knowledge network/infrastructure by publishing their results in open access journals. However, the statement preceding the policy recommendations from the PUBLICATIONS work group begins with "IF the EC wishes to make publications which stem from the Innovation Union freely available...." This IF, reflects a lack of common vision within the publications arena regarding the future of research publications in the digital era, a point we will return to.

One common point of convergence within the area of scientific publications is the current widespread acceptance of the article or, in the case of Humanities and some Social Science disciplines, the book, as a unit of output. In contrast, the area of scientific data is challenged by not only the volume of data but the very wide variation in what constitutes data, how it is structured, etc.

The great emphasis placed upon defragmentation during the workshop discussions reflects the European Commission's stated role to engage in capacity building and act as a supporting body. The European Commission has expressly sought to fund infrastructures in the areas of access, dissemination and preservation, as well as support networking activities. Clearly these aims have been necessary for the progress that has already taken place. However, the recommendations from the different groups and the overall importance that this theme had within the workshop also indicate a continued need for the European Commission to play – and perhaps expand – this role. The European Commission is uniquely positioned to coordinate activities and efforts across Member States, across stakeholder groups, across disciplines and across the areas of access, dissemination and preservation. Moreover, given the aims of the ERA, the European Commission is also best positioned to coordinate efforts in these areas with other European policy goals as well as coordinate European efforts with global efforts.

As a final observation, it is worth noting that the structure of the workshop contributed to illustrating the new phase described here. As noted above, the workshop brought together diverse projects that were funded through different programmes within the European Commission. The workshop also gave opportunities for identifying convergences across these projects and work groups and allowed individuals to network outside of their usual area of activity.

[21] Ibid.

Research communities

A second theme that emerged in the discussions, and which is also reflected in the key questions and policy recommendations is the recognition of different research communities with different needs, coupled with a concern to design and create solutions and systems that are based on user/researcher needs. The key question identified by the SCIENTIFIC DATA group summarizes this diversity across disciplines, asking: "How does each discipline/community of practice find their standards, rules, attitudes, 'licensing' compatible at the global level?" The discussions regarding the project descriptions in particular, brought to light that this is not always a straight forward question to answer as both explicit and latent user needs exist.

To understand the nature of user needs, it is useful to consider various communities' relationships to transformations taking place in the digital age. Roughly speaking, research communities can be placed in one of two categories: those that are driving change in the digital age, and those whose research is being driven by changes brought about by the digital age. This theme also relates to that of defragmentation, because as I will illustrate below, those research communities who are driving change are also characterized by a higher degree of consolidation. See Figure 2 for an illustration.

In line with the European Commission's aim to create the free movement of knowledge ("fifth freedom") within the ERA, communities of researchers are emerging that are centred upon research interests rather than upon regional boundaries. In some cases the community and its interest in cooperating and collaborating predates the infrastructure or tools that are currently being developed to support the flow of knowledge (e.g. as in the case of EUROVO-AIDA), and predates the European Commission's interest in promoting the free movement of knowledge. In other cases, the creation of the infrastructure or set of tools is contributing to community building.

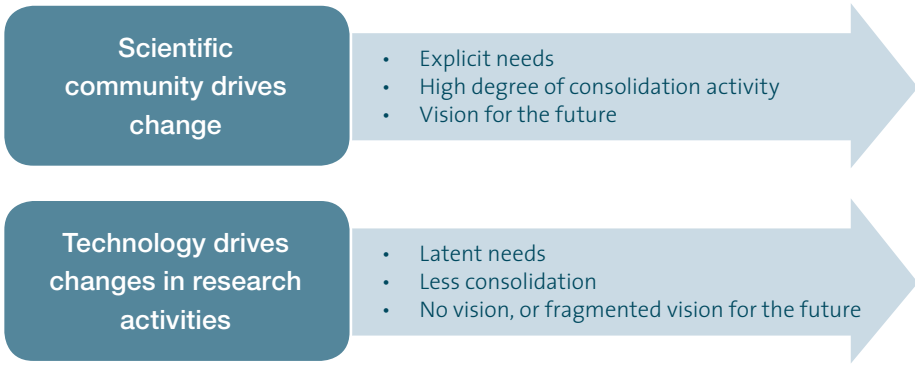
It is beyond the scope of this report to review all disciplines, but some examples drawn from the discussions at the workshop can illustrate these points. Among the communities who drive technological advancements to support science in the digital age are astronomers and physicists.

Recalling the description of the European Grid Initiative (EGI), this project was born out of the specific needs of higher energy physicists at CERN who required more powerful computing technology to manage analysis and simulation. Physicists were also pioneers in the area of preservation and repository work, having established Arxiv in 1991. Physicists were also early adapters of open access, using Arxiv as a means of distributing research results quickly in a field that is generating knowledge more quickly than publishers can keep pace with. In Europe, the large number of FP7 projects that CERN is involved in provides further evidence that this community is driving change.

Like high energy physicists, astronomers require expensive equipment to carry out their research work and high energy computing to analyse data. Moreover, the astronomers' data, by nature, are spread around the globe. These conditions have led to consortia and cooperation in different capacities to deliver needed technology to support the community. The Virtual Observatory concept, which the EUROVO-AIDA project is working

to move into an operational phase, is pushing forward advances in data deposition and archiving technology. Beyond the research community, EUROVO-AIDA is also working to pioneer delivery of scientific data to younger students and to amateur astronomers, demonstrating a further commitment to sharing research data.

FIGURE 2: Research communities and their relationship to transformations in the digital age



Among those communities who are driving change, there may be debates over which standards to adopt and the best path to achieve goals, but there is a great deal of convergence around recognized needs. Moreover, and perhaps of critical importance, these communities tend to have a clear vision for the future of their field to direct the coordination of efforts. In contrast, within those groups whose work is changing due to new technology, debate may be centred on needs as such, as it might not be clear what technology can offer. Moreover, many fields tend to be much more diverse than high energy physics and astronomy in terms of methods and subject matter, making consolidation and harmonization more challenging. It is against this backdrop that DARIAH refers to its mission as that of creating infrastructures – plural – rather than as creating a single infrastructure. In the Arts and Humanities, the creation of infrastructures also involves the creation of procedures, standards, capacities and conceptualizing of possibilities, in addition to technological architectures. Projects like DARIAH begin by immersing themselves in the everyday research situation of relevant scholars to identify latent needs that can be met through technology.[22] In this sense, the resulting tools will contribute to transform the conditions of research and scholarly investigation among these communities.

The workshop reported on here provided an opportunity for those working with these different types of communities to share experiences. While physicists and astronomers may represent unique communities, it may be worthwhile to better understand how and why they are able to formulate common visions for the future, which in turn allow one to move towards convergence rather than fragmentation.

[22] Drivers of change can also be found within the Arts & Humanities. For example, architectural sciences have driven developments in 3-D modeling.

Critical issues in the digital era

A third theme that emerged in the workshop is the presence of a series of recurrent issues that face different activities in order to achieve change and transformation in scholarly communications in the digital era. Here the different themes are delineated according to the relevant weight these were given by the different groups on the final day of the workshop.

TABLE 1: Key issues and their significance across areas of activity

Key Issues	Publications	Scientific Data	Preservation	Infrastructures
Technological	LOW	HIGH	HIGH	MEDIUM
Sociological/ Behavioural	HIGH	HIGH	MEDIUM-HIGH	LOW
Financial Sustainability*	HIGH	MEDIUM	HIGH	MEDIUM
Legal	LOW	LOW	HIGH	LOW
Overall Vision	HIGH	HIGH	HIGH	HIGH

Technological issues

Electronic publishing emerged during the late 1990s and is today commonplace with very few exceptions. Publishers have tackled most of the necessary technological issues to enable electronic documents to be read, accessed and used online. Some challenges have been in the area of depositing researchers' work in repositories. While this is generally regarded as the responsibility of the individual researcher, publishers do need to provide the proper meta data, proper file formats, etc. The PEER project notably is making further contributions in this area. Archiving and preservation are also relevant to the work of publishers and the PARSE project, not a participant in this workshop, found that over 90% of the journals that were considered in that study had well-established strategies and practices in place for journal articles.[23] Perhaps reflecting the fairly unproblematic state of technology from the publisher perspective, the publications group did not discuss technical issues nor identify technical issues when formulating key questions or policy recommendations.

As we move towards scientific data a different picture emerges. Given the inter-related activities between scientific data, preservation and infrastructures, the technological challenges related to data preservation were discussed across these work groups. The SCIENTIFIC DATA work group noted the "plethora of infrastructures, data models and data languages" as well as the need to support the entire data lifecycle. In its policy recommendations, the INFRASTRUCTURES group pointed out the need to direct efforts to lower barriers to sharing, noting technology as one issue. During the project presentations such technical issues as interoperability and the challenge of ensuring that all types of

[23] http://www.parse-insight.eu/ . See alos Smit, E., Van der Hoeven, J and D. Giaretta (2011) Avoiding a DigitalDark Age for data: why publishers should care about digital preservation. Doi: 10.1087/20110107.

objects are preserved and are migrated to new formats as these emerge in the future, were mentioned. It was also pointed out in the context of a few project discussions that data management is to some extent more difficult than the creation of infrastructures, for example, in part due to the heterogeneity of data as well as to the large number of stakeholders that are involved. Standards are also mentioned across work groups and projects as critical to define in order to direct technology and solutions.

Sociological/behavioural challenges

All work groups and project discussions noted the presence of not only technical challenges but also what were referred to as sociological, human factors and behavioural issues. Inciting researchers to embrace and integrate the new tools and services being developed through these projects into their own personal workflows and everyday research activities was mentioned by three of the work groups and was naturally addressed explicitly in the AWARENESS work group. This challenge was also remarked upon in several project presentations.

The PRESERVATION work group listed six recommendations on incentives for researchers to deposit data. Recalling these recommendations, they are:

- Set up a citability system for data
- Make data first class publication
- Ensure that the academic merit system takes data publication into account
- Set up career structures for data scientists
- Training for data scientists as “data users”
- Provide support to make publication of data easier (or possible in the first place) and select what data deserves to be preserved and for how long (e.g. what is patrimonial data).

These recommendations can loosely be categorized as serving one of two purposes. They either contribute to building recognition around the collection of data as a valued scientific endeavour, or to assisting researchers to make the action of deposition easier to manage and maximize.

The underlying theme of recognizing scientific contributions can also be detected among the recurring reference to the need to re-evaluate the rewards and assessment systems that are currently in place across European institutions and beyond. It was pointed out in several work groups and in several project descriptions that this system is a critical barrier to success. At present, researchers and the success of research projects are overwhelmingly assessed on the basis of publications (in most fields in the form of a journal article), and in turn using the very blunt instrument of the impact factor to analyse the significance of those publications. While there was consensus that quality is important, many participants emphasized that new tools and measurements, emanating from the digital opportunities that exist, should be considered and introduced as appropriate. The SISOB project is an example of an attempt to revise this system and introduce new instruments.

Particularly where projects and efforts contribute to meeting latent needs of users or to leveraging latent opportunities within the overall system of scholarly communications, the added value will need to be communicated and demonstrated to the broader community in meaningful ways. This was noted, for instance, by the DARIAH project group as being critical to success in relation to infrastructures for the Arts and Humanities.

It is worth noting in this context that most of the projects included a training element in their overall delivery. For example, EUROVO-AIDA brings together European astronomers to provide training on how to use the advanced functionalities of the Virtual Observatory. EUROVO-AIDA is also concerned with the dissemination of research data beyond the research community. In an effort to bring science to the classroom, the program has created teaching materials using original data but in a format that can be used and understood by teenage students. NECOBELAC is a project whose raison d'être is precisely to provide training by training the trainer. The project might offer a template for other projects looking to disseminate knowledge for users.

Financial Issues & Sustainability

Although a few participants reflected that it might be (more) useful to arrange smaller discussions among more similar projects, a benefit of the mixed group of participants was that one gained insight into the lifecycle of research outputs. The EGI project was already completed at the time of the workshop and a foundation had been established to carry the European Grid work forward with a permanent staff. Similarly, the OAPEN project was in a process of becoming registered as a foundation, with the project now completed. In contrast, other projects, like ODE, had only recently been launched.

Some projects, such as PEER and SOAP, as well as ODE, by their nature as research projects will have an endpoint at which it is unlikely that the project will continue in a new format. However, the majority of projects were centred on developing a concrete solution, tool(s), or other product, which if successful, will or should transition to implementation on a more permanent basis.

In this context, reference to *sustainability* was made across groups, and is reflected in the key questions that are posed by the work groups and even emerges in the policy recommendations. In a general sense, sustainability referred to the fact that the advancements made by projects could be lost if it was not possible to discover a new format for the continued existence of the work that had been initiated. Sustainability referred both to the need to identify new business models and possible new forms for partnership, but also financial issues.

All work groups referred to financial sustainability. The CLARIN project participants noted, for example, that projects tend to be funded by the European Union but upon implementation, attention shifts to Member States for financial contributions. This seems to have left many projects with an uncertain financial stream. To name a few additional examples, out of the PUBLICATIONS work group the following recommendation was made: *Long term funding structures which support the new structures within reasonable budgetary constraints.* The PRESERVATION work group proposed *Ensure the funding of a "preservation layer" on top of the developing e-Infrastructure.* While the SCIENTIFIC DATA

work group identified as a key question/issue: *Funders: Business Models (rewards), Who shall/should pay whom such that data is published, linked, preserved, etc. in a sustainable and trustworthy manner.*

Legal Issues

While not the most discussed issue, legal concerns did surface during the workshop. For example, legal issues in relation to copyright and commercial use were raised in the PUBLICATIONS group as well as in the SCIENTIFIC DATA group. For the PUBLICATIONS group, legal issues look to be important enough to be the focus of one of the major issues addressed in key questions and in policy recommendations. Some groups, such as the project discussion on EGI, noted the improvements on legal issues that have been made, for example with the introduction of ERIC. Legal issues received slightly more attention in the PRESERVATION work group, who included as a strategic question, "How to unify legal frameworks across Europe?" and "How to facilitate preservation (e.g. copyright ownership)?"

Overall vision

Recalling the discussion above regarding consolidation and defragmentation, the presence of an overall vision looks to be critical to achieving coordination and consolidation. This overall vision can be most clearly detected within the areas of SCIENTIFIC DATA and INFRASTRUCTURES. Participants in the SCIENTIFIC DATA group referred to the "Riding the wave", a final report from the High Level Expert Group on Scientific Data.[24] This report presents a series of scenarios to illustrate the vision for the future that is presented later in the report for an infrastructure to support the storing and sharing of scientific data. Moreover, this vision links data with needed infrastructures. GRDI2020 is specifically working towards a common vision by preparing a roadmap for data infrastructures. The existence of a road map or common vision provides a baseline against which proposed policies and activities can be considered.

Those working in the area of PRESERVATION look to share a common vision or understanding of the future as evidenced by the large number of cross-national cooperative efforts that are being made, the OpenAIRE project being just one of a large number. An overall vision may be difficult to define at this time. Although there are some important challenges to be addressed in the area of data preservation, the existence of a common vision for scientific data should make it easier to move forward on these challenges. However, preservation also concerns the preservation of publications, and it is this area of activity that looks to be furthest from formulating a common vision. The divide over access issues that was clearly revealed in the wake of the commission's 2007 communication is still present to some extent today.

The need to formulate a common vision for research outputs, including publications, for Europe is now imminent given the broader goals of the European Union for the European Research Area and the Digital Agenda that has been announced. The European Research

[24] Riding the Wave. How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission, October 2010.

Area Board has called for a "new renaissance". This vision involves a united ERA directed towards solving the Grand Challenges of our time, in collaboration between public and private sectors and in interaction between science and society. Promotion of cohesion and an encouragement of excellence are also key pillars in this vision. An open scientific environment is regarded as the key component in achieving this vision, and the ERAB issued an "urgent appeal" that links this to the survival of our species: "If we succeed in creating a truly open environment for research and innovation to flourish across the ERA, we will fulfil our obligations to catalyze the new Renaissance and improve our species' chances of survival. If we do not, if we fragment into competing disciplines, industries, nations and regions, we will miss our historic responsibility to develop Europe and the world."[25] There is a clear call for coordination and consolidation across borders, sectors and disciplines and the second report from the ERAB, presented in 2010, moves this vision forward by formulating recommended actions. Action 6.1, in particular, will have an impact on a vision for publications "All outputs of publicly funded research are available via 'open access' to all interested parties, and universities undertake a broader role in science communication."[26]

European Commission communications presented in 2010 on A Digital Agenda for Europe and the Innovation Union also have consequences for scientific publications. The former communication states that "[...] publicly funded research should be widely disseminated through Open Access publication of scientific data and papers" and that "[...] the Commission will appropriately extend current Open Access publication requirements [...]". The communication on the Innovation Union specifically notes in Commitment 4 that the ERA framework that will be proposed in 2012 "[...] will notably seek to ensure through a common approach [...] dissemination, transfer and use of research results, including through open access to publications and data from publicly funded research." Open access to publications is also addressed in Commitment 20: "The Commission will promote open access to the results of publicly funded research. It will aim to make open access to publications the general principle for projects funded by the EU research Framework Programmes [...]."

The policy direction and vision for the ERA, and with it the Digital Agenda and Innovation Union, suggest that it will be important for the European Commission to formulate a vision for publications/scientific communications in the forthcoming communication on scientific information in the digital age or to provide a framework for creating this common vision. A vision for research outputs must be coordinated with other policy areas related to the realization of the ERA and broader economic development goals of the European Union. This type of coordination can only be facilitated at a higher (European) level, with input from necessary stakeholder groups.

The key questions posed by PUBLICATIONS group at the workshop provide a starting point for formulating a vision, particularly "What has to be accessible, to whom and how?" Once these questions have been answered, it will be possible to move forward

[25] European Commission (2009), Preparing Europe for a New Renaissance. A Strategic View of the European Research Area, First Report of the European Research Area Board, p.5.

[26] European Commission (2010) Realizing the New Renaissance. Policy proposals for developing a world-class research and innovation space in Europe 2030. Second Report of the European Research Board.

with the recommendations of the PUBLICATIONS group to carry out an assessment of the existing system (to understand what transformations need to take place to the current infrastructure and across all stakeholders), to identify the infrastructural barriers and innovation necessary to achieve the vision, as well as to understand what will constitute a sustainable model for the future and how to fund this. The process followed by the High Level Expert Group on Scientific Data might provide some inspiration for formulating a vision, as both publications and data constitute the key outputs of scientific endeavour.

It is also worth noting that despite the continued controversy over access issues, in the four years since the European Commission presented its communication on scientific information in the digital age much has happened to change the publishing landscape. In particular, the open access publishing sector reveals a very different picture today than in 2007. The Public Library of Science has demonstrated a positive financial result, and BioMed Central became profitable and was furthermore acquired by Springer, one of the largest subscription publishers. The financial stability of these publishers can be added to that of Hindawi Publishing Corporation. Smaller enterprises such as Copernicus Publications and Co-Action Publishing have also demonstrated that open access models can provide the basis for a successful small or medium enterprise in Europe. It is also significant that Springer announced its in-house open access portfolio Springer Open (which is quickly growing), followed shortly by an announcement earlier this year by Wiley-Blackwell about the launch of their own open access portfolio. Other publishing houses such as Nature Publishing Group have also announced experiments with open access publishing. In fact, a large number of members of the Open Access Scholarly Publishers Association (OASPA) are mixed model publishers, operating both subscription and open access portfolios.[27] This new constellation, as well as the research emanating from the SOAP project and PEER can provide additional input and support to the process of formulating a vision.

[27] These publishers include the American Institute of Physics, the American Physical Society, The BMJ Group, Institute of Physics (IOP), International Union of Crystallography, Oxford University Press and Sage Publications, in addition to Springer Science + Media.

CONCLUDING REMARKS

The workshop on EC-funded projects on scientific information in the digital age was unique in that it brought together a diverse group of individuals working on diverse projects, who might not otherwise have had an opportunity to share and discuss their projects and the future of scientific information together. What looked to be a strange mix of company provided an opportunity to gain insight into the overall state of scientific information in the digital age. The discourse that emerged during the two days gave an opportunity to identify points of convergence, the relevant importance of different issues for different areas of activity, and to contrast different scholarly communities in relation to transformations taking place in the context of digital technology.

An important theme that emerged is that scientific information is entering a phase of consolidation, in which a greater degree of coordination and efficiency can be detected. However, this process is progressing at an uneven pace across areas of activity within scientific information, between scholarly communities and across countries. As demonstrated through the discourse of the workshop some scientific communities are able to explicitly envision and state future needs and drive change in the systems surrounding scientific information to meet these needs. In contrast, other research communities are being transformed by technological advancements, and must be convinced that solutions indeed meet needs they might not be aware they had. An important question in this context is how can we learn from those communities who are driving change? This suggests that the European Commission can continue to play an important role in supporting developments and coordinating efforts across member states and across scientific communities, as well as across stakeholder groups and areas of activity.

When juxtaposed against the fields of scientific data, preservation and infrastructures, it is striking that a vision is still lacking for one of the key outputs of research activity: publications. As argued above, the broader political goals of the European Union suggest that it will be important that such a vision be formulated. While there is clearly more contention amongst the stakeholders in this area of activity than in others, there is nonetheless a need to define the role publications can play in the future, what needs to be accessible and how. Without such a vision, there is a risk that the overall system will develop and begin to enter a third phase of consolidation among areas of activity, without a clear understanding of the role to be played by scholarly communications. Moreover, recent policy developments linked to the ERA place further pressure upon the need to create a common understanding of open access to scientific outputs for stated policy goals to be achieved.

Finally, key issues surfaced in the discourses of the workshop. These were technological, sociological/behavioural, Financial Sustainability, Legal and Overall Vision. Like the other themes, these issues are unevenly distributed in terms of their relevant importance to the areas of activity. However, all of these issues do surface in some way in all areas of activity. Financial sustainability, in particular, was of great importance to all projects and areas of activity. The European Commission can play a continued role in working with stakeholders to address these issues through dialogue, policy and other means.

ANNEXES

ANNEX 1- Participant list

Participants	Institution	Project
Amanda Wren	European Cancer Organisation	EUROCANCERCOMS
Andi Aschenbrenner	Goettingen University	DARIAH
Beatriz Barros	Malaga University	SISOB
Birgit Schmidt	Goettingen University	OPENAIRE
Bjørn Henrichsen	Norwegian social science data service	CESSDA
Carl-Christian Buhr	European Commission CAB INFSO	
Carlos Morais Pires	European Commission DG INFSO, F3	
Caroline Sutton	Co-Action Publishing Rapporteur	
Celina Ramjoué	European Commission DG RTD, B6	
Christoph Bruch	Max Planck Digital Library	PEER
Costantino Thanos	Centro Nazionale Ricerche – ISTI	GRDI2020
David Giaretta	Science and Technology facilities Council	APARSEN
Deborah Kahn	BioMed Central	SOAP
Eelco Ferweda	Amsterdam University Press	OAPEN
Fabio Casati	Trento University	LIQUIDPUB
Francesco Fusaro	European Commission DG RTD, B6	
Françoise Genova	Strasbourg Observatory	EUROVO-AIDA
Gilles Laroche	European Commission DG RTD, B6	
Hans Jørgen Marker	Swedish National Data Service	CESSDA
Hans Pfeiffenberger	Helmholtz Association	ODE
Hilary Hanahoe	Trust-IT services Ltd	GRDI2020
Ignasi Labastida	Barcelona University	COMMUNIA
Ingrid van den Neucker	European Cancer Organisation	EUROCANCERCOMS
Jarkko Siren	European Commission DG INFSO, F.3	
Jean-François Dechamp	European Commission, DG RTD, B6	

>>>

Participants	Institution	Project
Juan Carlos de Martin	Nexa center for internet & society	COMMUNIA
Juan Pellegrin	European Commission, DG INFSO, E4	
Julia Wallace	International Association of scientific, technical & medical publishers	PEER
Leif Laaksonen	IT Centre for Science	E-IRGSP2
Matthieu Kleinschmager	European Commission DG HR, B3	
Michael Chatzopoulos	Athens University	OPENAIRE
Michael Mabe	International Association of scientific, technical & medical publishers	PEER
Natalia Manola	Athens University	OPENAIRE
Paola de Castro	Istituto superiore sanità	NECOBELAC
Patricia Reilly	European Commission CAB RTD	
Per Öster	IT Centre for Science	EGI
Peter Doorn	Royal Netherlands Academy of Arts	DARIAH.
Peter Wittenburg	Max Planck Institute	CLARIN
Rossend Llurba	The Netherlands Organisation for Scientific Research	E-IRGSP3
Salvatore Mele	CERN	CERN, SOAP , ODE
Saskia C.J. de Vries	Amsterdam University Press	OAPEN
Steven Krauwer	Utrecht University	CLARIN
Wim van der Stelt	Springer	LIQUIDPUB
Wouter Schallier	Stichting LIBER	ODE

ANNEX 2 – Responses to Key Learning on Day 1

Day 1 of the workshop closed with a session that asked each participant to express what his or her key learning was from the day. These expressions were recorded on cards and read aloud to the group. The recorded statements are listed below and grouped according to topic addressed. The grouping was provided by the workshop facilitator:

On the diversity of all projects:

- There are many projects doing similar things but there is a need to connect them
- Differences among disciplines, data sharing, short duration for validation and sustainability, projects with policy recommendations
- Impression of diversity and communicability of projects
- Diversity of activities= long way to convergence. Terminology: What is Infrastructure?
- 45 people, 18 projects, 1 EC: more similarities than it seems ... tear down the barriers for e-Science
- Wide range of projects addressing wide range of issues of diverse magnitude. Greater need to compare like with like
- Projects that address different levels of system with different roles and stakeholders at each level
- Different views and approaches regarding the management of some phases of the life cycle of scientific information
- Projects provide interesting features about the dissemination of scientific information

On our own projects:

- More confidence that EGI is timely and will have a relevant offer for many areas of science
- New solution for sustainability of OAPEN

On the future of scientific information in the digital age:

- OA publishing will go full speed ahead, the business obstacles are not difficult to overcome
- The question about sustainability: what, why, how?
- The need of a more generic approach to scientific repositories development
- Need to pool existing resources more effectively
- Words will fill the time available, Projects will use the time available, Evaluation is difficult
- I have never fully thought about the full life-cycle of scientific data. Time to start
- More coordination
- Research and scientific information have the same problems and challenges in everywhere: how to share? How to measure? How to distinguish the quality of results?
- A better understanding of how complex policy making is for this area

- Importance of feedback and exchange of information from stakeholders in understanding the process behind certain decision making leading to the financing of certain projects
- Policies on data access need to be established, researchers will follow
- Need for sustainability + evidence-based decision making
- There is material to provide the EC with recommendations for future policy initiatives

On our process:

- This workshop is a good initiative
- Insight in diversity & problems
- Speaking is power!
- Communication is fundamental for acquiring and sharing new knowledge
- Story telling broadens perspectives and helps to discover similarities, differences and important issues
- A brave try to analyse and connect very different projects, results
- Share common problems, learn from different solutions, and establish useful contacts
- Listen interactively to knowledgeable colleagues tell about their projects and understand what is the most valuable for me. Less excited about the meta-question: too little time for that
- Learned many useful details, learned by floating that everybody is asking the same question as I am
- Good to learn about other projects, more session like this but shorter, telling about a project is excellent, doing something is more than excellent.
- Researchers' behaviour
- Networking v Barriers, Sharing
- Divergence by nature, convergence by will (work in progress)
- A sense of being at half-course in our process; how will today seed tomorrow?
- Users in mind.
- Uncertainty

ANNEX 3 – Surfacing individual questions on Day 2

Individual questions were raised by participants at the start of Day 2 and recorded on cards. The following is a list of these questions as they were stated on the cards:

- How can we efficiently and freely share and find all forms of quality scientific knowledge, data and opinions?
- Where is quality information in the digital age? Where is the thinking ability?
- Missing (e-)infrastructures for the data process. What is data? Clear responsibilities scientists ←→ society and society ←→ scientists
- How to apply the great advances of ICT on really improving scientific age? Find simple solutions to what seem to be big problems (sustainability, real access to data, communication, quality data)
- What will happen to quality control / peer review in the transition to an OA scientific society where research will alter too?
- Awareness: how to increase awareness researchers about sharing?
- How to get the data out of the cardboard boxes of scientists (USB-sticks) (Nature editorial 2005)? Incentives – Resources for additional efforts – Clearly understood best practise
- How can we create an environment that encourages and reassures scientists to share and access data?
- How to achieve the commitment of researchers' communities for sharing? Putting policies into practise: sharing of research results needs the commitment of research communities
- How is it possible to handle effectively and efficiently the scientific information and in particular the scientific data without defining formal models for representing the different kind of information (meta-data, provenance context, uncertainty, quality)?
- How can we ensure that all scientific information can be found, assessed, preserved, and trusted?
- Data and information itself: many aspects including incentives, trainings, etc but not only. The key is to have data and information in a form which allows reusing it: formats, metadata, semantics, etc... disciplinary and interdisciplinary discussions, progressive bottom-up approach (complementary to top-down).
- Linking data: How can we link information in the silos of the repositories, especially research data, publications and research information (about persons, institutes and projects)?
- How to get from collections of data to an information infrastructure?
- How can we use unfamiliar data (time/discipline) in an automated way? Integrate the data into new/different s/w
- How to find relevant / trusted data among the mass of things out there?
- With the incredible increase of heterogeneous information (digital libraries, repositories, archives and the web), how can we ease the researchers' life in searching and especially in cooperating with scientists in interdisciplinary areas?
- How do we organise ourselves to secure easy access to data and information?
- How do I get open access to data and when I do, what tools do I have at my disposal so that I can use it efficiently and trustworthy?

- How to make and keep data “fluid”? Counter movement needed in response to the strong movement towards lock-in and confinement of data, e.g. some aspects of cloud computing?
- How to unleash the power of digital technology to optimise the process of communicating, evaluating, searching, (re-)using and archiving scientific information? So far we are perhaps at 10% of potential?
- How to manage the huge amount of data, keeping it available, accessible, affordable and sustainable? Will a pan-European storage/data infrastructure help?
- Evidence-based providing a framework for the development of an open access infrastructure
- Sustainability of scholarly ecosystems
- How do we ensure sustainable solutions with buy-in from key stakeholders?
- What overall system do I want to see implemented for researchers and the society of the next generations?
- How can we ensure that research publications are freely available as a resource for diverse (re-)usage for profit and not for profit?
- What can the EC do to ensure all publicly funded research is made openly available to all without barriers?
- How do we make scientific information open access in the digital age without incurring additional costs?
- How can the academic community make the transition to Open Access?
- How to organise the transition to new sustainable access models without disrupting the system; what is the role of the government in that? Keep on developing new opportunities in the system?
- What are the roles in European infrastructures for scientific information and how do the roles of existing players change? e.g. data centres, libraries, research institutions, etc.

European Commission

Sharing knowledge: EC-funded projects on scientific information in the digital age

Luxembourg: Publications Office of the European Union

2011 – 49 pp. – 17.6 x 25.0 cm

ISBN 978-92-79-20451-7
doi:10.2777/63780

How to obtain EU publications

Publications for sale:

- via EU Bookshop (http://bookshop.europa.eu);
- from your bookseller by quoting the title, publisher and/or ISBN number;
- by contacting one of our sales agents directly. You can obtain their contact details on the Internet (http://bookshop.europa.eu) or by sending a fax to +352 2929-42758.

Free publications:

- via EU Bookshop (http://bookshop.europa.eu);
- at the European Commission's representations or delegations. You can obtain their contact details on the Internet (http://ec.europa.eu) or by sending a fax to +352 2929-42758.

The report presents and analyses discussions that took place at a workshop primarily organized by the European Commission Directorate-General for Research and Innovation, with the participation of the Directorate-General for Information Society and Media. Leaders from eighteen projects supported under different European Commission programmes, including ICT Policy Support (formerly eContentPlus), Science in Society, and Infrastructures (under the FP7 Capacities programme) attended the workshop. The overarching purpose of the workshop was to discern learning from the projects and to reflect upon how this learning can be translated into policy recommendations and concrete actions on scientific information in the digital age.

KI-32-11-771-EN-N

doi:10.2777/63780

ISBN 978-92-79-20451-7

Publications Office