



European Public Sector Information Platform

Topic Report No. 2014 / 10

OPEN RESEARCH IN EUROPE - OpenAIRE

Author: Najla Rettberg

Published: October 2014

Table of Contents

	Keywords:	3
	Abstract/ Executive Summary	3
1	What is Open Access?	4
2	Open Access in Europe	7
3	What is OpenAIRE?	8
4	Interoperability and use of standards	11
5	What are the Services and who uses them?	13
	5.1 Who are the Users?	14
6	Heading to Horizon2020	16
7	Added value services	17
8	The outlook for OpenAIRE	18
9	Conclusions and recommendations	19
	About the Author	21
	Copyright information	21



Keywords:

Infrastructure, Open Access, European Commission, Mandate, Open Data

Abstract/ Executive Summary

Neelie Kroes, Vice President of the European Commission, puts it: we are "entering the era of open science", and "the impact will be good for citizens, good for scientists and good for society" (Kroes, 2013).

OpenAIRE is a European Open Access (OA) infrastructure, which means it gathers the research output funded by the EC (and other funders), to optimize visibility and provide free access to peer reviewed publications and associated research. This article describes how a European infrastructure can support the EC's open research agenda, the benefits that open research can bring to innovation, and how it plans to sustain the diverse services it provides.



1 What is Open Access?

Research outcomes, resulting from taxpayers' investment into science and research are a common good and should be made openly and freely available for all. According to the European Commission, open access (OA) is defined as "the practice of providing online access to reusable scientific information that is free of charge to the end user". This primarily refers to the peer-reviewed outcomes of research process, but is now widening its scope to include research data and other research outcomes (e.g. lab notebooks, grey literature, abstracts, conference proceedings).

The open access 'movement' has been growing over the years and its champions are far and wide. Traditionally, the outcome of research has been copyrighted by publishers and licensed to users and providers such as libraries. The starting point of the interest in OA was the Budapest declaration in 2002, the main aim being that research results should be reused without legal, financial or technical barriers². There are many stakeholders involved in open access publishers, policy makers, research institutions, libraries, and ultimately researchers not all of whom embrace OA, however attitudes to OA are changing,³ for example, a recent survey by a publisher concluded that attitudes to OA are changing and business opportunities can be seen in the landscape and that the belief that a sense of opportunity can be sought in a new open environment.

Open Research provides new opportunities: Open Research provides new opportunities, and arguments for open scientific practice in Europe are strong: a more open playing field leads to more cross-disciplinary research. OA does not have to be limited to the public sector — open results can also lead to richer innovation, particularly in the commercial sector. The exploitation of valuable research can be good for society at large and lead to wider business opportunities. Seen from the other perspective, the costs of cutting off access to research in the private sector could lead to a barrier in innovation. For example, the knowledge-based

¹ Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide en.pdf

² Budapest Declaration (2002). Budapest Open Access Initiative: http://www.budapestopenaccessinitiative.org

³ Learned societies more confident about future – and a 'new pragmatism' on Open Access: http://blog.alpsp.org/2014/08/learned-societies-more-confident-about.html



industry has free access to only 50 % of the public-funded research (Archambault et al., 2013)⁴, current access to the latest public-funded research and data is less than optimal, as a recent study revealed in attitudes to how levels of access are perceived. (RIN, 2011) which can be seen in the figure below⁵.

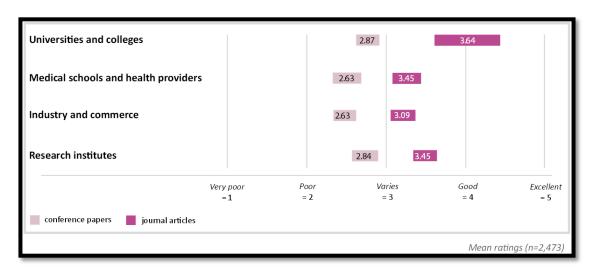


Figure 1. Current level of access to journal articles and conference papers by broad sector (only those for whom the resource is important). Source: RIN, 2011

Measuring the 'impact' of research: The ability to gather research output and the flexibility to carry out analysis such as text mining is also an added bonus of OA. The potential for tools to measure research impact in funding streams, such as citation analysis and usage pattern analysis are useful to future funding predictions and planning. In short, many value added services can come out of OA, and new communities and sectors of society can benefit accordingly.

Open access is not without its challenges. The role of the scientist in OA and their overall acceptance and uptake is probably the biggest challenge the open access movement faces. The most frequent question is that of the intellectual property rights over their own work, which is often signed away when a contract with a publisher is made. However, there are other options for a more open licensing approach. The Commission encourages authors to retain their copyright and grant adequate licenses (like Creative Commons CC-BY licenses). Not all research might be openly available immediately, often due to economic interests and other constraints. However, as soon as the decision has been taken to release a set of results to the public, the

_

⁴ Archambault, E. *et al.* 2013. Proportion of OA Peer-Reviewed Papers at the European & World Levels 2004-2011 at http://www.science-metrix.com/pdf/SM EC OA Availability 2004-2011.pdf

⁵ RIN, 2011 http://www.rin.ac.uk/node/1172



default set at "open" serves the interests of the wider scholarly community and the general public.



2 Open Access in Europe

The last decade has also seen a surge in interest at policy and funder level, and policy makers are realizing the benefits of opening up research.⁶ At the European level, many national policies at research funder and ministry levels have started to endorse OA via their policies. This has led to international declarations and recommendations for the implementation of open access (Global Research Council, 2013, and G8, 2013). One of the first relevant steps taken by the European Research Council (ERC) was a statement supporting Open Access (2006), shortly followed by guidelines for researchers funded by the ERC (2007) stating that all peer-reviewed publications from ERC funded projects should be made openly accessible shortly after their publication. Those guidelines were revised in October 2013, reinforcing the mandatory character of the requirements and expanding them to monographs⁷.

The European Commission took it first global political initiative in August 2008, launching an Open Access Pilot, covering seven areas (corresponding to about 20% of the total funding) of the 7th Framework Programme (FP7). All the grant agreements signed after August 2008 contained a clause (Special Clause 39) requiring beneficiaries to deposit articles resulting from FP7 projects into a repository and to make their best efforts to ensure open access to these articles

Implementing OA is the next step forward. The community needs tools, which will simplify the process and enable workflows to be embedded in local practices. It is in this context that we can look to the establishment of a research infrastructure to support these needs.

ePSIplatform Topic Report No. 2014 / 10 , October 2014

7

⁶ Suber, P (2012) Open Access (the book) http://cyber.law.harvard.edu/hoap/edit/Open Access (the book)

⁷ Open Access Guidelines for researchers funded by the ERC - http://erc.europa.eu/sites/default/files/document/file/ERC Open Access Guidelines-revised 2013.pdf



3 What is OpenAIRE?

OpenAIRE is an FP7-funded initiative which has built a European scholarly communication infrastructure that enables the aggregation of open access publications and research data catalogues, and importantly, has linked them to linking them with funding streams. On top of this graph of information, it provides services.

The origin of OpenAIRE lay in DRIVER (Digital Repository Infrastructure Vision for European Research)⁸, which created an e-Infrastructure of connected repositories, usually located in research institutions and libraries. OpenAIRE is a network of Open Access repositories, archives, and journals that supports Open Access policies. Importantly, the infrastructure follows a participatory approach with a European footprint employing 32 National Open Access Desks (all member states and 5 associate countries).

OpenAIRE harnesses the contents of publication and data repositories. It also collects information from OA Journals and other aggregators. At a technical level, it goes beyond the traditional publication aggregators by interconnecting several diverse entities related to scholarly communication and the overall research lifecycle, namely publications, research data, funding programmes, people, organisations, data sources, and eventually, software and instruments as well. The project allows users to navigate through a rich information space and provides a wide range of services, from research object deposition to statistics visualisations.

The OpenAIRE services and infrastructure function two-fold:

Human element: As mentioned above, OpenAIRE is a network of Pan-European advocacy nodes: key to OpenAIRE is understanding the users and consumers of the system and OpenAIRE works with a dedicated support office in each member state. Their role is to reach out to researchers and project coordinators in receipt of EC project funding, to make them aware of the EC open access mandate. The open access situation across Europe is very heterogeneous, and at different states of maturity. Thus the representative on the ground understands the local issues and is able to reach out with relevant messages, often in the local language. Europe has a wide and diverse repository landscape and the implementation of open access is very heterogeneous both at policy and infrastructural level. Member states are at different levels of repository development, some very advanced such as Germany and Britain.

⁸ http://www.driver-support.eu



In the Nordic countries, much research output is stored in research systems such as CRISs, and some countries take a central aggregation approach, such as the Netherlands and France. The good news is that there has been a recent growth in mandates, in 2013 there were 33, at both national funder level and institutional which will help support open access at national levels. By being part of OpenAIRE there is evidence that Europe has been a leader with its policy on OA and a catalyst for change in some countries. For example, Spain and Portugal's national research and development projects now mirror the EC's mandate, by encouraging deposit in repositories. In some countries, such as Croatia and Lithuania the EC open access policy has incentivized the establishment of local repositories.

Technical infrastructure: OpenAIRE is built on a fully open source infrastructure framework — OpenAIRE is based on the D-NET Software Toolkit, conceived to facilitate designers and developers in the construction and maintenance of scientific data infrastructures. D-NET implements an open source service-oriented framework where a rich set of services for the collection, processing and provision of metadata and files can be customised and pipe-lined to implement the internal workflows of data management applications. As proven by the several installations (Argentina and Poland national repository infrastructures) and its adoption in a number of European projects beyond OpenAIRE (EFG, ESPAS, EAGLE, HOPE), infrastructures realised with D-NET are easily customisable, extensible, scalable, and sustainable. ⁹ Its customizability and extensibility makes it a suitable candidate for creating aggregative infrastructures mediating between different scientific domains and therefore supporting multidisciplinary research. In terms of numbers, OpenAIRE is growing and currently gives access to over 8.5 million open access documents, 85 thousand of which are linked to EC funding information. It also has over 460 data providers, that is the repositories, both institutional and disciplinary and the publishers who provide the information on publications.

_

⁹ <u>Paolo Manghi</u>, <u>Michele Artini</u>, <u>Claudio Atzori</u>, <u>Alessia Bardi</u>, <u>Andrea Mannocci</u>, <u>Sandro La Bruzzo</u>, <u>Leonardo Candela</u>, <u>Donatella Castelli</u>, <u>Pasquale Pagano</u>, (2014) "The D-NET software toolkit: A framework for the realization, maintenance, and operation of aggregative infrastructures", <u>Program:</u> electronic library and information systems, Vol. 48 Iss: 4, pp.322 - 354

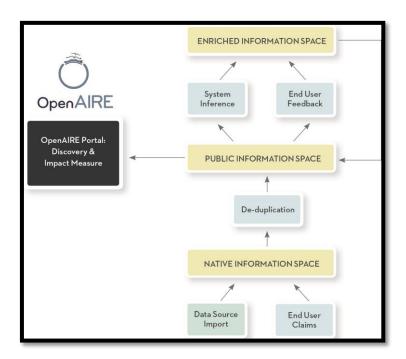


Figure 2 - Image courtesy of NISO / $\rm ISQ^{10}$

 $^{^{10} \; \}underline{\text{http://www.niso.org/publications/isq/2013/v25no3/}}$



4 Interoperability and use of standards

It is essential that OpenAIRE can interact with other data sources and infrastructures. It needs to consume accurate and standardized information. To this end, it has established its 'Guidelines' that data suppliers can use as a low-barrier set of requirements to implement and expose data to OpenAIRE. The *guidelines for literature repositories*¹¹ now available in version 3.0, address the use of controlled vocabularies and formats to facilitate interoperability. Take for example, information about funding bodies and project information; this information needs to be encoded consistently to allow searching across repositories.

OpenAIRE has also started to link together institutional publication repositories with other rich data sources. By extending its guidelines further, OpenAIRE can also harvest from data repositories. To this aim, it has drafted guidelines similar to the literature guidelines. All data resources coming into OpenAIRE have to be registered, validated and collected. The OpenAIRE data guidelines are heavily based on the DataCite metadata scheme¹². A set of low-barrier properties makes interoperability as simple as possible.

The OpenAIRE infrastructure can now handle a range of typologies via its data model. By supporting the CERIF¹³ data model for example, OpenAIRE can harvest institutional and people information from research information management systems, thus giving an extra contextual layer to the research publication. Standards in use in the context of research administration such as CERIF (Common European Research Information Format)¹⁴ are compatible with the metadata model used in OpenAIRE. This enables data exchange at the local and funder level.

The level of interoperability depends on which level of the OpenAIRE guidelines are adopted. Not all the data/repository provider organizations are in the position to choose high level of engagement, as it usually requires the intervention of skilled staff and higher costs of development. A good solution to overcome this problem has turned out to be the implementation and distribution of specific plug-ins dedicated to support the guidelines as part of the distribution of the main software repository tools. This solution, developed and promoted by the University of Minho for DSpace and openly available to the repository community, has indeed simplified and favored the compliance with the OpenAIRE guidelines.

ePSIplatform Topic Report No. 2014 / 10 , October 2014

¹¹ http://www.guidelines.openaire.eu

¹² http://schema.datacite.org/meta/kernel-2.2/index.html

¹³ http://www.eurocris.org/Index.php?page=CERIF-1.5&t=1

¹⁴ http://www.eurocris.org/Index.php?page=CERIFreleases&t=1



Further plug-ins are available for the repository platform EPrints and the journal platform Open Journal Systems (OJS).



5 What are the Services and who uses them?

OpenAIRE continues to create value added services on top of its curated contents.

Portal services: The portal enables easy discovery and access to this additional heterogeneous information, independently of where it is physically maintained. Project information is obtained from authoritative registries directly provided by the funding agencies, like for example, the EC-CORDA database in the case of EU-funded projects. In short, OpenAIRE offers services to project coordinators and research managers by generating project publication lists to embed in their own project reporting lists. These are also incorporated into their CORDIS webpage. As a further aid in simplifying the enrichment process, the researcher can also exploit services that automatically identify these links through a mining process on the full texts of articles. The identified link is proposed and the researcher is then asked to validate it.

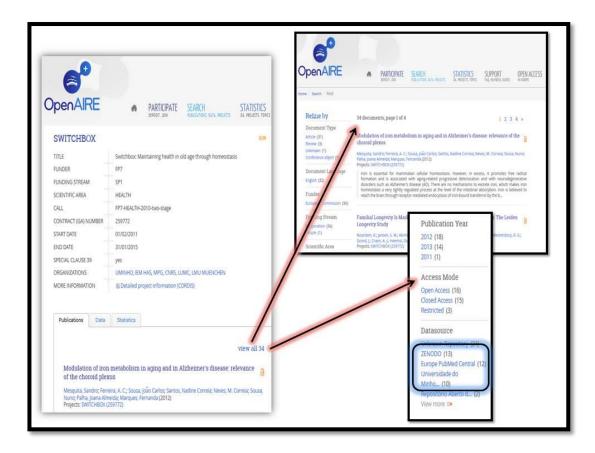


Figure 3: Example project Overview via the OpenAIRE portal



Zenodo: Another of the services is ZENODO¹⁵, a digital repository for EC funded data and publications, hosted by CERN. It offers a valuable level of trust and a high level of service quality and an easy to use workflow.

APIs: Other infrastructures and providers can also consume the OpenAIRE data. Any third-party application provider (including other e-Infrastructures) that agrees to comply with the OpenAIRE terms and conditions of use can build its own service by relying on those made available by the OpenAIRE e-Infrastructure. For application providers, using these services will reduce their own development and maintenance costs, cut down on time to reach their user market, as well as being guaranteed a high quality service.

5.1 Who are the Users?

Researchers: The portal offers researchers services for retrieving and accessing the open access publications of their peers and for disseminating, enriching and monitoring their own ones. For those researchers who are not served by a reference repository they can upload their publications and research data to ZENODO, the OpenAIRE supported repository. Through the portal, researchers can also transform deposited publications into richer information objects equipped with additional contextual information. In particular, through the facilities made available in the portal they can link their papers to the datasets documenting the applied research process and to information on the projects/funding programmes. Researchers can also use OpenAIRE to obtain information on the impact of their open access scientific production, like citation reports as well as usage and impact measures as they are made available to funders.

Data providers are usually institutional and disciplinary repositories managers. OpenAIRE provides them with a channel to disseminate these resources to a wider range of users. It also makes available a variety of services such as the enrichment and cleaning of metadata, the generation of aggregated statistics on the open access usage and notifications about publications in other repositories.

Project coordinators of EC or nationally funded projects can exploit tools that support the management, reporting and monitoring of project outcomes. OpenAIRE creates publication lists with access conditions can be easily created and pasted into the activity reports required

¹⁵ http://www.zenodo.org



by the funding agencies. Usage statistics and information about the performance of these publications can also be easily generated.

Research administrators can obtain detailed usage statistics for funders, at present the European Commission, but this will soon be expanded to other national funders. Based on specific knowledge of the programme funding schemes, statistics can be provided at different levels: programme, repository, and publication. These statistics give a measure of impact based on the usage numbers provided at publication or programme level.



6 Heading to Horizon2020

Building on the results from those first initiatives, and on the growing political support to Open Access, the Commissioners Neelie Kroes (Vice-President of the European Commission and responsible for the Digital Agenda) and Máire Geoghegan-Quinn (Commissioner for Research, Innovation and Science), presented in July 2012, three fundamental documents¹⁶, establishing Open Access as the "default" for the European Research Area, and the new Framework Programme, Horizon 2020. The underlying principle of Open Access to the research results funded in the Horizon 2020 announced in 2012, has already been defined and anchored on the documents and regulations of the new Framework Programme¹⁷,

Open access to all publications: all beneficiaries of H2020 funding are required to deposit, as soon as possible and at the latest on publication, a copy of their peer reviewed journal articles into a repository, and must ensure open access to the deposited publication — via the repository — as soon as possible, but no later than six (or twelve for publications in the social sciences and humanities) months of publication. These conditions have to be made very clear, as well as the relevant workflows, and put into briefing papers to save researchers' time.

The EC's open data pilot: A novelty in Horizon 2020 is the Open Research Data Pilot¹⁸ which aims to improve and maximise access to and re-use of research data generated by projects. It will be monitored with a view to developing the European Commission policy on open research data in future Framework Programmes.

OpenAIRE now has to step up as the infrastructure that will support the stronger mandate for OA to publications: this mandate is a chance for OpenAIRE to support this OA policy via its infrastructure. The OpenAIRE team also has to create a range of resources for stakeholders to make clear workflows and information as to how to meet this mandate.

¹⁶ Communication on a reinforced European Research Area partnership for excellence and growth (http://ec.europa.eu/research/science-society/document library/pdf 06/era-communication-partnership-excellence-growth_en.pdf); Communication Towards better access to scientific information: Boosting the benefits of public investments in research (http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf); Recommendation on access to and preservation of scientific information (http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf);

¹⁷ See Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (http://ec.europa.eu/research/participants/data/ref/h2020/grants manual/hi/oa pilot/h2020-hi-oa-pilot-guide_en.pdf) and Fact sheet: Open Access in Horizon 2020 (https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Open_Access.pdf).

¹⁸ https://www.openaire.eu/oa-in-h2020/h2020/h2020-oa-data-pilot



7 Added value services

The next step is to integrate open access to funding results and make open science the focus of research funding policies. OpenAIRE aims to provide services that ease deposition and publishing workflows as well as cleaned and reliable data. OpenAIRE plans to offer technological support through well-defined APIs and Service level agreements. Activities also include legal frameworks for the provision of advances services to all stakeholders involved (publishers, repositories and data archives) in order to enable content and service provision based on metadata and file processing.

Linking research publications to research data: a large part of the scientific workflow is creating research data. Often this data is not shared or made public, however it would be of great benefit to see the underlying data for a research publication. OpenAIRE is able to support these links, as well as a concerted outreach effort to enable the publishing of data, and making sure that scientists get credit for sharing and having their data cited.



8 The outlook for OpenAIRE

Setting up and maintaining open access e-infrastructures such as OpenAIRE comes with a range of issues such as how to transform projects and prototypes into long-lasting infrastructures with reliable services, government and financial models (Lossau, 2012)¹⁹.

OpenAIRE has been granted funds for 3.5 years to run the project, 'OpenAIRE2020', starting in January 2014. Within this timeframe, OpenAIRE will be set up as a legal entity. As a leading infrastructure provider in Europe, it has to deliver a round the clock service and trusted content, and to demonstrate its commitment in the future it makes sense to have the stability of a business operation via a legal framework with a set of governance measures.

The consortium has been creating a business model in order to pave the way forward in planning its smooth running and gain an income. One place to start has been to measure the benefits such a service brings to society, and add an accountable 'weighting' to the services it can provide. For example, open access cannot be financially measured per se, but can be seen as a common good and it is worth measuring it by non-marketing mechanisms. OpenAIRE has to understand its stakeholders and how willing they are to engage, as well as to project ahead as to how it will be used in the future. It is also looking to streams of revenue to maintain its services. For example, reuse of its content by service contracts for third parties, contribution from member states and the EC to whom such monitoring services are very valuable. The initiative has already collaborated with a range of organisations that provide data and services to others (e.g. EBSO) as well as mining services to publishers (e.g Copernicus) and other infrastructures for research and monitoring services (e.g EGI). Its robust infrastructure has also been rolled out in international instances (e.g. Argentina).

In looking at what exactly should be sustained, it is clear that OpenAIRE is two faceted: it has a network of outreach nodes which promote open access and open science. It also runs a robust technical infrastructure, as well as Zenodo. Together these elements need to be maintained for the infrastructure to meet the growing and changing needs within the scholarly communication landscape.

_

¹⁹ Lossau, N. (2012). An Overview of Research Infrastructures in Europe - and Recommendations to LIBER, *Liber Quarterly*, Vol 21, No. 3/4, <u>URN:NBN:NL:UI:10-1-113632</u>



9 Conclusions and recommendations

Moving forward with open access: It is clear that open access is here to stay: it is growing as a favoured mode to disseminate research outputs at a global level, and this has been recognized by funders across Europe, who are setting the default to open. However there are a number of barriers to make this happen:

- Scientific evaluation criteria is still very much bound to the journal impact factor. Researchers are having to make the choice between publish or perish or adhere to funders' open access mandates. On the other hand, open access repository tools and new research platforms offer researchers new opportunities, such as providing statistics and altmetrics, and collaborative ways of working. Making research available via other platforms other than journals is still to be embraced, but could support this transition to a more open way of doing science.
- Publishers often add confusion to open access conditions, such as conditions for deposit and embargo periods. They should be urged to move to be more adaptable to the current climate, and to serve the real needs of science, by working with more sustainable charging models. In addition, there is a need to distinguish between those publishers whose sole role is to disseminate science and those who are led by huge profits, regardless of whether it is channeled back into producing more services for scholarly communication.
- Within the open access field itself, lie convergences of opinions: the criteria and different
 grades of openness, such as open vs reusable (platinum, diamond and gratis). A consensus
 in the community is needed, but is hard to reach.

At the same time, scholarly communication is changing: entirely new ways of sharing and communication science are being developed. The role of an alternative, more open peer review is also growing among a new generation of scientists and this could provide a boost to the push for open access. Additionally, the research lifecycle is dependent on researchers being able to analyse all research results. Yet many are reluctant to share their data. Efforts to promote data citation and efforts such as the EC's Open Research Data Pilot will help movements in this direction.

Impact for PSI: OpenAIRE is one of many research infrastructures that are being built. While it is cross-disciplinary in its approach, it serves a number of domains, and its interaction with

EPSI platform

OPEN RESEARCH IN EUROPE - OpenAIRE

other initiatives, such as PSI information is critical for the future growth and innovation, made possible by interoperability with other data sources. This will be beneficial to the diffusion and acceptance of the open access culture, as some researchers naturally use infrastructures of their discipline while others rely more on local infrastructures for small scale or interdisciplinary projects and activities.

Lessons learned in establishing an infrastructure: OpenAIRE is working hard to offer technological and coordination support for an open access mandate. As this mandate intensifies, 100% of all peer-reviewed articles during the Horizon2020 funding period, so do the services and infrastructure upgrades of the OpenAIRE infrastructure. Lessons learnt during the process of establishing an infrastructure are as follows:

- The effort into the successful uptake of the infrastructure should not be underestimated.
 Outreach and policy back-up are crucial as well as an understanding of the differences in infrastructure implementation across the member states.
- Build trust by providing good quality services. These services don't have to be fully fledged from the start, and the steps can be incremental.



About the Author

Natalia Manola is the project manager of OpenAIRE. She is based at the University of Athens. A senior software engineer, she has led several EC research and development projects as well as working as a project manager in the information technology fields.

Paolo Manghi is the technical manager of the OpenAIRE infrastructure. He is based at the Institute of Information Science and Technologies at CNR. His expertise lies in digital libraries and data infrastructure for science.

Najla Rettberg is the scientific manager of OpenAIREplus. She is based at the University of Göttingen in the Electronic Publications Unit and has a background in scholarly communication and digital preservation, having worked in a variety of organisations around Europe.

Copyright information

© 2014 European PSI Platform – This document and all material therein has been compiled with great care. However, the author, editor and/or publisher and/or any party within the European PSI Platform or its predecessor projects the ePSIplus Network project or ePSINet consortium cannot be held liable in any way for the consequences of using the content of this document and/or any material referenced therein. This report has been published under the auspices of the European Public Sector information Platform.



The report may be reproduced providing acknowledgement is made to the European Public Sector Information (PSI) Platform.