OpenAIRE — Building a Collaborative Open Access Infrastructure for European Researchers

Najla Rettberg

Göttingen State and University Library,
University of Göttingen, Germany,
najla.rettberg@sub.uni-goettingen.de

Birgit Schmidt

Göttingen State and University Library,
University of Göttingen, Germany,
bschmidt@sub.uni-goettingen.de

Abstract

This paper outlines the efforts of the OpenAIRE networking team to establish a Europe-wide open access initiative. OpenAIRE is an effort to realize the open access policies of the European Commission, and has built an infrastructure to support the widest possible dissemination of project results within a certain funding area, FP7. The purpose of the paper is to highlight how such a service can be established through the work of a successful network of European open access desks and by effective communication with a range of stakeholders. The paper also outlines the flexible technical infrastructure and research activities within the project. Not without its challenges, the approaches to tackling existing barriers, such as achieving repository interoperability, are explored. The paper also introduces the aims and initial activities of the continuation project, OpenAIREplus.

Key Words: Open access, e-infrastructures, OpenAIRE, repositories.

Najla Rettberg and Birgit Schmidt

Introduction

The overarching goal of OpenAIRE is to provide an infrastructure and support network for enabling free and open access to peer-reviewed research publications, which specifically result from seven disciplinary areas[1] within the European Commission's (EC) Seventh Framework Programme (FP7) tranche of funding. Now in its second phase, the project will grow its range of publications, and link out to associated data and funding information.

A pan-European project, OpenAIRE[2] is based on a network of librarians, repository managers and technicians from across 33 European countries involving all EU Member States and in addition, Norway, Iceland, Croatia, Switzerland and Turkey. OpenAIRE combines the strengths of both a distributed and a centralized approach, linking up virtually remote repositories and bringing them together into a central searchable environment. Its versatile technology is fully supported by a network of European advocacy offices, which champion open access, disseminate information and provide guidance. This networked team has successfully set up an infrastructure and promotes open access in all partner countries, in particular access to peer-reviewed publications resulting from EC-funded research. More generally, the need for a systematic and joined up infrastructure for research is becoming increasingly vital in today's e-environment, a challenging feat in the emerging repository landscape which has a history of both institutional, centralized and disciplinary repositories (Aschenbrenner et al., 2010). The success of OpenAIRE will be measured by how researchers interact with the service, how usable they find it, and how the mandate is adhered to (Hagerlid, 2010).

OpenAIRE is strategically placed as an e-Infrastructure to support research communities, as Europe heads towards Horizon2020, the EC's next funding programme (2014–2020) which will fully support open access and initiatives which contribute to an open scientific environment.[3] It will also serve to make the European library community, and research administrators aware of open access, data practices and the need for robust scholarly communication management. OpenAIRE is a truly European venture and a key building block in the European research e-Infrastructure (Lossau, 2012).

This article describes the project's aims and achievements to date, from the point of view of the networking team. It highlights how its existing challenges

are being tackled, and looks forward to the next stage of its development, via the OpenAIREplus project.

A European Commonwealth of Learning

At the launch of OpenAIRE in Ghent, December 2010, Neelie Kroes, Vice President of the European Commission responsible for the Digital Agenda, made reference to open access encouraging and enabling greater engagement of society in science. This was reiterated by John Willinsky with reference to John Locke's philosophies about "a commonwealth of learning". In short, the community should reach for a continent-wide embrace of open access (Willinsky, 2011).

The aims of OpenAIRE are three-fold:

- To provide a unique e-infrastructure for scientists to access open access publications. This is via self-deposit either directly to OpenAIRE or to their repository of choice, be it institutional or subject oriented.
- To support the establishment of open access policies at a European level.
- To make usage statistics available to policy makers, and in doing so, link publications to project information.

In a continuation project, OpenAIREplus builds on this to link publications and projects to associated research data (more details below).

These aims are also aligned to support the open access policies established by the European Commission. In December 2007, the European Research Council published its Guidelines for Open access, as a follow-up of the 2006 Statement on Open access. In August 2008, the European Commission launched the Open access Pilot in FP7 that will run until the end of the Framework Programme.[4] The pilot covers 20% of the FP7 budget, and commits researchers from seven thematic areas to deposit their research publications in an institutional or disciplinary open access repository via a mandate called Special Clause 39 (SC39). Open access to the publication has to be provided no later than within six to twelve months depending on thematic area. This 'green' approach to open access is combined with the opportunity to use FP7 project funds to publish in open access journals.[5]

From Local to Global

Populating the OpenAIRE portal with the metadata of peer-reviewed open access publications within seven specific areas of FP7 is based on connecting local institutional or subject-based repositories for metadata aggregation. It also allows researchers to upload a copy of their peer-reviewed articles in these repositories. If researchers are not equipped with an institutional or disciplinary repository, they can submit to the OpenAIRE 'orphan' repository, which is aimed at those researchers who have no institutional repository of their own in which to deposit publications. CERN (a partner in OpenAIRE) maintains this purpose-built repository.

While research is a global undertaking, any support structures such as policies, workflows and standards will have to be found at a local level. One of the main challenges OpenAIRE faces is to aggregate at a global level this local information, taking into account socio-cultural and regional differences of the communities involved (Lossau, 2012).

It has approached this challenge from the start by structuring itself around a localized networking and outreach community. Over 40 European partners, many of which are libraries, now take part in the project and work at local level to support population of a global system.

The back-bone of the project are the National Open access Desks providing support to researchers, repository managers and other stakeholders based on their knowledge of the national research environment. Each participating country contributes representatives of a dedicated institution which enables the network to cover all EU member states and five associate countries (see Figure 1).

Engagement with Different Stakeholders

The national contacts liaise with stakeholders in their countries to make them aware of the EC's open access policies and OpenAIRE's support measures, in particular reaching out to universities, libraries, research organizations and funders, projects and researchers. Moreover, through the networking partners, these activities are aligned with other open access and repository-related activities in Europe and internationally. For example, LIBER[6] SPARC[7] Europe

Fig. 1: The OpenAIRE Network. Red indicates new partners in OpenAIREplus.
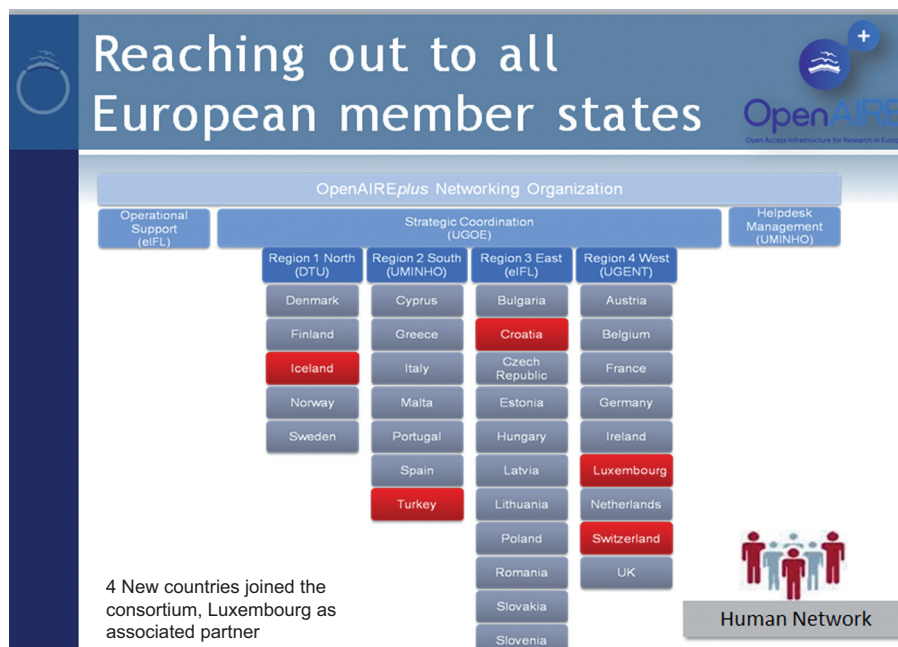
and COAR[8] promote repository networks and services and are developing a Roadmap for Repository Interoperability. They exploit this organizational structure to efficiently disseminate best practices, guidelines, initiatives, and events related to open access among local decision makers and research organizations (and vice versa).

In an effort to gain the commitment of a variety of stakeholders (which include research administrators, managers, repository managers, publishers, and project coordinators), a number of different approaches have been developed by the networking team. The OpenAIRE website holds information such as guides, FAQs,[9] as well as workflows to contribute information about publications resulting from FP7-funded projects. Services for project coordinators are also being developed, such as generating open access publication lists within project websites.[10] Additionally, to help ease the process of reporting open access publication numbers, tools for project administrators are also being developed, which analyse usage at project and publication level.

To address research institutions and researchers, toolkits have been created to assist in implementing viable open access policies and procedures. The recommendations are based on best practices from a range of leading universities and research institutions in Europe. The identification of publications resulting from EU-funded projects remains a challenge as there is no regular procedure to report about publications during the project period and beyond (some publications resulting from the project are published after the end of the FP7 project).

Repository managers will find FAQs addressing technical questions. Detailed instructions for joining a repository to OpenAIRE (also known as 'compliancy') have been published on the portal. OpenAIRE also provides support structures for deposition of papers by leading to the individual repositories, see Figure 2. To maximise impact, this has to be supplemented by local advocacy and support activities combining the efforts of libraries and institutional research support offices which administer externally funded research activities.

Both repositories and journals can now be registered and validated within the OpenAIRE information space. Add-ons for DSpace and EPrints are available. A plug-in supporting OpenAIRE compliance for Open Journal Systems (OJS) was released in February 2011.

Publishers and journals are also invited to join and to contribute information about publications resulting from EC-funded research. Journals based on the OJS platform can use the above-mentioned plug-in to adapt their system. Various publishers have been and will be approached to explain the benefits of joining and what OpenAIRE can offer to enhance workflows (such as an

Fig. 2: Researcher's Workflow.
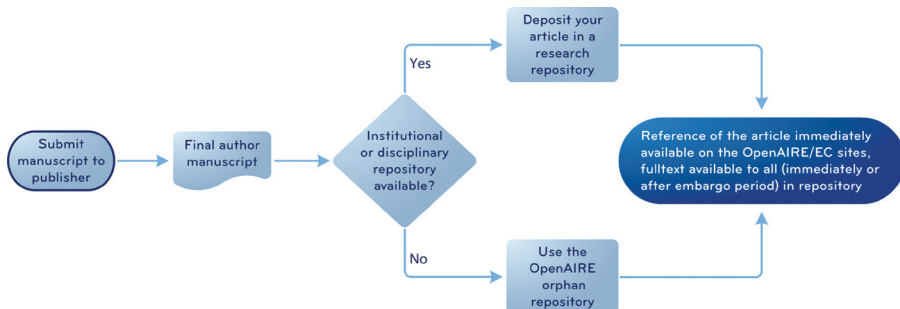
API to integrate project information and tools for mining publications in relation to projects). In the future, data journals might provide data sources for OpenAIREplus.

The project focuses on the EC as the main funder, but as the project progresses, other funding schemes at the national level will be targeted. Currently, clear communication by EC officers towards the funded projects is essential. Even if awareness is growing, open access as such is not familiar among all research areas. Sometimes there might be confusion on how open access can be realized. For example, awareness should be raised that simply adding publication lists to project websites is a non-sustainable way of providing information because they cannot be harvested by OpenAIRE and are not stable.

Benefits for Users

On top of this infrastructure, OpenAIRE has built a number of services for the user. The OpenAIRE portal is the backbone of the project. Constantly developed and refined, it offers functionalities for project administrators, anonymous and registered users to manage their publications linked to project funding data, and an Information Space of publications, together with their connections with funding projects (from the EC and national agencies) and in the future with research datasets. The networking team works to promote these services via its outreach work across Europe.

The OpenAIRE portal additionally reflects the state of open access in all the participating countries, and is translated into respective languages. Sections are targeted to specific stakeholders: researchers, project coordinators, and repository managers.

Clarifying copyright issues comprises a section of the portal as this is a major concern of researchers when it comes to deposition of publications. The majority of publishers and journals allow self-deposit of articles but some publishers demand time embargos which may vary from journal to journal. OpenAIRE provides copyright guidance, and sample letters to publishers.

This is complemented by a helpdesk[11] based on the network of national contacts dealing with issues related to deposit and open access publishing.

In addition, OpenAIRE offers online and on-site trainings. These aim to explore approaches and strategies in how to communicate with different types of researchers, how to use social media to attain OpenAIRE's goals and how to identify and track publications (by using tools like Zotero and Web of Science). Interactive tutorials are also held on a regular basis. Topics covered so far have recently targeted repositories and journals to become OpenAIRE compliant.

The Technical Approach

OpenAIRE is based on a uniquely flexible infrastructure. It is capable of harvesting and processing data from a range of sources: compliant open access institutional and disciplinary repositories, journals and publishers as well as project information from the EC CORDISdatabase.[12] The range of data sources will be successively extended in OpenAIREplus covering research management systems (CRIS) and metadata from dataset archives.

The software services of the OpenAIRE data infrastructure are based on the D-NET Software Toolkit[13] which enables a number of services such as the ability to manage heterogeneous data sources, normalising data (i.e. cleaning and transforming data) and ultimately giving access to users as well as third party systems such as search engines, and aggregators. Other services that will be developed and are of interest to third party providers will be: citation analysis, content classification, text mining to infer links to research data from publications and working with integrating IDs for authors via ORCID, and data, via DataCite.

In order to enrich the metadata, it is possible to automatically infer links between projects and publications by text mining publications. OpenAIRE then offers these links to project coordinators, asking them to confirm the automated 'guesses'. In future, this mining will be extended to other national funding schemes.

So that it can deliver reliable statistics, the infrastructure can also deal with duplications by merging and disambiguating publications. All the information is available via the OpenAIRE portal, which offers the user advanced search, browsing mechanisms and, in a future portal release, statistics of usage per publication, see Figure 3.

Fig. 3: Usage statistics at publication level.

Achieving Interoperability

Achieving interoperability between repositories is another recognized challenge area (Lossau, 2012). Part of OpenAIRE's mission is to promote and contribute to the increase in the number of OAI-PMH (Open Archives Initiative, Protocol for Metadata Harvesting) compliant repositories as a basic infrastructure to support the implementation of open access policies and mandates. These figures have been growing substantially over the last years with now nearly 1,200 open access repositories across Europe (up by over 40% since January 2011, data from OpenDOAR,[14] and ROAR[15]).

The OpenAIRE Guidelines (OpenAIRE, 2010) provide repository managers and journals with instructions to define and implement their local data

management policies in compliance with the open access demands of the European Commission. They comply with the technical requirements of the OpenAIRE infrastructure which is being established to support and monitor the implementation of the FP7 Open access Pilot and ERC Guidelines for Open access.

Moreover, for developers of repository and journal publication platforms, the guidelines provide guidance on adding supportive functionalities for authors of EC-funded research in future versions. Specific add-ons are available for repository (DSpace, ePrints) and journal platforms (Open Journals System).

A 'Weak Mandate' and Other Challenges

Adherence to the open access mandate based on the Special Clause 39 is the main barrier to growing publication numbers. While many researchers support open access, many are unclear about how to deposit and about copyright issues. Incentives for depositing are not strong enough and clearer messages of the benefits of open access need to be made. A multi-pronged approach could help to make the mandate stronger while providing incentives and support. For example, funders should equip project officers with a clear message that the mandate is compulsory, and provide clear guidelines for researchers. Lists of journals that support the mandate should be promoted and in some cases agreements with publishers for automatic deposit could be set up. The OpenAIRE initiative can support all these issues by providing an infrastructure for user-friendly deposit.

An important issue of discussion and work in this third year has been how to get OA repositories in Europe to register and engage with OpenAIRE. A number of concerted efforts have been made by the networking team such as webinars on 'compliancy', and detailed guides (mentioned above). To identify the reasons behind low compliancy, the OpenAIRE community was surveyed in December 2011. Barriers were identified such as the number of other compliancy activities (CERIF, etc) that have yet to be undertaken. Repository managers also expressed scepticism that there aren't enough FP7 publications evident as yet in local repositories, reducing the necessity for compliancy.

With this in mind, another solution to growing publication numbers is text mining to identify publications' relationships to FP7 projects. Large

repositories play a significant role in the dissemination of research results in a number of research areas and prominent repositories and networks such as ArXiv, UKPMC, and DRIVER will be mined.

Research Activities

The Community Development Report (Schmidt and Kuchma, 2012) provides a unique examination of national awareness of open access in 26 member states and one associate EU member (Norway). Moreover, it describes funder and institutional OA mandates in Europe, national implementation strategies, repository infrastructures, support networks and outreach and dissemination strategies practiced so far.

Another key research area in OpenAIRE comprises a series of studies on subject specific requirements for open access infrastructures. Given that research communities are diverse in practice, building a horizontal supporting infrastructure comes with challenges: "…any roadmap for OA infrastructure must address this natural tension between diversity and infrastructure" (Meier zu Verl and Horstmann, 2011, p. 360). Exemplifying the fields in the EC open access pilot, the study put forward the requirements that result within each field for research data. This is with a view to creating a 'new generation' information service that could exploit open access principles from a disciplinary perspective and support funding requirements.

A series of legal studies will be undertaken which will examine the legal implications in reuse of data. This will investigate and address the legal requirements for different kinds of usage of research data and metadata in an open access e-infrastructure focusing on publication-data links.

The 'Plus' in OpenAIRE

The OpenAIREplus project (Dec 2011–May 2014) continues and extends the scope of OpenAIRE to promote and monitor Open access to a wider audience and to more research output types. More specifically, it aims at growing a richer graph of data, covering material from all research disciplines and further countries, and including projects from national funding schemes, non-peer-reviewed publications, and research datasets.

Ultimately, OpenAIRE will extend its service to push beyond the realm of publications within FP7 and to widen the publication scope, but also to link publications to associated research data. This interlinking and reuse of research is seen as crucial to growth and innovation as Europe heads towards Horizon2020 and supporting an infrastructure for Open Science, for which "Interoperability is the key. It's the key to global, multi-disciplinary science, supported by reliable and high-performance data infrastructure" as Neelie Kroes recently pointed out (Kroes, 2012).

The first phase is largely exploratory. On the networking side, the project has started to raise awareness among OpenAIRE members about data management and about identifying appropriate data content from among the distinct geographic regions of this collaborative project. One of the first tasks is to use the existing European community base to scope the research data management landscape in Europe. Six questions were asked, primarily about regional data management initiatives both at funder and institutional levels. On first sight, the results display a concerted lack of robust data management practices within the 33 countries surveyed. Out of 21 responses, only three respondents answered 'yes' to being aware of any data policies at institutional level within their country.

Dissemination activities will be intensified and aim to bring in more stakeholders, for example, publishers, large data repositories, and national funders. The project will hold at least four community events over the next 24 months, the first being in June on policy for research data: "Linking Open access publications to data — policy development and implementation"[16]. Further workshops will focus on different topics such as interoperability, linking publications, and legal issues. They will aim to educate the community and involve experts from the open access and the research data management community.

This is also linked to local activities. For example, the University of Göttingen, as Scientific Coordinator of the project, recently held a data management workshop within the life sciences faculty. This kind of awareness raising events will feed into future training manuals on data management. Typical reactions from researchers include that some might see little connection between their work and the theoretical notion of sharing data, and feel powerless to change the traditional publication process. Researchers' comments also reflected the changing nature of the scientific process: "forget PDFs, imagine an ideal publication . . . where you can contribute

and discuss . . . and later update or correct parts of a paper in subsequent versions. . . . Many look back on their work and after a year, they already see it differently" (N. Rettberg, personal communication, April 4, 2012).

At a technical level, one of the objectives is to experiment with interlinking datasets and publications. This will be done via the construction of so-called "enhanced publications" from across different disciplines, see Figure 4. The ability to link within a publication out to a citable database, or to other research material, will enable users to find, view, interact, and also create their own relationships between different information objects.

As mentioned earlier, creating a generic, 'horizontal' infrastructure capable of supporting a cross-disciplinary infrastructure is in itself challenging due to the fact that each scientific field deals with data in a different way (Meier zu Verl and Horstmann, 2011). OpenAIRE plus will build on the research scenarios outlined above, and by building example prototypes, the project will explore how links between text-based publications and research data are managed in different scientific fields. Three different scientific partners are involved in order to gather a deeper understanding of how data is managed and linked: European Bioinformatics Institute[17] (EBI-EMBL), British Atmospheric Data Centre[18] (BADC), and The Data Archiving and Networked Services[19] (DANS).

Fig. 4: Linking Publications to Data and Funding.
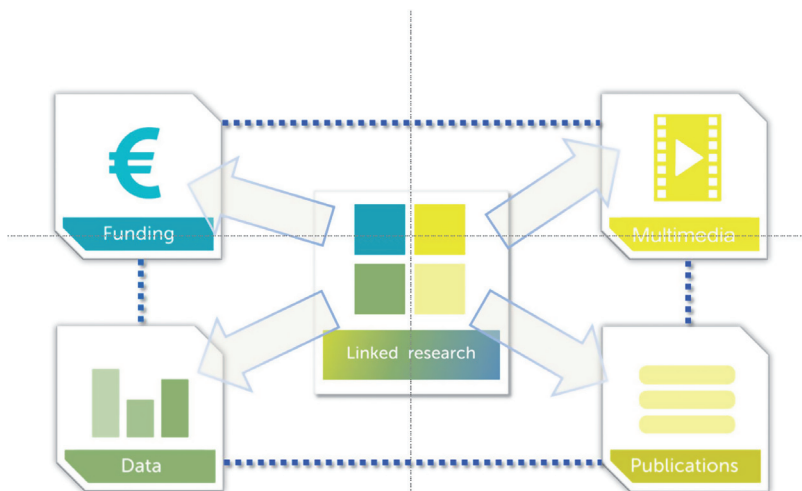
Conclusion

OpenAIRE has come far to contribute to the open access movement and to raise awareness at a European level, encouraging deposit at a local level, while also stimulating and enhancing the overall repository landscape. Now in its second phase, it will expand its value-added services towards research data and it will take steps to make researchers, librarians and funders aware of opportunities to make use of this interlinked information, in particular, publications, research data and funding information.

OpenAIRE has also highlighted challenges of setting up a network of repositories and gathering information across Europe. Identifying publications resulting from EC-funding is a challenge that the project is tackling in a multi-faceted way. The initiative has to deal with the issue of a 'weak' mandate, and the (mostly social) challenges of interoperability and compliancy. In OpenAIREplus, we will see a combination of a generic approach serving all disciplines with disciplinary prototypes, gaining an insight into how scientific communities approach data sharing. Additionally, scoping training requirements for librarians as tomorrow's 'data scientists' will lead the project fittingly towards Horizon2020.

The establishment of a community of open access repository advocates, who support researchers, is an invaluable ingredient to the project's progress and dissemination across Europe. As substantial technical infrastructures continue to develop, it is clear that this growth has to be complemented by networking activities that give out clear, simple messages, and support the very people the system is built for: scholars and researchers. Above all, the experience of the librarians who work with repositories and researchers will find opportunities to further their knowledge about open access in practice and to enable linking of scholarly communication outputs to projects and other information. In this sense, the truly community-led flavour of OpenAIRE makes it a promising future service for tomorrow's researchers.

Acknowledgements

References

Aschenbrenner, A., Blanke, T., Küster, M., & Pempe, W. (2010): 'Towards an Open Repository Environment'. *Journal of Digital Information*, 11(1). Retrieved August 6, 2012, from http://journals.tdl.org/jodi/article/view/758/646.

Hagerlid, J. (2010): 'OpenAIRE – European repositories jointly supports the EC open access pilot'. *SciecomInfo*, 6(1). Retrieved May 30, 2012, from http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/3551/3125.

Kroes, N. (2012): Neelie Kroes Vice-President of the European Commission responsible for the Digital Agenda making Open access a reality for science Publishing and the Ecology of European Research (PEER) project conference Brussels, 29 May 2012. Retrieved May 30, 2012, from http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392&format=HTML&aged=0&language=EN&guiLanguage=en.

Lossau, N. (2012): 'An Overview of Research Infrastructures in Europe – and Recommendations to LIBER'. *LIBER Quarterly*, 21(3/4), April 2012. Retrieved August 6, 2012, from http://liber.library.uu.nl/index.php/lq/article/view/8028.

Meier zu Verl, C., & Horstmann, W. (Eds.) (2011): *Studies on Subject-Specific Requirements for Open access Infrastructure*. Bielefeld: Universitätsbibliothek. Retrieved August 6, 2012, from http://pub.uni-bielefeld.de/publication/2445229.

OpenAIRE (2010): *OpenAIRE Guidelines 1.1 – guidelines for content providers of the OpenAIRE information space*. Retrieved August 6, 2012, from http://www.openaire.eu/component/attachments/download/79.

Schmidt, B., & Kuchma, I. (2012): *Implementing Open Access Mandates in Europe – OpenAIRE study on the development of open access repository communities in Europe*. Amsterdam University Press.

Willinsky, J. (2011, January 12): The Enlightenment 2.0: An Open Letter on OpenAIRE. A message posted on SLAW. Retrieved May 30, 2012 from http://www.slaw.ca/2011/01/12/the-enlightenment-2-0-an-open-letter-on-openaire/.

Notes

1 The areas are: Energy, Environment, Health, Information and Communication Technologies, Research Infrastructures, Science in Society, Socio-economic sciences and the humanities.

2 http://www.openaire.eu/.

[3] http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf.

[4] http://www.openaire.eu/en/open-access/open-access-in-fp7.

[5] There are currently two complementary strategies to achieve open access to scholarly literature: The green route to open access refers to depositing research output (e.g. peer-reviewed journal articles) in institutional or subject repositories (access can sometimes be delayed due to publisher's embargo period). The repositories do not carry out the peer review, but make research output freely and openly available and increase visibility.
The gold route is for an author to publish in an open access journal. Open access journals use a funding model that does not charge readers or their institutions for access (subscription or access fees) and do no longer invoke copyright to restrict access to and use of the material they publish. Instead they use copyright and other tools to ensure permanent open access to all the articles they publish. Publication costs are usually carried by institutions or funding agencies.

[6] LIBER: http://www.libereurope.eu/.

[7] SPARC Europe: http://www.sparceurope.org/.

[8] Confederation of Open access Repositories (COAR): http://www.coar-repositories.org/.

[9] http://www.openaire.eu/support/faq?lang=en#moofaqCat-78.

[10] http://www.openaire.eu/en/component/content/article/76-highlights/398-project-coordinators-its-for-you.

[11] http://www.openaire.eu/en/support/helpdesk/.

[12] CORDIS is the Community Research and Development information service of the EC. Its purpose is to provide an portal to all EU funded projects and results. http://cordis.europa.eu/.

[13] D-NET: http://www.d-net.research-infrastructures.eu.

[14] www.opendoar.org.

[15] http://roar.eprints.org.

[16] A full write-up, including videos of this event can be found at the OpenAIRE portal, http://www.openaire.eu/en/component/content/article/76-highlights/393-openaires-first-workshop.

[17] European Bioinformatics Institute (EBI-EMBL): http://www.ebi.ac.uk/.

[18] British Atmospheric Data Centre: http://badc.nerc.ac.uk/home/index.html.

[19] The Data Archiving and Networked Services (DANS): http://www.dans.knaw.nl/en.