EUROPEAN UNION

EUROPEAN RESEARCH AREA
AND INNOVATION COMMITTEE

– ERAC –
Secretariat

Brussels, 3 February 2016
(OR. en)

ERAC 1202/16

NOTE

From:	ERAC Secretariat
To:	Delegations
Subject:	ERAC Opinion on Open Research Data

Delegations will find attached the ERAC Opinion on Open Research Data, as adopted by written procedure.

ERAC Opinion on Open Research Data

Executive Summary

This report is a product of the ERAC Task Force on Open Access to Research Data.

The introduction focuses on the terminology, the opportunities and the challenges in open research data and data sharing.

The core of this report consists of 11 recommendations in the field of open research data, including data sharing and the reuse of data. These recommendations cover 4 categories: Training of stakeholders and awareness raising, Data quality and management, Sustainability and funding and Legal issues.

'Training of stakeholders and awareness raising' covers recommendations 1 – 4 which focus on what is needed to be done to change the current culture in research from a "publish or perish" ethos to Open Science and data sharing. Awareness raising, training and education, incentives and reward systems and monitoring are key aspects in these recommendations.

'Data quality and management' (recommendations 5 – 8) focuses on technical aspects of reviewing and assessing data quality, identifying and citing data and management of projects.

Key words for these recommendations are Data Management Plans, metadata standardisation, Digital Object Identifiers, peer-reviewing and quality assurance.

'Sustainability and funding' recommendations (9 and 10) focus on the appropriate infrastructures and funding considerations. Different categories of costs and the FAIR data principle (findable, accessible, interoperable and reusable) are highlighted in these recommendations.

'Legal issues' (recommendation 11) covers intellectual property rights, security issues and applying for industrial property rights, the idea being that open research data should be "as open as possible, as closed as necessary".

The next steps sections discusses the level of urgency and interlinkage of the recommendations, the assessment of supportive systems and the European Open Science Cloud. I also covers the set up of a mechanism for regular exchange, public/private partnerships and consideration of a spot on the horizon.

The report is concluded by a list of references and 7 annexes, providing additional insight into subject.

Preface

The move towards "openness" in science gained momentum over the past few years. This is in part due to digital developments and new possibilities in available in a data-driven economy, increased public engagement and transparency as well as the demand for synergies and the reduction of duplication of research.

Following the Communication of the Commission of 17 July 2012, the European Research Area Committee (ERAC) requested expressions of interest from Member States and associated countries to work together on an assignment regarding Innovation Data given by the end of 2013. The Netherlands accepted to lead this work at the ERAC meeting of February 2014. An informal Task Force convened, involving experts from 12 Member States[1], later broadened to 14 countries[2] with the involvement of representatives of the European Commission. At the end of 2014, the Task Force provided a preliminary report shedding light on the work achieved, detailing potential work-streams on Open Access to research data in relation to the shift towards "Open Science" within the Commission. In December 2014, the ERAC supported the wish of the Task Force to continue its work into 2015 on the topics of the preliminary report.

The main focus of this report is open research data from publicly funded research. The report details 11 optional recommendations to consider when implementing open access to research data policies.

A consultation version of the report was distributed among researchers and experts both in public and private organisations, as well as policy makers, funders, librarians and publishers.

We aim at and hope that the recommendations will help policy makers, experts and researchers at an EU and national level to gain insight into this subject.

[1] Austria, Belgium, Croatia, Denmark, Finland, France, Germany, The Netherlands (lead), Portugal, Spain, Sweden, United Kingdom
[2] Luxemburg, Switzerland

Reading note

When reading the recommendations, please consider that one size does not fit all.

Data is broad and diverse, encompassing different scientific disciplines and methodologies.

This is not, should it be read as, a statement of policy, but rather as a list of recommendations which could help to build national policies and can give insights in different experiences in the field of open research data. We realise the transition to Open Science will only work holistically, involving the research community, government and stakeholders. The recommendations will have to be adapted to fit the needs of the local research system and communities. In addition, the recommendations consist of numerous options, with different stakeholders to be targeted for each. Target stakeholders include policy makers, researchers and funders, working on institutional, national and international level.

Yet due to national and institutional differences, we need to consider that stakeholder groups will function differently in each country. Ultimately Open Science should be implemented by Member States/countries in the best and most appropriate way to suit their national infrastructure and their particular circumstances.

Table of contents

Introduction

Terminology and scope

Open Science encompasses different subjects and terminology[3]. The OECD report "Making Open Science a reality" (OECD, 2015a) offers an explanation of the different terms. Open Science parallels other contemporary research methodologies such as Text- and Datamining, big data, research integrity, open educational resources and open peer-review.

This report deals with Open Access to research data with the focus on publicly funded research. To be consistent with currently used terminology, Open Access to research data will be referred to in this report as open research data. Open research data as such is part of the drive towards Open Science and should lead tot the optimal reuse of research data.

For this report, open research data relates to all collected and created data to support and complete research, and all data underlying an existing or future research publication and all data essential in replicating the results. The Task Force acknowledges that the nature and form of the data varies between disciplines. The recommendations in this report do not apply to sensitive or confidential data, for instance data that could be considered a security risk or is subject to privacy laws and intellectual property rights (including trade secrets). Also special attention has to be given to research that is funded by combined public and private resources and the need to respect the position of all participants involved (see next step 4).

We believe it is essential that openness adheres to ethical principles and respects judicial operating environments; it is for researchers to ensure their data is always legally and contractually cleared before release.

The ERAC Task Force endorses the findings and recommendations provided in the RECODE report. (RECODE, 2014) and we encourage stakeholders to read information published by RECODE.

[3] The various terms include: Open Access (gratis, libre, gold, green, hybrid), Public access, Open Data, Research Data and Open Government data.

Central to this report is a series of recommendations in four key categories: Training of Stakeholders and Awareness Raising, Data Quality and Management, Funding and Sustainability and Legal issues.

The recommendations cover context and detail and, if available, good practice examples.

The report aims to contribute to the progress of open research data and the optimal reuse of research data from publicly funded research. It furthermore aims to shed light on the potential of the policy, while removing obstacles and misunderstandings early in policy development.

The report presents a list of recommendations that consider the national, European and global state of play of open research data. The list is not exhaustive and can be considered as a reflection of current thinking at the time of publication. Nevertheless, we hope it acts as a useful overview of the many initiatives in the field of open research data.

Open Access to publications and open research data

The initiative of Open Access to scientific publications has started earlier and developed faster than open research data. Where more and more Member States are developing policies on Open Access to publications, not all of them are considering open research data (yet). We believe, however, the continued development of Open Access to publications will also create opportunities for open research data. In the digital age, we expect lesser distinction between digital publications and digital access to data. However, open research data is fundamentally different from Open Access to publications. For a publication, the author made a conscious decision to publish in the first place; this is not the case with open research data.

For more information with a schematic overview, see Annex 3.

Data sharing: between opportunities to seize and challenges to take up

Opportunities

Open research data offers many opportunities at scientific, economic and societal levels and will become a fundamental part of the future of Open Science.

1. The immediate or early availability of publicly funded research data, which is accessible for all, could accelerate and improve the creation of new scientific knowledge. Data releases could reduce international duplication of data collation and make datasets more directly reusable. This is also the case for research data giving evidence of non-significant effects or failed tests.

   Striking examples of this are already found in data sharing activities that reacted to the Ebola outbreak such as GenBank (see go.nature.com/aotpbk) and DataONE (enabling federated access to ecological data from different initiatives) projects.

2. Opening research data could – alongside other systems in place - enhance the quality of data through the development of innovative open peer-reviewing practices, annotations, linked open data and protections against fraud.

3. Open research data may unlock an equality of access to essential research information avoiding exclusion of researchers from countries and/or institutions that cannot afford to pay for access to scientific data.

4. Open research data can raise domestic awareness of its national research and provide international visibility, that can aid in showing how investments of research funders can contribute to the production of knowledge, products and services.

5. It is expected that open research data has a positive effect on economic impact and the value of curating and preserving research data. Furthermore, it is likely that there are many forms of social and economic benefits. Capturing them all is not straightforward, although there are reports showing net benefits (Beagrie and Houghton, 2014).

6. Considering publicly funded research data, by default as a public good and making data available with as few restrictions as possible could contribute to the social and economic benefits of spreading knowledge more effectively.

7. Policy makers could consider also using open research data to evidence-base processes of political decision making.

8. Although data sharing is far from common practice, some disciplines such as astronomy and oceanography already have a long tradition of sharing data. The use of open research data and data sharing varies between disciplines. These examples be reviewed as examples for showing the benefits of open research data to others.

9. Modern technology and contemporary research habits have anticipated the move towards open research data, research data in digital form being increasingly (re)used in research.

Challenges

There are many opportunities for open research data, the reuse of data and data sharing. However it brings with it specific challenges that need stakeholders consideration to ensure researchers, policy makers, industrial partners and citizens can benefit from the aforementioned opportunities.

1. One of the main concerns towards open research data is the fear of losing competitive advantages. Therefore industries are often reluctant to share proprietary information (or perceived as trade secrets). This can also be because the benefits of open research data are not clear to (or supported by) them.
   Citizens may have strong privacy concerns (Manyika et al., 2014) relating to some specific fields of research (such as biomedical and social sciences).

2. Researchers tend to be reluctant to share data if they do not receive due credits. Some are therefore ambivalent towards new ways of dissemination (Open Access) and assessment (Open Peer-reviewing) of research outputs (Lefèbvre, 2006, 2008, and 2010; Creaser et al., 2011; Housewright et al., 2013). This may contirbute as to why Open Access is not widely considered as a quality benchmark when selecting journals for publication (Davis et Walters, 2011). Nevertheless, attitudes towards Open Access vary across the disciplines (Creaser et al., 2011) and the perception of Open Access tends to become more positive as it develops over time (Frass et al., 2014).

3. Generally speaking, several disciplines historically lack a data sharing culture, with many researchers not used to sharing and disseminating publications outside of a research team (Candela et al., 2014; Van den Eynden and Bishop, 2014). There are a number of reasons for this:

   - Some researchers consider data a private affair, closely linked to particular methodologies and theoretical frameworks, and hard to reuse in other research contexts;
   - Some fear a reanalysis of datasets could reveal errors or produce alternative conclusions (Wicherts et al., 2011);
   - Data management may be perceived as a particularly time and money consuming, distracting from R&D activities;
   - As datasets are not generally considered truly assessable research outputs, and due to the competitive "publish or perish" context in which most current academic research is produced and assessed, researchers tend to be "possessive" of research data they collect. They may aim at capitalising on it later in order to publish (or co-publish) as many papers as possible on the basis of the data, so that it is not "scooped" by others (McKiernan, 2014), which can and does happen;
   - Research data is considered an asset with a competitive advantage and therefore key to collaboration, co-authorships and in attracting students and researchers to work for institutions;
   - Collaboration with private partners can include non-disclosure agreements, preventing researchers from sharing data;
   - The lack of evidence in whether free and worldwide data sharing has benefits for a given country or organisation funding the research.

4. There is currently no quality assessment and peer-reviewing of data in a comparable way to that existing for (peer-reviewed) publications, since it is not common practice (yet) to publish the data.

5. There is currently a lack of attention for open research data policies and infrastructures at an institutional level, compared with Open Access to publications (Archambault et al., 2014).

6. There is no generally accepted business model covering the costs relating to open research data curation, open research data infrastructures and related training.

7. The acceleration of technology, and particularly of software, has led to numerous publication formats of datasets; many are difficult to access and, at worst, require the user to purchase additional software for access and reuse.

8. There are fears of the free-rider effect; released data could be farmed by countries or organisations with no interest in Open Science for their own gains.

The ERAC Task Force on Open Access to Research Data – recommendations

TRAINING OF STAKEHOLDERS AND AWARENESS RAISING

1. Promote a better understanding of open research data through communication & awareness raising

Communication and awareness raising activities should be organised to promote open research data as broadly as is the case with Open Access to publications.

Within awareness raising and communication, the focus should consider the following:

- Communicate to researchers that open research data improves the accessibility, visibility and reusability of data:
  - Accessing an online open research data set is easier with possibly lower transactional costs, than personally asking an author for his/her data (Savage and Vickers, 2009);
  - Higher visibility and discoverability of datasets (Piwowar, 2013);
  - Higher citation rates to related scientific publications (Piwowar, Day and Fridsma, 2007; Piwowar and Vision, 2013);
  - Improved access for small labs and (developing) countries with reduced research funding to international datasets (Arzberger et al., 2004);
- Communicate to researchers that open research data can lead to improvements in data quality:
  - Comprehensive evaluation of a research project requires access to the underlying data. The data would be considered an integral part of the methodology and peer-reviewers should able to easily access them;
  - Through the possibility of reproducing and reusing the data (also relating to research integrity);

  - Better quality assurance of the data via innovative Open Peer-Reviewing which potentially opens the process of peer-review to the entire scientific community;
  - Subsequent strengthening of the credibility of scholarly publications, researchers and research institutions following recent media focus on scientific fraud (at the expense of public confidence in science) (Grieneisen and Zhang, 2012);

- Promote awareness that scientific data which is funded by the taxpayers can be seen as a publicly owned good;
- Communicate to researchers that public availability of research data can improve public understanding of science, evidence-based policy making and citizen-science initiatives (Kowalczyk and Shankar, 2010);
- Promote proper data management as a method of allowing the long term preservation of the data and encourage its integration into funding policies;
- Promote assets linked to the Open Innovation paradigm to the private sector (Spiegler, 2007; Houghton and Sheehan, 2009; Houghton, Rasmussen and Sheehan, 2010; Houghton, Swan and Brown, 2011; CEBR, 2012; Royal Society, 2012; Beagrie and Houghton, 2014; OECD, 2015a). Specifically, that Open Science and open research data can allow even SMEs without (large) R&D departments to access cutting-edge scientific information at no cost;
- Raise awareness that when working with open research data, as a default principle, it does not mean all of the data must be published. Access to data could be restricted, embargoed or even closed. In such cases, the reasons why the data cannot be open should be made clear and be consistent. This can be due to privacy issues (such as in medical and biological research), confidentiality and, where relevant, national security, IPR or industrial interests. We recommend this to be summarised as: “as open as possible, as closed as necessary”.
- We encourage research funding and delivery organisations to disseminate widely and to reward – e.g. through awarding or funding - all types of good practices behaviours where possible (see also RECODE, 2014).

Good practice example

The Lodlam Open Data prize is an award for Linked Open Data in the field of Library Archives and Museums which is given at the LODLAM Summit in Sydney, Australia (http://summit2015.lodlam.net/challenge/). In 2015 the prize was awarded to researchers from Centrum Wiskunde & Informatica (CWI), VU University (VU), Erasmus University Rotterdam (EUR), Delft University of Technology (TU Delft) and the Netherlands Institute for Sound and Vision, for the Polimedia research project. Polimedia aims to stimulate and facilitate large scale cross media analysis of political events and is beneficial for researchers in communication science, politics, history and journalism.

2. Establish training and education programs on Open Science

Studies suggest a shortage of data-related skills in the training researchers currently receive (OECD, 2014; OECD, 2015b). Open Science, including open research data should form an integral part of researcher, student and stakeholder training.

We recommend:

- Focusing in particular - but not exclusively - on Open Science training and education programmes that target young researchers and (masters)students in teaching them to recognise the complexity of data and the value of good data stewardship. As they are still developing their research habits this can be instilled early in their doctoral training;

- Providing dedicated training for librarians and data managers/stewards, in coordination with professional data centers (RECODE, 2014);

- Taking into account the diverse needs of relevant stakeholders when developing training and education programs (RECODE, 2014). Training and education activities, both informal (e.g. workshops) and formal (curriculum development in data management skills and data-intensive research) should be developed (RECODE, 2014);

- Including Open Science/Open Data modules within European doctoral programmes and/or summer schools;

- Introducing properly trained data-officers to scientific research project teams;
- Open research data training would be advantageous and can be given in multiple ways, including modules on how to manage and annotate data in a way that will allow sharing and reusing of the data;
- Considering dedicated funding needed for training and education programs in Open Science.

Good practice example

The current FOSTER project (Facilitate Open Science Training for European Research) (https://www.fosteropenscience.eu/), funded under the Seventh Framework Programme for research and technological development, offers training resources for Open Science practices. FOSTER hosts online and face-to-face training packages that help researchers comply with Open Access policies in line with the recommendations and expectations of the Horizon 2020 framework. While face-to-face packages adopt a sustainable 'train the trainer' approach, a dedicated FOSTER Portal has also been created to support e-learning, blended learning, self-learning, dissemination of training materials/contents and Helpdesk.

3. Establish a reward system for data sharing activities

Considering researchers are mainly credited and appreciated for having their published papers appear in international "top journals", stronger career incentives should be established to reward data sharing and open research data habits.

- Research councils and foundations could provide incentives or mandates for data sharing, taking Open Data as the default requirement in the dissemination of research data (OECD, 2015a). Governments should also provide ad hoc enablers regarding technological infrastructures and training;
- Open research data has the potential to promote the reputation of creators or data collectors as well as funders, publishers and third parties, provided the data can be cited by users in a suitable manner;

- Data dissemination should be considered when evaluating researcher performance: “Formal acknowledgement of research data as a legitimate output is expected to bring gradual change in practices. Such formal recognition should be accompanied by the development and use of metrics that allow the collection and tracking of data use and impact” (RECODE, 2014);

- An “Open Science” dimension may be integrated as one of the criteria into institutional rankings of universities – and particularly into the multidimensional ones, such as the EU-funded U-Multirank - with an indicator that can be based on open research data;

- Data Management and stewardship activities could be considered an eligible cost;

- Where applicable, we encourage taking the particular interests of individual researchers participating in a research project into account (Roche et al., 2014) – including their “right of first use” of the data, alongside the customs of different disciplines. More flexibility may be allowed for researchers to extend deadlines to open up datasets that underlie the research project, and which could be made open only later, e.g. by the end of the project or after a reasonable period of monetarisation. It can be argued that not all researchers can be expected to systematically release their whole set of data in open access in the same way as it happens with publications (Harnad, 2012). While in case of Open Access to publications, the researchers have already decided to publish the given content, in the case of open research data it can be that a specific part of the data has not even been worked with yet. Therefore they may be given sufficient time to analyse “their” data first, having invested effort into collation/generation;

- Appropriate author attribution of datasets and data citation standardisation will help to fight against scooping (see recommendation 7).

On an EU-level assistance and advise can be given to funding organisations and other key players in the drafting of research data policies and (model) DMP’s and resources can be provided for the establishment of transnational frameworks.

Good practice example

In the UK, the Research Councils UK and the Wellcome Trust have developed policies for research data and supporting services, though they constitute an exception for Europe. Specifically, the Wellcome Trust requires "applicants whose proposed research will generate data that hold significant value as a resource for the wider research community to submit a data management and sharing plan as part of the application process" (http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Data-management-and-sharing/index.htm).

4. Ensure sound monitoring

The implementation and impact of open research data needs appropriate monitoring.

Sound and possibly innovative indicators of open research data should be developed. They would encourage evidence based policy making, facilitate evaluations and provide quantitative information on the assets of open research data for all parties involved.

Technical barriers restricting data citation and related metrics should be properly addressed, for instance: software and hardware incompatibility; data file structures; data storage and management (Groves, 2010). This will enable the monitoring of researcher compliance against funding and national policies.

Good practice example

An Open Access Barometer is being developed in Denmark. The barometer will measure the extent to which Open Access targets, outlined in the Danish National Strategy of Open Access to scientific articles, are met annually. The first measurement will be made available in early 2016 from the Danish National Research Database (http://www.forskningsdatabase.dk).

DATA QUALITY AND MANAGEMENT

As yet, there is no standard data quality assessment protocol similar to that which exists for scientific publications (Brase et al., 2009). However to be (re-)usable by scientists and other stakeholders alike, data must meet a minimum quality criteria, on a technical and scientific/content related level (Dallmeier-Tiessen et al., 2011; Royal Society, 2012).

5. Make data identifiable and citable

Data identification is necessary to ensure it is citable and reusable. One particularly efficient way to make datasets identifiable involves the use of persistent, unique and interoperable types of identifiers, such as the Digital Object Identifiers (DOIs). Another – but much less stable and persistent – way is through referencing the URL of the dataset. We recommend publishers, research funders and research performing organisations consider mandating the use of identifiers such as DOIs as a standard practice in the new Open Science environment.

Data citation refers to the practice of providing a reference to the research data in the same standardised way researchers provide bibliographic reference to published research results, allowing authors to be credited and recognised for their work (Mooney and Newton, 2012; CODATA-ICSTI, 2013). Appropriate data citation needs to ensure data can be recognised as a primary research output and be duly attributed to the researcher. This would enable the development of citation related metrics.

Researchers should follow agreed data citation guidelines, listing citations to data sets in the relevant reference/bibliography in their publications (Kotarski et al., 2012). Publishers, research funding and performing organisations alike should be encouraged to provide clear guidelines for data citation and support efforts made by dedicated organisations aiming to improve data citation and providing recommendations regarding data citations - for instance Datacite which recommends the following standards for the citation of datasets:

- Creator (PublicationYear): Title. Publisher. Identifier (DOI or URL);
- Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier (DOI or URL).

Data citation may also include licensing information (e.g. Creative Commons licenses), preferably machine readable, which clarifies the flexibility of the use of the data. (RECODE, 2014).

Good practice example

The Horizon 2020 Guidance recommends that identifiers are "persistent, non-proprietary, open and interoperable" (European Commission, 2013).

6. Promote metadata standardisation and production of metadata

Metadata is data that describes other data. It is information detailing the context, content and structure of a dataset along with information about the management and processes needed throughout a dataset's lifecycle. Quality metadata is essential for the discovery and use of the data. Discipline specific vocabularies and ontologies are used to help make quality metadata. Research data and public sector information (PSI) are similar in the sense that both require commonly agreed international standards to enable capitalisation of their full potential. To promote cross-border and cross-sector data exchange and on-line availability and usability, interoperability should be promoted by metadata standardisation, allowing an easier exchange of information. Ideally a standardised model should include:

- Legal metadata. A common understanding of concepts and vocabularies describing the legal state of the data set;
- Semantic metadata. Precise meaning of exchanged information which is preserved and understood by all parties;
- Technical aspects. Building interoperability for different metadata for interfacing and linking services for the preservation of metadata, such as checksum. A standardised definition of a (comprehensive) metadata scheme for Quantities, Dimensions, Units and Uncertainty will improve research data management: each scientific discipline or community manages data according to its own habits while apparently not looking so much at data interoperability across disciplines and decades to come. Misinterpretation of numerical data can be the consequence.

In practice, enabling cross-portal search for datasets and making research data available across borders and sectors through descriptions of datasets in data portals, is highly valued. Technical and operational efforts to develop and operate a central portal sharing interoperable assets (i.e. standards) are essential. Standards for data archiving do exist and should be applied wherever possible to help researchers. The EU has proactively supported a standard creation for data. The Data Catalogue Vocabulary Application Profile (DCAT-AP), for example, enables aggregation of and search for datasets across data portals.

Good practice example

Aiming at the higher end of the interoperability of research data and better use of research resources, the DG CONNECT/DCAT-AP has a possible standard for a data vocabulary for research data catalogues. Research practices and special operational environments, however, set requirements and conditions for metadata management and modelling.

7. Promote innovative models for (open) peer-review and processes of quality assurance

The process of opening datasets encourages researchers to assess the quality of their data. Sharing data beyond the research team stimulates researchers to ensure the data is accurate and properly documented.

The quality of open research data should be assessed via the development of innovative peer-review processes and the development of specific quality assurance procedures and tools.

Innovative data (open) peer-reviewing processes

The assessment of the quality of research papers – as well as, to a lesser extent, monographs - has been fully integrated into the work routine of researchers through the process of peer-reviewing. There is no such standardised reviewing procedure concerning data. We encourage the development of innovative peer-review processes for research data, at pre- and post-publication stages (RECODE, 2014; Archambault et al. 2014: 9).

Current ways of providing peer-review to data include: the peer-review of data in data journals, peer-reviewing the soundness of the data collation through traditional peer-reviewing of articles (the editor opening access for the peer-reviewers to the data sets on which the article is built), the extension of the innovative "open peer-reviewing" practices developed by several Open Access journal publishers to the assessment of data, and crowd sourcing of the assessment of the quality of the data by the mere fact of sharing them.

This does not necessarily mean open peer-review has to be connected to open research data. It can be that another method is chosen for peer-reviewing or that open peer-review is used for other reviewing processes than that of open research data.

Quality Assurance

Specific quality assurance procedures for data repositories (including accreditation systems) and data centers should be developed (RECODE, 2014).

Accompanying dedicated tools

Accompanying dedicated tools, such as re3data.org, the global registry of research data repositories, could be developed and supported (also financially) allowing researchers and other stakeholders to easily check the quality of open research data repositories. On the other hand, generic Open Access tools such as ROAR (Registry of Open Access Repositories) and OpenDOAR (Directory of Open Access Repositories) could index significantly more open research data repositories than it currently does (Archambault et al., 2014: 9).

Good practice example

"re3data.org" is a global registry of research data repositories covering different academic disciplines. It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. re3data.org promotes a culture of sharing, increased access and better visibility of research data. The registry went live in autumn 2012 and is funded by the German Research Foundation (DFG)" (http://service.re3data.org/about).

8. Strongly promote the use of data management plans

In research projects, planning how data will be managed throughout of a project and beyond, is vital. A Data Management Plan (DMP) is an important tool to handle research data, including, but not limited to, provisions on which data is wholly or partly Open Access. It is also a helpful tool in building an up-to-date inventory of datasets.

A DMP can design, put into practice, and follow up on how research data is collected, organised, used, stored and preserved to achieve the highest quality and long-term sustainability. It can have its focus on both the current and the future use of the data.

Planning helps focus resources and funding to where it is really needed and will clarify individual and institutional roles, such as data ownership. A DMP identifies the specific factors that can promote, limit or prohibit data sharing before any data has been collected. DMP's can take differences between disciplines into account and help a negotiation process between the concerned stakeholders. Potential limitations to data sharing include IPR-issues, the presence of sensitive information in data, proprietary file formats, or poorly documented data (Corti et al., 2014).

A DMP should address the following topics, though each funder/university may have additional and specific requirements (Corti et al., 2014):

- What data will be created during the research project?
- Which policies apply to the data, such as legal, institutional and funding requirements?
- What data standards are used, including metadata standards?
- How will the data be documented and identified (European Commission, 2013)?
- Ownership, copyright and intellectual property rights in data;
- Data security aspects;
- Data storage and back up measures and required equipment or infrastructure;

- Plans for sharing data; Who will be able to access and reuse? Are there embargoes or restrictions?
- Data management roles and responsibilities;
- What costs and extensive resources needed over and above usual research will be used?
- Usability – the presentation (format) and distribution of data.

A DMP should be completed early in a research project (ideally on application). This will make the research process easier (Jones et al., 2013), though adjustments to the DMP during the project, on funding, are perfectly acceptable.

In the USA, Australia and the Netherlands, funders often demand a DMP as an integral component of research grant applications (Væring Andersen et al., 2014).

However, in a recent study of 19 participating Research Funding Organizations (RFO) across countries in Europe (Knowledge Exchange, 2016 (*in prep*)), only five of the RFOs stated that their organisation requires a DMP. We therefore encourage more European funders to require a DMP in order to promote awareness of good data management practices, including an increase in open research data.

In the UK and the USA, IT tools have developed to guide researchers through the different requirements in the data management plan. DMPonline (https://dmponline.dcc.ac.uk), Digital Curation Centre, Scotland and DMP tool (https://dmp.cdlib.org), University of California (USA), are among the tools that help researchers write DMPs that meet funder requirements. DMPonline is a flexible tool, organised by target funder and it can be structured according to whichever stage in the tender has developed (Jones, 2013).

Implementing IT tools, such as DMPonline may possibly promote data management planning among European funders and researchers. However, online tools may not be sufficient on their own and should be combined with other means, for instance training.

Examples of concrete tools for training in DMP are MANTRA and RDMRose. MANTRA's target audience are researchers and it is an online research data management training course with quizzes, videos and software tutorials. RDMRose's target audience are librarians.

Good practice example

In the Open Data pilot in Horizon 2020, participating projects are required to develop a DMP in which they specify what data will be open (Guidelines on data management in Horizon 2020). This DMP does not have to be created at the application stage but within the first six months after the project has started. It should be updated during the duration of the project. Not all areas of Horizon 2020 are participating in this pilot and applicants have the possibility to opt-in or opt-out.

SUSTAINABILITY AND FUNDING

9. Ensure the existence of FAIR (findable, accessible, interoperable and reusable) open research data infrastructures

A cornerstone for the dissemination of research data is securing long-term preservation, including metadata, to ensure interoperability (having standards for storage, analysis and access to data), and to set up rules for access and reuse of the data, encompassing authentication and modes of access. These characteristics are summarised as FAIR: Findable, Accessible, Interoperable and Reusable.

This will require a technical infrastructure and tools for dissemination, but also expertise and support from competence centres. These infrastructures could have a federated structure, which may avoid inefficiencies and the inadequate proliferation of data. The emphasis and priorities might differ by discipline, as some research domains experience annual doubling rates in data growth, sometimes privacy and data protection issues are prevalent, and other challenges may include enormous amounts of data at experiments produced instantaneously.

There will also be an increasing demand for advanced user support (e-Science experts, data librarians: the human e-infrastructure) within nearly all scientific fields. Needs are diverse and comprise of the whole lifespan of the research data and will need to include new factors from existing user communities.

In addition, new communications and procedures within the research and publication processes may evolve, such as the creation of Open Access Data Journals, or the development of Open Data policies for journals.

Hence initiatives should focus on the creation of data platforms, as well as on their support. We need policy coordination with respect to these infrastructural and organisational activities.

The responsibility of ensuring an adequate and sustainable data infrastructure varies, depending on the highly fragmented European and global research data landscape. Hence the challenge could be best summarised as:

– How can adequate and sustainable infrastructures and dissemination tools for open research data be ensured;

– Who will be responsible for setting up local and national infrastructures, and how to organise coordination on a European or global scale?

Due to the variation among disciplines, institutional and national contexts, it is important to have to have a bottom-up approach. An efficient way of addressing this challenge is to set up a system for Data Management Plans to create awareness among researchers, to introduce standardised protocols (by disciplines), to have efficient processes (data seal) for submitting and archiving data and to set up protocols for reuse of data (including privacy and protection of sensitive data). It is also important to have adequate cost estimates and to ensure these costs are eligible for funding. This includes the costs of long term preservation.

If research data infrastructures are shared over national boundaries, responsibilities, coordination and costs could be shared between respective Member States and handled through statutes and/or membership fees. This implies likewise identifying and reducing barriers to inter-institutional, inter-disciplinary and international collaboration among research institutions, industry and citizen groups.

A diversity of infrastructures

Currently a diverse range of dissemination tools and infrastructures are available to researchers wishing to make their research data open access: Open Access data journals, collaborative Web tools for Scientists, online networking and bibliographic management tools, research data repositories, institutional repositories, personal and independent websites as well as cloud storage infrastructures[4]. In annex 4 an overview of these different tools and infrastructures can be found.

Priority to trusted repositories

Given this huge diversity of dissemination tools and infrastructures, we recommend that researchers deposit their data in a recognised, fully interoperable, open and "trusted" repository. This can be in an institutional repository, a (multi)disciplinary repository or any other type of sustainable data infrastructure, depending on the disciplinary practices, the national or the institutional mandates.

In a non-restrictive manner, data in trusted repositories should be: supported within their scientific community, findable, sustainable, qualified to provide professional data curation as well as stable identifiers for all deposited datasets, interoperable with other data infrastructures and, most importantly, free of charge and publicly accessible without any unnecessary restriction, allowing proper re-usability of the data.

We also encourage the publication of "data papers" in "data journals", themselves preferably in Open Access, in a complementary manner to the deposit of the data in trusted Open Access repositories, in order to foster the quality and the re-usability of the datasets. In most cases the data has to be deposited in an Open Access repository[5]. This is, for example, the case with the data journal *Scientific Data*, published by Nature since 2014 (http://www.nature.com/sdata/about), which peer-reviews "data descriptor" articles in the STEM disciplines, with details about the methodology, arguments about the validity of the data and advice on how to use them. An Open Access publisher such as Ubiquity Press also mandates that all of its data journals (e.g. the *Journal of Open Archaeology* Data: http://openarchaeologydata.metajnl.com) link the published data papers to openly archived data, according to appropriate open standards.

[4] Each of these dissemination tools and infrastructures present some benefits as well as some drawbacks which we discuss in Annex 5, with regard to their discoverability, accessibility, usability, sustainability and the quality assurance they provide.

[5] It should be avoided in any case that a data journal becomes the owner of the data on which the published data papers are based.

Thought can be given, on top of the repositories, to a service layer like OpenAIRE (https://www.openaire.eu), which allows funders to monitor the compliance to mandates and disseminate the data. This could link data with other information being registered in CRIS (Current Research Information Systems) or provide API's (Application Programming Interfaces) and LOD (Linked Open Data) services.

Good practice example

In consultation with large data producers and managers, the Dutch data archive DANS documented 16 guidelines that determine whether a data repository qualifies for the Data Seal of Approval. The Data Seal of Approval may be of interest to research institutions and organisations that archive data and to users of that data. It can be granted to any compliant repository that applies for it via the assessment procedure. (http://datasealofapproval.org/en).

10. Ensure sufficient funding for open research data and for data sharing activities

Context

The costs for open research data and data sharing activities need to be analysed for a number of reasons. Annex 6 gives a schematic overview of the elements involved in this complex field.

The cost structure of open research data is complex

Research publications are mainly produced and prepared for dissemination according to long adopted standard types and formats. Metadata, infrastructures and procedures for storing, preserving and providing research publications have standardised over the years. Costs of Open Access to publications, even if far from resolved, have been more identifiable and somehow measured, unlike costs for open research data.

Research data encompasses a greater diversity of formats, comprising as raw basic material not only text and figures but also images, audio files, organic and inorganic materials and software code, among other formats. Likewise, data handling procedures, metadata and even infrastructures can assume an unparalleled variety in the research publication landscape. Adding up to the heterogeneity of research data landscape, disciplinary singularities can ensure the same category of

data type is handled very differently, making it difficult to reap the potential benefits from data cross-fertilisation. Also, the nature of cutting edge research causes standard-setting agreements difficult to endure, since the discovery of new data types and data-related procedures or techniques are irregularly evolving, and at an increasingly fast pace. These new procedures and techniques could require building and installing new types of data infrastructures.

Research publications are the product of an entire system for disseminating research results, used much before Open Access became a hot topic. In contrast, research data typically has not been shared or disclosed unless necessary to support the findings of a publication.

The complexity of costs makes it difficult to determine a realistic budget

A first step to help determine costs related to the sharing and management of research data is to break down the activities, procedures, services and infrastructures engaged and try to attribute the costs associated with each. Four cost categories are described below.

Next to getting a breakdown of the costs in general, it is also useful to know which habits from the current practice of keeping data exclusively within the research team can be diverted to a new system of Open Access to data.

1. Overarching costs

Managing research data for sharing and reuse, requires a set of highly time-consuming activities and also depends on a professionally trained workforce, endowed with the necessary level of skills. Time factors and those pertaining to the high level of specialisation of human resources, such as training and capacity and skills-building, represent overarching high costs present throughout all stages of data management and sharing.

2. Infrastructural costs

Regarding infrastructures, these include all data hardware, but also the required software. Costs depend on the options chosen for the technical infrastructure for data-management (hardware) and software (specifically designed by the research team for the project, supplied by third-parties by acquiring them as goods or having them outsourced as services, or a combination). It also depends on whether the choice is to set up a new infrastructure or to adjust existing data infrastructures.

A significant part of costs with data infrastructures is associated with setting up, maintaining or adjusting infrastructures for storage, long-term preservation, curation and further reusability of research data, such as data archives or repositories. Since the time required for those infrastructures to deliver such services can be long, perhaps even decades, a particular challenge is to designate a responsible owner of the infrastructure, establish a sustainable governance framework effectively engaging the relevant part of the research community and other stakeholders and to ensure an adequate and sustainable funding stream so as not to jeopardise the enduring efforts.

Alongside this, often data has safety, security and privacy requirements. The specific data infrastructures used to store, curate and make available, upon certain conditions, data with special privacy or security constraints, such as personal data (for example medical and financial records) and national security-related data, among others, are characterised by the need to implement tight security measures, often mandated by law. This data require specially designed infrastructures and handling procedures, making access to them more difficult and the costs for storage and handling are higher. Examples of such infrastructures include data safe havens and clearinghouses.

Lastly, the need to achieve interoperability of repositories in order to access each others data, asks for specific requirements and has more costs involved. A data archive should no longer be seen as an island but it should be integrated on a repository network standing on robust common principles and guidelines in order for a minimum level of service to be provided by the network.

3. Handling costs

Different types of costs are related to the activities required for handling and managing the data. These include the costs of collecting data, quality-assurance costs, costs driven by selecting which data should or should not be kept for further reuse, archiving, curation, description and costs due to reusability procedures. These reusability procedures can include converting data produced in proprietary formats to interoperable formats or migrating data archived in formats bound to become obsolete to other formats, or to think about how to sustain the (outdated) software, in order to guarantee data usability in the long run.

Due to their fundamental importance in data management processes, two of these activities are worthy of additional scrutiny, namely quality assurance of data and selection of data for long-term preservation. The quality of decisions on these activities impacts on the costs or savings related to the accessibility of research data. One of the difficulties is to decide which data might still be of use in the future, when techniques (e.g. for analysing) will evolve further. These functions are tightly connected to the management of research data infrastructures and the people responsible for them are sometimes referred as *gatekeepers, custodians or stewards.*

Another category of handling costs is connected to the costs resulting from the application of procedures and techniques destined to ensure that personal data is duly anonymised or pseudonymised, in parallel to the specific costs associated with safe haven data infrastructures.

4. Legal costs

A further source of potential costs for open research data arises from legal requirements related to research data.

One type of legal cost is relates to data itself: its ownership. Certain research cannot produce or collect data freely on their own and rely privately owned or curated data, works protected by copyright or, even when the data belongs to public institutions, the access or reuse of data invokes paying a fee of varying amounts.

Even when data itself or its access has no charge, a different type of legal costs can be described as administrative costs arising from clearance and authorisation procedures, for instance, when a favourable opinion from a Data Protection Authority or an Ethics Committee should be delivered for data to be accessed, as can often be the case for research in medical or social sciences.

Another case occurs when research involves cross-border collaboration and exchange of data. Different legal frameworks regulating the handling of data may apply on a research team and result in unforeseen costs so that requirements of different jurisdictions can be fully complied with so that the data can be freely exchanged.

This set of legal activities go hand in hand with the need for researchers to resort to specialised legal assistance, and this need too involves extra costs.

Who will pay?

From the description above, it is clear costs are involved in making fully open research data a reality. This immediately raises the question of who will be the paying party: who holds responsibility and who will make the most use of this data? Will this be European or nationally organised, on a governmental level, through a funding organisation or through higher educational institute libraries?

This also raises the question as to whether the data should be made available freely of costs or if it is legitimate/justified to charge end-users for access to data, though charging for open research data certainly limits their use and potential benefit and involves a contradictory component that may prove to be self-defeating?

An extra complication in this aspect is that assurance has to be given for sustainable/long term funding, to ensure the necessary activities and infrastructures for selecting and preserving the potentially useful data is available.

We as a Task Force feel that, although it is debatable whether costs for depositing, (long term) preservation, value adding or other types of actions to make the data (better) reusable are justified, costs for access in itself does not fit the principle of open research data at this moment.

Funding opportunities for researchers

The necessity to democratise knowledge means researchers must be sure costs will not be an obstacle or impediment to access data.

Furthermore, it should be assessed whether costs involved in realising open research data could be eligible in different funding schemes/for different funding organisations. This can refer to costs for setting up a DMP, for storing/depositing data, for being able to reuse the data and for assuring long term preservation of the data.

Conclusion

While it is possible that for some disciplines a whole new system – from standardisation of data types and formats, to handling procedures and infrastructures constructed to disseminating research data for further reuse - must be developed by scientific communities, it could be that for some disciplines adjusting of the existing repositories/infrastructures is sufficient.

The basic requirements on metadata and data tagging vary widely across sub-disciplines and assurance of potential usefulness of data requires involvement of researchers and other specific stakeholders in defining such requirements. So potentially new financial means are necessary for setting the system up as well as to keep it running.

Sources of potential costs for open research data are wide ranging. Selecting and applying methodologies to determine the cost structure of research data should be one of the first and primary tasks of data managers, although not exclusively for them, so that sustainable funding arrangements can be catered for. It will be necessary to think about the eligibility of different categories of costs within (existing) funding schemes.

Supporting research data must not be looked at only from the perspective of costs since significant overall savings - considering the greater research context - could be achieved due to long-term preservation and preparation of data for further reusability.

LEGAL ISSUES

11. Make IPR issues insightful

Protection and licensing as a basis for open research data

The preferred form of intellectual protection for open research data, which encompasses sharing rights, must consider the interests of different stakeholders and funding bodies. In cases where commercial suppliers provide data as part of a research project, flexibility could be considered between funding bodies regarding the terms of the license. Making research data available immediately could threaten the involvement of private partners in research and therefore may not benefit the public.

A balanced approach in either opening up or protecting data is recommended: “as open as possible, as closed as necessary”.

Current state of protection

One of the most common ways to protect information and data is by keeping the data secret on a temporary or permanent basis. In any research project you have to consider what data/information should be kept secret an what will be open to the public (by means of open sharing and/or Intellectual Property Rights protection including licensing). Generally individual datasets do not hold copyright protection, provided the individual elements are not in themselves under copyright.

Depending on the national copyright system, research data may be protected under the copyright law, because of some original/distinctive elements. It could be that in most cases this will not apply to research data, where the data is derived from sources in a given form and merely assessed rather than sorted and compiled.

Where countries have implemented the European Database Directive, database protection exists if the data is systematically or methodically arranged into a database consisting of pieces that can be individually retrieved, with the compilation requiring appropriate investment. The legal or natural person which has made the substantial investment is seen as the database producer and enjoys the exclusive rights to distribution, reproduction and granting access to the database.

The protection for research data under Intellectual Property Right (IPR) rules is in general sufficient to protect against free-riding and exploitation. It is to be noted that levels of protections vary heavily between the Member States.

Registering industrial property rights

Open research data and the registering of industrial property rights/a patent should not exclude each other. Therefore whenever a researcher wants to register for an industrial property right on certain research data, he should be given sufficient time in order to be able to do so. While information that has to be kept secret, for any kind of reason, could also prevent the application of an intellectual property right. These aspects need to be further analysed and elaborated.

Keep datasets trustworthy

Legal certainty contributes to the quality and rapidity of due diligence research, decisions on the registering of patents and the setting up of business cases.

It is therefore important to know when/at which date the data has been released, to determine e.g. what the status of the technical development was at that time. To have and maintain trustworthy data, it has to be clear what the current state of the data is and that these data are not subject to changes afterwards. And if changes to the dataset have been made over time, this should be explicitly indicated.

Enable the use of licensing by creating a level playing field in the Digital Single Market

Open licenses (e.g. Creative Commons 4.0) are recommended for open research data and for fluidity should be considered with clear harmonised policies. Differences between the protection levels among Member States may hamper the free flow of research data within the Digital Single Market and make the implementation of cross-border licensing schemes difficult.

Create a level playing field for Text- and Datamining (TDM)

TDM are important functions in modern research. Depending on the actual technique used, and the materials analysed, TDM may already be permissible under current copyright laws within the limitations of scientific use. Yet in certain cases it could contradict copyright laws. More work on copyright in relation to TDM is recommended.

Next steps

This report is a first step in raising awareness in the complex field of open research data, data sharing and the reuse of research data. We believe it is important that the recommendations are followed up. Here are five steps we believe could help progress open research data.

1. Determine the level of urgency and interlinkage of the recommendations (on national and EU-level)

Although each recommendation holds a specific purpose, they can not be seen as completely independent of each other. Some recommendations are conditional and have to be (partly or fully) implemented in connection with other recommendations, such as recommendation 10 (Ensure sufficient funding for open research data and for data sharing activities). For instance, for certain elements of open research data, it needs to be identified who will be responsible for paying the costs, before they can be implemented.

The following recommendations can also be interlinked or dependent: recommendation 9 (Ensure the existence of FAIR open research data infrastructures) can only fully work if data is made identifiable (recommendation 5) and metadata is standardised (recommendation 6).

Some recommendations take longer to implement than others and some are more urgent then others; recommendation 8 (Promote the use of data management plans) is practical on more than one level, can be implemented quickly and can be beneficial beyond the scope of Open Science. Recommendation 1 (Promote a better understanding of open research data through communication & awareness raising) however is much more difficult to execute because of the broad scope and the underlying culture change.

Last but not least, some recommendations are very complex to implement or and could be burdensome to take forward (such as recommendations 4, 7 and 11).

2. Discuss the European Open Science Cloud and assessing the usability of other systems

Constructing this report the Task Force noticed many different (technological) systems relating to infrastructure, hardware, software or other functions necessary to work with open research data. For example, the enterprise architecture (EA) method was mentioned as a useful tool for the planning of IT infrastructure and services using.

A communication on the European Open Science Cloud is expected early 2016. The European Open Science Cloud initiative aims to position the EU in a leading role in scientific infrastructures globally and to ensure European stakeholders reap the full benefits of data-driven science and services for the digital economy and wider society.

It is however to be noted that excisting approaches should be taken into consideration. Therefore the opportunities of the Cloud should be assessed and discussed with Member States and stakeholders.

The Cloud should support the transition to Open Science in the context of the Digital Single Market and facilitate data access and reuse, while reducing the data storage costs.

It would be helpful if the Cloud increases awareness of the value of data and the potential of Open Science and when it will look into the structure of incentives and interoperability, as well as considers that the EU as a whole compromises of different countries with different research and innovation landscapes.

3. Set up a mechanism for regular exchange between stakeholders

Open research data involves many diverse stakeholders, including public funding agencies, research infrastructures, policy makers, universities, scientific publishers, the private sector and the general public. At a European and global level, different initiatives featuring these stakeholders have been developed to define the open research data agenda and aid its progression (this ERAC Task Force, for instance, is such an initiative).

Annex 7 includes a - non exhaustive - list of similar initiatives identified so far by this Task Force . We noticed that other initiatives indicate similar constraints and opportunities regarding open research data and propose both a number of comparable options as well as a few alternative solutions.

It would be of advantage to policy makers, and probably not only policy makers, to foster exchanges among the varied initiatives at European and global level. For instance discussion targeted towards specific audiences could be conducted at one level, such as the Global Research Council or OECD-GSF, whilst global aspects relevant to wider groups of stakeholders could be discussed and examined at an annual platform or forum initiated and moderated by the European Commission.

A system similar to the network of National Points of Reference (NPR) that is currently in place under Article 8 of the European Commission Recommendation on access to and preservation of scientific information from 17 July 2012 C(2012), could be proposed for other geographic areas. The OECD, for example, is considering this issue. Given the initiatives above, a platform for regular exchange between stakeholders to harvest and assemble various initiatives and reports will be helpful.

Such a mechanism, system or platform/forum does not necessarily need to lead to additional meetings, it could be implemented as a (partly) digital way of exchange.

4. Consider the possible difficulties of open research data, data sharing and the reuse of data in relation to public-private cooperation.

While the focus of this report is on publicly financed research and the data thereof, much of the research projects and programmes are done in a consortium with both public and private partners. The funding of these projects are often (partly) brought onto the projects by private partners who want a return on their investments and have legitimate business interests. They want to register industrial property rights and keep trade secrets.

Although within Open Science the possibilities exist to take the concerns of the stakeholders involved in public/private partnerships into account, major concerns remain, especially (but not only) in the private sector. We recommend that sufficient and broad attention is given to these concerns private partners may have and to the possibilities Open Science provides. Meetings and discussions with these stakeholders will be helpful and necessary in this process.

5. Determine the spot on the horizon and influence the strategy in its direction.

At some point in the transition towards working in Open Science and working with open research data, it is recommended to consider whether to define a long-term vision on open research data (the spot on the horizon), so that instruments and funding can be aligned to aim for this spot on the horizon.

We advise ERAC to monitor the ongoing process on open research data and stay on top of the developments in this area in the upcoming years, as well as to stimulate that progress will be made on the next steps described in this report.

References

ARCHAMBAULT, E., AMYOT, D., DESCHAMPS, P., NICOL, A., PROVENCHER, F., REBOUT, L., & ROBERGE, G. (2014). *Proportion of Open Access Papers Published in Peer-reviewed Journals at the European and World Levels, 1996-2013. RTD-B6-PP-2011-2: Study to develop a set of indicators to measure Open Access*. Science-Metrix. http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf.

ARZBERGER, P., SCHROEDER, P., BEAULIEU, A., BOWKER, G., CASEY, K., LAAKSONEN, L., & WOUTERS, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, *3*, 135-152.

BEAGRIE, N, & HOUGHTON, J. (2014). *The value and impact of data sharing and curation: A synthesis of three recent studies of UK research data centres*. London: JISC. http://www.cni.org/news/jisc-report-value-impact-of-data-curation-and-sharing/, accessed on 11 June 2015.

BRASE, J., FARQUHAR, A., GASTL, A., GRUTTEMEIER, H., HEIJNE, M., HELLERF, A., PIGUET, A., ROMBOUTS, J., SANDFAERF, M., & SENS, I. (2009). Approach for a joint global registration agency for research data. *Information Services and Use*, 29(1), 13-27.

CANDELA, L., CASTELLI, D., MANGHI, P., & TANI, A. (2014). Data Journals: A Survey. *Journal of the Association for Information Science and* Technology. DOI: 10.1002/asi.23358.

CEBR (2012). *Data equity: Unlocking the value of big data*. London: Centre for Economic and Business Research. www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf, accessed on 10 June 2015.

CHAVAN, V., & L. PENEV (2011). The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12, Supplement 15, S2.

COAR (Confederation of Open Access Repositories) (2015). *Promoting Open Knowledge and Open Science Report of the Current State of Repositories*. https://www.coar-repositories.org/files/Promoting-Open-Knowledge-and-Open-Science-Report-on-the-current-state-of-Repositories.pdf.

CODATA-ICSTI (Committee on Data for Science and Technology - International Council for Scientific and Technical Information) Task Group on Data Citation Standards and Practices (2013). Out of cite, out of mind: The current state of practice, policy and technology for the citation of data. *Data Science Journal*, 12, 1-75.

CORTI, L., VAN DEN EYNDEN, V., BISHOP, L., & WOOLLARD, M. (2014). In D.A. Thurston (Ed.). *Managing and Sharing Research Data. A Guide to Good Practice.* (pp. 24-27). London: SAGE Publications.

CREASER, C., Fry, J., GREENWOOD, H., PROBETS, S., SPEZI V., & WHITE, S. (2011). PEER Behavioural Research : Authors and Users vis-à-vis Journals and Repositories. Final Report. http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf, accessed on 30 July 2013.

DALLMEIER-TIESSEN, S., et al. (2011). Highlights from the SOAP project survey: "What scientists think about Open Access publishing". http://arxiv.org/abs/1101.5260, accessed 10 June 2015.

DAVIS, P.M., & WALTERS, W.H. (2011). The impact of free access to the scientific literature: a review of recent research. *Journal of the Medical Library Association: JMLA*, 99(3), 208.

EUROPEAN COMMISSION (2012). *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Towards better access to scientific information: Boosting the benefits of public investments in research*. https://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

EUROPEAN COMMISSION (2013). *Guidelines on Data Management in Horizon 2020, Version 1.0*.

FECHER, B., FRIESIKE, S., & HEBING, M. (2015) What Drives Academic Data Sharing? *PLoS ONE*, 10(2): e0118053. doi:10.1371/journal.pone.0118053.

FRASS, W., CROSS, J., & GARDNER, V. (2014). Taylor & Francis Open Access Survey. http://www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf, accessed on 30 June 2014.

GRIENEISEN, M. L., & ZHANG, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLoS ONE*, 7(10): e44118. doi:10.1371/journal.pone.0044118

GROVES, T. (2010). The wider concept of data sharing: View from the BMJ. *Biostatistics*, 11(3), 391-92.

HARNAD, S. (2012). *Open Access to Refereed Research Publications and Open Access to Research Data: A Crucial Strategic Distinction*. http://vimeo.com/17302549, accessed on 17 November 2014

HODSON, S., & MOLLOY, L. (2014). *Current best Practice for Research Management Policies, A memo for the Danish e-infrastructure Cooperation and the Danish Digital library*

HOUGHTON, J., RASMUSSEN, B., & SHEEHAN, P. (2010). *Economic and social returns on investment in open archiving publicly funded research outputs*. *Report to the Scholarly Publishing and Academic Resources Coalition (SPARC)*. Center for Strategic Economic Studies. Victoria University.

HOUGHTON, J., & SHEEHAN, P. (2009). Estimating the potential impacts of Open Access to research findings. *Economic Analysis and Policy*, 29(1), 127-42.

HOUGHTON, J., & SWAN, A. (2013). Planting the green seeds for a golden harvest: Comments and clarification on going for gold. *D-Lib Magazine*, 19(1/2).

HOUGHTON, J., SWAN, A., & BROWN, S. (2011). *Access to research and technical information in Denmark. Technical Report*. School of Electronics and Computer Science. University of Southampton.

HOUSEWRIGHT, R., SCHONFELD, R.C., & WULFSON, K. (2013). Ithaka S+R USFaculty Survey 2012. http://www.sr.ithaka.org/sites/all/modules/contrib/pubdlcnt/pubdlcnt.php?file=http://www.sr.ithaka.org/sites/default/files/reports/Ithaka_SR_US_Faculty_Survey_2012_FINAL.pdf&nid=502.

JONES, S. (2011). *How to Develop a Data Management and Sharing Plan*. Digital Curation Centre. Scotland.

JONES, S., PRYOR, G., & WHYTE, A. (2013). How to develop Research Data Management services – a guide for HEIs. In *DCC How-to Guides* (pp. 11-12). Edinburgh: Digital Curation Centre.

Knowledge Exchange (2016). *Funding Research Data Infrastructures, Landscape report*. In prep. preparation.

KOTARSKI , R., REILLY, S., SMIT, E., & WALSHE, K. (2012). *Report on best practices for citability of data and on evolving roles in scholarly communication. Opportunities for Data Exchange*. www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf.

KOWALCZYK, S., & SHANKAR, K. (2010). Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45, 247-94.

LEFÈBVRE, M. (2006). Vers une remise en question des modalités traditionnelles d'évaluation des savoirs scientifiques ? Séminaire « Communication scientifique et valorisation de la recherche à l'heure d'Internet ». Urfist-SCD Toulouse 1.

LEFÈBVRE, M. (2008). L'évaluation des savoirs scientifiques : modalités et enjeux. In J. Schopfel (dir.). *La publication scientifique : analyses et perspectives* (pp. 299-316). Paris : Hermès.

LEFÈBVRE, M. (2010). Rendre public le processus d'évaluation de la recherche. *Réseaux*, 164(6), 71-96.

MANYIKA, J., CHUI, M., FARRELL, D., Van KUIKEN, S., GROVES, P., & DOSHI, E.A. (2014). *Open Data: Unlocking Innovation and Performance With Liquid Information*. McKinsey & Co Report. http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

MARTIN, S., FOULONNEAU, M., TURKI, S., IHADJADENE, M., PARIS, U., & TUDOR, P.R.C.H. (2013). Risk Analysis to Overcome Barriers to Open Data. *Electronic Journal of e-Government*, 11(1), 348-359.

MCKIERNAN, E.C. (2014). *My concerns about PLOS's new Open Data policy*. Personal Blog. https://emckiernan.wordpress.com/2014/02/26/my-concerns-about-ploss-new-open-data-policy/

MOONEY, H., & NEWTON, M.P. (2012). The anatomy of a data citation: Discovery, reuse, and credit. *Journal of Librarianship and Scholarly Communication*, 1(1), eP1035. http://dx.doi.org/10.7710/2162-3309.1035, accessed on 10 June 2015.

OECD (2014). *Measuring the Digital Economy: A New Perspe*ctive. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264221796-en.

OECD (2015a). *Making Open Science a reality*. https://www.innovationpolicyplatform.org/sites/default/files/DSTI-STP-TIP%282014%299-REV2-for%20declassification%20-%20US%20and%20other%20comments%20addressed-numbers-removed.pdf

OECD (2015b). *Data-Driven Innovation for Growth and Well-Being*. Paris: OECD Publishing.

PIWOWAR, H. (2013). Altmetrics: Value all research products. *Nature*, *493*(7431), 159-159.

PIWOWAR, H., DAY, R.S., & FRISMA, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE*, 2(3).

PIWOWAR, H., & VISION, T.J. (2013), Data reuse and the open data citation advantage. *PeerJ*, 1:e175. http://dx.doi.org/10.7717/peerj.175.

RECODE Project Consortium (2014). *Reports Policy recommendations for Open Access to research data*. http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf.

ROCHE D.G., LANFEAR R., BINNING S.A., HAFF T.M., SCHWANZ L.E., CAIN K.E., et al. (2014). Troubleshooting Public Data Archiving: Suggestions to Increase Participation. *PLoS Biol*, 12(1): e1001779. doi:10.1371/journal.pbio.1001779. http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001779

Royal Society (2012). *Final report: Science as an enterprise*. Royal Society Science Policy Centre Report. https://royalsociety.org/policy/projects/science-public-enterprise/Report/, accessed on 11 June 2015.

SAVAGE, C.J., & VICKERS, A.J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PloS one*, 4(9), e7078.

SPIEGLER, D.B. (2007). The private sector in meteorology: An update. *AMS Journals Online*. http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-88-8-1272, accessed on 11 June 2015.

VAN DEN EYNDEN, V., BISHOP, L. (2014). *Incentives and motivations for sharing research data, a researcher's perspective*. A Knowledge Exchange Report. knowledge-exchange.info/Default. aspx?ID=733

VÆRING ANDERSEN, A. (2014). *FIF – Fælles Infrastruktur for Forskningsdata*. DEFF.

VETENSKAPSRÅDET (2015). *Proposal for Swedish National Guidelines for Open Access to Scientific Information*. Swedish Research Council. https://publikationer.vr.se/en/product/proposal-for-national-guidelines-for-open-access-to-scientific-information/

WICHERTS, J.M., BAKKER M., & MOLENAAR, D. (2011). Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. *PLoS ONE*, 6(11), e26828. doi:10.1371/journal.pone.0026828

Glossary and appendices

Abbreviations

API	Application Programming Interfaces
CC-license	Creative Commons License
COAR	Confederation of Open Access Repositories
CRIS	Current Research Information Systems
DANS	Data Archiving and Network Services
DataONE	Data Observation Network for Earth
DCAT-AP	The Data Catalogue Vocabulary Application Profile
DMP	Data Management Plan
DMPonline	A tool that assists researchers to produce an effective Data Management Plan (DMP)
DOIs	Digital Object Identifiers
EA	Enterprise Architecture
ERAC	European Research Area and Innovation Committee
EUDAT	European Data Infrastructure
FOSTER	Facilitate Open Science Training for European Research
GenBank	A database that contains publicly available nucleotide sequences
Github	Web-based Git repository hosting service
IDOA	Immediate-Deposit & Optional-Access

IPR	Intellectual Property Right
LOD	Linked Open Data
LODLAM	Linked Open Data in Libraries, Archives and Museums
MANTRA	Research Data Management Training
NPR	National Points of Reference
OECD	Organisation for Economic Co-operation and Development
OECD-GSF	OECD Global Science Forum
OpenAIRE	The European Open Science Cloud for Research
OpenDOAR	Directory of Open Access Repositories
PLOS	Public Library of Science
PSI	public sector information
RDMRose	A JISC funded project to produce taught and continuing professional development (CPD) learning materials in Research Data Management (RDM) tailored for Information professionals
re3data.org	global registry of research data repositories
RECODE	The Policy RECommendations for Open Access to Research Data in Europe
RFO	Research Funding Organizations

ROAR	Registry of Open Access Repositories
SME	Small and Medium-sized Enterprizes
STEM	Science, Technology, Engineering, and Mathematics
TDM	Text- and Datamining
URI	Uniform Resource Identifier
ZENODO	An open dependable home for the long-tail of science, enabling researchers to share and preserve any research outputs in any size, any format and from any science

Annex 1 – The Task Force

People involved in (the meetings of) the Task Force and/or in realising this report

Austria – Peter Seitz, Claudia Leutgeb, Florian Pecenka

Belgium – Bart Dumolyn, Eric Laureys, Marc Vanholsbeeck

Croatia – Ivana Pavlaković

Denmark – Hanne-Louise Kirkegaard, Jonas Bak

Finland – Sami Niinimäki, Pirjo-Leena Försström

France – Mark Asch, Cécile Heriard, Marie-Pascale Lizée

Germany –Bettina Klingbeil, Philipp Zimbehl, Hans-Josef Linkens

The Netherlands – Dries van Loenen, Ron Dekker, Paul Ymkers, Davy Pieters, Francien Heijs, Marjan van Meerloo

Luxemburg – Lynn Wenandy

Portugal – Ana Neves, Vasco Vaz

Spain – Clara Eugenia Garcia

Sweden – Eva Stensköld, Viktoria Mattsson

Switzerland – Myriam Cevallos

United Kingdom – Tim Higginson, Philip Walker, Gavin Copeland, Ron Eggington

European Commission – Daniel Spichtinger, Celina Ramjoué, Jean-François Dechamp

Meetings (in Brussels)

13 June 2014

24 October 2014

24 March 2015

3 July 2015

8 October 2015

Formulation of the original wishes of ERAC

In December 2013, ERAC asked for more insights into the then upcoming subject of Open Innovation Data. This was also in line with the Communication of the Commission on Open Access in 2012 (European Commission, 2012).

Annex 2 - Categories of open research data/Research data sharing

In a 2015 study by Fecher et al. (Fecher et al. 2015) the lack of motivation to share research data was explored. As a results of a systematic literature review and a survey, the authors identified a framework depicting research data sharing in six descriptive categories:

6. **Data donor**; comprising factors regarding the individual researcher who is sharing data. The sub-categories are sociodemographic factors, degree of control, resources needed, and returns.

7. **Research organisation**; comprising factors concerning the crucial organisational entities for the donating researcher, being the own organisation and funding agencies (e.g., funding policies).

8. **Research community**; comprising factors regarding the disciplinary data sharing practices (e.g., formatting standards, sharing culture, scientific value and publications which are regarded as the primary currency in academia).

9. **Norms**; comprising factors concerning the legal and ethical codes for data sharing (e.g., copyright, confidentiality).

10. **Data recipients**; comprising factors regarding the third party reuse of shared research data (e.g., adverse use).

11. **Data infrastructure**; comprising factors concerning the technical infrastructure for data sharing (e.g., data management system, technical support).

The study discovered that the topic of usability is a common theme in the literature. The analysed material suggests that service providers need to simplify the sharing process and the tools used in the process. There is also strong support for easy guidelines on how to share data in addition to technical support for researchers.

To create an effective policy for data sharing, it is essential to understand the position and motivation of the involved parties. Policy measures need to provide incentives for sharing but also impede researchers not to share. Formal recognition and financial reimbursement for the extra work have been suggested as well as understandable and clear legal basis on what can and cannot be done with the collected data, which is especially important in disciplines with personal data. University level education on data curation, sharing culture, data documentation, and security should be considered and the data infrastructure needs to be up-to-date and easy-to-use.

Annex 3 - Open Access to publications vs open research data

From the preliminary report

It is important to realise that there are also clear distinctions between publications and data, which ask for a different approach, both on the policy level and the more operational level. On a preliminary basis, the Task Force has identified some of these differences, which should be kept in mind while developing policies on Open research data. In the table below, these differences are explained in a simplified way.

	Publications	Research data
Type of proces	Transition from an old to a new system	Construction of a totally new system
Challenge during the process	Passing through transition phase with different models at the same time	Urgent need for policy convergence and operational standardisation in order to avoid fragmentation
Degree of choice	Mostly either 100% open or not open, but some "in between" options (relating to copyrights and re-using)	Many intermediary options for access to only part of the data
Number of models	Limited (Green, Gold, etc) and (less preferably) several hybrid models))	Many options; no 'one size fits all'

Costs	Aim for budgetary neutral or maybe lower costs (apart from possible transition costs)	Building up of new infrastructure leads to additional costs. This is even more so because of the different sorts of data depending on the discipline
Costs for the user	Open Access = free of charge to the end-user	Who will pay the additional costs? Is it feasible to introduce a user fee or is it preferable to offer really Open Access (= free of charge to the end-user)?

Annex 4 – overview of possible dissemination tools and infrastructures

Some repositories are specifically designed to archive research datasets such as 3TU datacentrum, the mission of which is "to ensure the accessibility of technical scientific research during and after completion of research to give a quality boost to contemporary and future research" (http://datacentrum.3tu.nl/en/about-3tudatacentrum/background-and-mission).

Some other repositories make the data that underlie scientific publications – as well as other research outputs - discoverable, freely reusable, and citable. Dryad is such an "international multidisciplinary repository of data underlying scientific and medical publications" which "is a curated general-purpose repository that makes data discoverable, freely reusable, and citable" (Wikipedia, s.v. Dryad, https://en.wikipedia.org/wiki/Dryad_(repository)). Figshare, operated by Macmillan Publishers, is another example of online digital repository where researchers can free of charge preserve and share their research outputs, including figures, datasets, images, and videos and which has partnerships with other major actors of the Open Access publishing industry (PLoS, Nature Publishing Group, Taylor and Francis and F1000). Also Zenodo (http://zenodo.org/) is a general purpose repository, which can be added to this list.

On the other hand, repositories, at European (i.e. b2share/EUdat: https://b2share.eudat.eu/) or institutional (university) level, may be used to archive and share data. In particular, institutional repositories are increasingly used to store, preserve, curate and disseminate a broad scope of research products – including datasets -, in the perspective of setting up "a global network of repositories that will act as the foundation for access, certification, quality assessment and further reuse of research outputs" (COAR, 2015).

Research data infrastructures may benefit too from being shared by different institutions. Dataverse Netherlands (http://service.re3data.org/repository/r3d100011201) is an example of such an open source web application allowing researchers to share, preserve, cite, explore and analyze research data, which is open for organizations associated with Dutch universities to deposit data.

Finally, it is noticeable that Web tools have also recently been developed which intend to help scientists collaborate and produce research. Some of these allow researchers to deposit and openly share data, such as LabArchives (http://www.labarchives.com) which has specific arrangements with BioMed Central Open Access journals, or Github which constitutes "a potentially great resource for researchers to make their data publicly available" and can be used to "store data in the cloud for future use (for free), track changes, make data publicly available for replication, create a website to nicely present key information about the data; and uniquely: benefit from error checking by the research community." (http://www.r-bloggers.com/data-on-github-the-easy-way-to-make-your-data-available/)

Annex 5 - The pros and cons of open research data infrastructures and dissemination tools

	Pros	Cons
Open Access journals with Open Data mandates (for "for underlying data")	• Excellent discoverability and accessibility; • Excellent usability of the data: validation, replication, reanalysis, new analysis, reinterpretation, or inclusion into meta-analyses; • Data dissemination is directly linked to the publication activity, which is traditionally valorised in the recruitment and the promotion of researchers; • Allows peer-reviewers of the papers to check the scientific quality of the underlying data.	• The notion of "underlying" data remains vague. In some cases all related data are needed to replicate a research; • Reluctance to put data in Open Access may discourage researchers from publishing in journals with Open Data mandates; • Effective implementation of the mandate is not assured; • Supposes the existence of relevant Open Data infrastructures ; • The questionable prestige rankings of scholarly journals could reflect upon the perceived quality of related data.

	• Allows peer-reviewers of the papers to check the scientific quality of the underlying data.	• Supposes the existence of relevant Open Data infrastructures ; • The questionable prestige rankings of scholarly journals could reflect upon the perceived quality of related data.
Open Access data journals	• As a proper publication, a data paper provides academic accreditation to researchers (including citations); • Dedicated peer-review of the quality of the data (scientific and technical aspects); • Maximisation of the opportunities for data-reuse because of the excellent description of the published datasets and the high discoverability and accessibility of the data papers;	• Not all data journals mandate that the data discussed in the paper are archived in certified Open Access repositories; • Different standards with regard to the accessibility and the identification of data (although most of data journals require at least a DOI or URI for the shared datasets); • Remain within the traditional paradigm of journal publication.

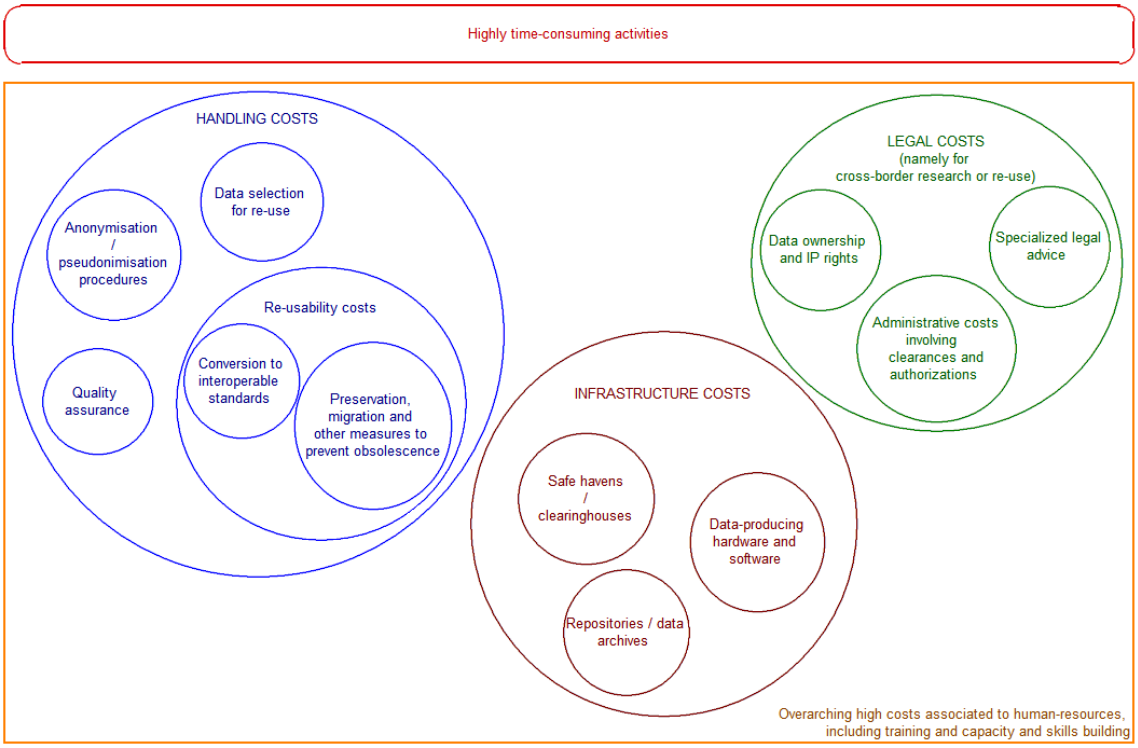
	• Possible usage of traditional bibliometrics (such as Impact Factors), as well as development of new data related metrics: "As data papers are becoming distinct publishing products, a number of data journals are also supporting alternative metrics (altmetrics), thereby enhancing further data publication." (RECODE, 2014).	
Web 2.0 collaborative tools for scientists	• Help scientists to include Open Access sharing of data into their daily routine and workflow; • Not linked to an institution: follows mobile researchers.	• Participate to the potentially inefficient multiplication of the places of repository; • Uncertainty about the sustainability of the tools.
Web 2.0 social networking and bibliographic management tools	• Well known tools; • Not linked to an institution: follows mobile researchers.	• Participate to the potentially inefficient multiplication of the places of repository; • Uncertainty about the sustainability of the tools;

		• "Dark Open Access" : access is only possible for the registered members (not real Open Access); • Not optimal discoverability; • Basic usage is free of charge but additional services and extra storage cost.
Research data repositories	• Dedicated tools for data archiving and data sharing; • Allow possible embargo periods; • Not linked to an institution: follows mobile researchers.	• Participate to the potentially inefficient multiplication of the places of repository; • Uncertainty about the sustainability of the tools.

Institutional repositories	• Existing mandates (IDOA mandates) at institutional and European level for Open Access could easily be extended to Open Data; • Inclusion of the data in the usage of institutional repositories to assess researchers and to steer universities; • Participate to the institutional visibility of universities; • Excellent discoverability; • Strong Open Access mandates; • Centralised metrics and altmetrics; • Extended possibilities of data mining (COAR, 2015).	• Even if it is technically possible to archive data in institutional repositories, current institutional repositories are mainly used to share papers than other types of research outputs (Archambault et al., 2014); • Librarians are not always the best trained specialists to curate data repositories.

Personal websites		• Lack of standardisation; • No quality assurance.
Cloud storage	• Usual way for teams of researchers to share project related data.	• Not open for not-members of the team (not real Open Access).

Annex 6 – Schematic overview of costs



Annex 7 - Ongoing initiatives

Global perspective

There are several international organisations who deal with open research data, primarily through non-binding policy guidelines, statements of intent and multilateral declarations. The majority of these are a result of a general Open Access policy that has gradually evolved since the beginning of the millennium. There has only been initiated few broader and more substantial specific worldwide open research data initiatives and none of these are politically anchored in an international organisation.

Chosen international initiatives and policies concerning ORD:

- **Budapest Open Access Initiative (BOAI) – Berlin Declaration (BD)**

  BOAI is an international initiative that was launched in 2002 by the private organisation Open Society Institute (OSI)[6]. The purpose of this initiative was to bring the international research community together around a common definition and a common statement of intent vis-à-vis OA in order to secure its expansion and implementation. Today, BOAI is considered by many sources as the precursor for the current more widely accepted understanding of Open Access[7]. In 2003 BD was launched by the Max Planck Society together with *"The European Cultural Heritage Online Project"*. BD is based on BOAI's understanding and implementation of OA but it has a deeper interpretation of Open Access, as it includes open research data as an independent element in its definition of Open Access. BD has since then been signed by more than 475 organisations[8].

[6] OSI is a part of the Open Science Foundation which is owned and lead by George Soros
[7] See: http://www.opensocietyfoundations.org/press-releases/scientists-foundations-libraries-universities-and-advocates-unite-and-issue-new
[8] http://openaccess.mpg.de/319790/Signatories

Links:

BOAI – Official homepage

Berlin Declaration 2003

- **G8**

The Science Ministers from the G8 Member States and the EU Commissioner for Research met during the G8 conference in June 2013 to discuss science policy including Open Access and open research data[9]. The discussions were followed by a written Policy Statement which was signed by all the participants. In the Policy Statement it is acknowledged that it is important to ensure, that the results of scientific research are as widely available as practical. The document is not binding for its signatories, but it should be seen as a statement of intent and as an encouragement for further international cooperation and coordination.

Links:

G8 Policy Statement

- **OECD**

In 2007, the OECD published a report on principles and recommendations for expanding open research data. The report has been prepared on the basis of the Member States' declaration vis-à-vis "Access to Research data from Public Funding" from 2004 which was signed by all the Member States at that time as well as China, Israel, Russia and South Africa. Neither the declaration nor the report contains politically binding agreements but the report is still in 2014 used as a political reference document[10]. The OECD also undertakes continual public monitoring of the Member States' initiatives.

[9] The meeting was a first of its kind. A similar group consisting entirely of Science Ministers called "the Carnagie Group" held meetings in the period 2005-10.
[10] See Damvad's report (in English) on ORD made for the Norwegian Research Council (2014)

OECD has recently (29 Sept 2015) published a the report "Making open science a reality" (OECD, 2015a), which gives a good overview.

Links:

Declaration on Access to Research data from Public Funding (2004)

Principles and guidelines for Access to Research Data from Public Funding (2007)

OECD's webpage for monitorering af medlemslandenes politikker

- **UNESCO**

UNESCO has, on the basis of their official Open Access strategy, prepared a report containing several non-binding policy guidelines about the implementation of Open Access (including open research data, though to a more limited extent). The report has been prepared by Dr. Alma Swan, a leading expert in the field of Open Access, in cooperation with WSIS Communities. In addition to its focus on Open Access in a research perspective, it also includes a more educational perspective.

Links:

UNESCO's strategy for open acess to science (2013)

Policy Guidelines for the development and promotion of Open Access (2012)

WSIS Commuinities

- **RDA – The Research Data Alliance**

  RDA is an independent alliance established with the aim to expand open research data. The alliance was formed in 2012 by an international group consisting of members from private companies, universities and funding agencies from USA, EU and Australia[11]. The alliance is an open bottom-up driven organisation where all different kind of actors (individuals as well as private companies and universities) is invited to take part in and contribute to the activities of the alliance. The organisation itself is still at an early stage but different types of events and symposiums are already being held including two yearly plenary sessions for all the members. The alliance is officially supported by the EU, USA and Australia.

  Link:

  Research Data Alliance (RDA)

- **Science 2.0**

  Science 2.0 is a new approach to general research practice. The approach is based on openness and is driven by new technological opportunities for sharing, especially the internet. In brief, Science 2.0 is best described as an approach where the researchers' work, ideas, data and research results are made available to the public. This makes it possible for fellow researchers to contribute or comment on the work/data. Science 2.0 is in other words an internet-based grass root phenomenon that has gradually grown over the years and thus become the subject of interest from several big research institutions and political organisations. In 2006 the private company ION Publications made a webpage with the same name and purpose as Science 2.0. In extension ION now has the copyright on the name.

  A consultation on Science 2.0 was held among Member States from June - Sept 2014.

  Links:

  About Science 2.0 on Wikipedia

  Webpage for ION Publications initiative

[11] Including Microsoft, Australian National Data Service and Karlsruhe Institute of Technology

EU

The work with open research data in the EU is split in two categories: 1) open research data in the EU's own programs and 2) EU-coordination of open research data as a practice and concept in the Member States.

The EU's most valuable instrument to promote open research data is the Union's Framework Programme for research and innovation (FP). The FP is managed by the European Commission's Directorate General for research (DG Research) with reference to several laws passed by the Member States (Competitiveness Council) and the EU Parliament. The most important of these being the Regulation of the European Parliament and of the Council establishing the Framework Programme, the Council Decisions establishing the Specific Programme(s) implementing the FP and the Regulation of the European Parliament and of Council on the rules for participation. Open Access is mentioned for the first time in history in both documents for the current FP (Horizon 2020) as an important area that should or has to be prioritised in the different parts of the program.

- **FP7 / Horizon 2020**

  The Commission has worked with Open Access in the form of pilot projects in two framework programs respectively FP7 and Horizon 2020. In FP7 a pilot project covered seven different sub areas equivalent to 20 % of the entire programme's budget[12]. As a result of their status as pilot projects the project agreements under these programs required that the participants should archive their articles in a parallel digital archive and do their utmost to ensure public access to their articles within 6-12 months depending on the discipline. The Open Access activities are carried on in Horizon 2020 where the principle from the pilot project is extended to the entire program. In addition to this, a new similar pilot project is added to Horizon 2020 with a focus on open research data. As for the Open Access pilot project from FP7 open research data is introduced as an integrated element in the contracts for the projects that are receiving grants in a number of sub programs[13]. Researchers who

[12] The thematic areas Energy, Environment, Health, SSH and ICT as well as the programs for Science in Society and Research Infrastructures
[13] FET, Research Infrastructures, LEIT-ICT, Energy, Climate, Inclusive society and Science with and for Society

participate in these programs are encouraged to include open research data as a part of their project implementation. However, they are allowed to opt out or not to include all data if they wish to protect sensitive data. In addition to the pilot projects the EU has given financial support to the projects OpenAIRE and OpenAIRE+ both under FP7 and Horizon 2020. The two projects have been designed to support the expansion of Open Access and open research data in Europe both through counseling and by working as a knowledge bank and Open Access web-portal.

Links:

FP7 OA Pilot

Horizon 2020 OA-ORD-pilot guidelines´

Openaire/openaire+

- **Coordinated Open Access activities**

  The EU does not have a mandate to impose the Member States specific regulations regarding Open Access or open research data in national research policies. However, since the establishment of the European Research Area (ERA) in 2007 both the Commission and the Council have issued notices and recommendations on the subject. The latest example of this dates from July 2012 where the Commission promulgated a recommendation to the Member States vis-à-vis the implementation of both Open Access and open research data in their national research policy. Subsequently in 2013 the consultancy firm Science Metrix on request from the Commission produced a comparative analysis of open research data in and outside of the European Research Area. Further, in the summer of 2014, the EU held an extensive consultation regarding Science 2.0 and its added value for ERA. The results have become available during the course of 2014.

Links:

Overview of the EU policy initiatives within OA

Science Metrix report regarding ORD policies (2013)

Commission recommendation (2012)

Riding the wave – ORD report from High Level Expert Group (2010)

- **Working groups / Networks**
  - ”informal network for dialogue on access to scientific information”
  - Digital Era Forum: (vorher e-Infrastructures Policy Forum)

    Subgroup on “Knowledge circulation – Open Access to data and to publications and to Digital ERA”
  - ERAC Working Group on Knowledge Transfer: (Ref. 223)
- **EU-funded projects**

  – **Pasteur 4 OA (**http://www.pasteur4oa.eu/project)

PASTEUR4OA supports the aim of encouraging the development of matching policies on Open Access and open research data in the European Union, according to the European Commission’s recent Recommendation on “Access to and preservation of scientific information” (July 2012) and in view of maximising alignment with the Horizon 2020 policy on access to the research funded by the Commission. The project helps develop and/or reinforce Open Access strategies and policies at the national level and facilitate their coordination among all Member States. It will build a network of centres of expertise in Member States that will develop a coordinated and collaborative programme of activities in support of policymaking at the national level under the direction of project partners.

The project builds on an already existing project, Mediterranean Open Access Network (MedOANet - http://www.medoanet.eu) capitalising on its work and an already established network within Mediterranean Europe. Further, it will take advantage of the experience and extensive networks of organisations such as EOS (Enabling Open Scholarship), JISC (Joint Information Systems Committee), SparcEUROPE, LIBER (Association of European Research Libraries), EIFL, as well as prominent funding organisations that participate in the consortium, to secure a European-wide engagement of bodies of authority of the Member States with the project's aims and extend its impact beyond Member States to neighbouring Accession States.

More specifically, PASTEUR4OA will improve coordination in developing OA policies that align to the European Commission's Recommendations and Horizon 2020 rules by realising the following measurable and verifiable objectives:

- The identification of Key Node organisations throughout Europe and in accession/associated states, on the basis of their institutional profile, record and ability to influence policymaking, and the development of a network of expert organisations
- The development of a programme for engaging policymakers
- A Europe-wide project meeting of national experts
- Establish the foundations of a Knowledge Net by the end of the project through continuous engagement of the Key Node organisations.
- Recording policies and policy types in order to develop a policy typology
- Policy analysis: effectiveness and growth
- A mapping of existing policies to policymakers
- Development of advocacy materials
- The identification of policymakers in the MS and accession/associated states

- Policymaker engagement
- The project's final conference which will bring together Key Node members and policymakers, as well as provide the opportunity for a wider presentation of project achievements

– **Foster: Faciliate Open Science Training for European Research** (http://www.fosteropenscience.eu/)

Open Access and Open Science principles are an essential part of knowledge creation and sharing. They directly support the researchers need for greater impact, optimum dissemination of research, while also enabling the engagement of citizen scientists and society at large on societal challenges. This two year project aims to set in place sustainable mechanisms for EU researchers to FOSTER OPEN SCIENCE in their daily workflow, thus supporting researchers optimising their research visibility and impact, the adoption of EU Open Access policies in line with the EU objectives on Responsible Research & Innovation.

– **Recode Policy RECommendations for Open Access to Research Data in Europe** (http://recodeproject.eu/)

The Policy RECommendations for Open Access to Research Data in Europe (RECODE) project will leverage existing networks, communities and projects to address challenges within the Open Access and data dissemination and preservation sector and produce policy recommendations for Open Access to research data based on existing good practice.

The RECODE partners (Trilateral Research & Consulting, The e-Humanities group, The University of Sheffield, The Stichting LIBER Foundation ,National Documentation Centre, National Research Council of Italy, Blekinge Institute of Technology, Amsterdam University Press) will identify relevant stakeholders, build upon and strengthen existing stakeholder engagement mechanisms. It will conduct studies of good practice and exchange good practice principles with relevant stakeholders and institutions during networking activities. The RECODE project will culminate in a series of policy recommendations for Open Access to research data targeted at different stakeholders and policy-makers.

- **Activities of the Science Community**

  - **Global research council** (http://www.globalresearchcouncil.org/)

The Global Research Council is a virtual organisation, comprised of the heads of science and engineering funding agencies from around the world, dedicated to promoting the sharing of data and best practices for high-quality collaboration among funding agencies worldwide.In May 2013 an Action plan on Open Access was developed.

  - Knowledge exchange group (http://www.knowledge-exchange.info/Default.aspx?ID=115)

Knowledge Exchange is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research.

The Knowledge Exchange partners are:

  - CSC - IT Center for Science in Finland
  - Denmark's Electronic Research Library (DEFF) in Denmark
  - German Research Foundation (DFG) in Germany
  - Jisc in the United Kingdom
  - SURF in the Netherlands

The Knowledge Exchange partners express a common vision based on our five national strategies:'To make a layer of scholarly and scientific content openly available on the Internet.' Making this vision a reality means creating the building blocks for a European information environment that is outstanding in its support of research and higher education, bringing Europe together while extending the reach of European research and teaching around the globe. The partners work together on activities in the field of Open Access, Research Data, Research Tools and Technologies and Interoperability Standards.

– **Science Europe** (http://www.scienceeurope.org/)

Science Europe is an association of European Research Funding Organisations (RFO) and Research Performing Organisations (RPO), based in Brussels. Its founding General Assembly took place in Berlin in October 2011. Science Europe promotes the collective interests of the Research Funding and Research Performing Organisations of Europe. It supports its member Organisations in their efforts to foster European research. It will strengthen the European Research Area (ERA) through its direct engagement with key partners. In doing so it will be informed by direct representation of all scientific communities in its reflections on policies, priorities and strategies.

– **EARTO** (http://www.earto.eu/)

EARTO is a non-profit international association established in Brussels, where it maintains a permanent secretariat.

EARTO Vision: a European research and innovation system without borders in which RTOs occupy nodal positions and possess the necessary resources and independence to make a major contribution to a competitive European economy and high quality of life through beneficial cooperation with all stakeholders.

EARTO Mission: to promote and defend the interests of RTOs in Europe by reinforcing their profile and position as a key player in the minds of EU decision-makers and by seeking to ensure that European R&D and innovation programmes are best attuned to their interests; to provide added-value services to EARTO members to help them to improve their operational practices and business performance as well as to provide them with information and advice to help them make the best use of European R&D and innovation programme funding opportunities.

The Association represents the interests of about 350 RTOs from across the European Union and "FP-associated" countries (90 direct members, some of which are associations regrouping several RTOs).

EARTO has released a position paper on the 13th of November 2015 titled: "Is there still room for Open Innovation in a Digital Single Market built op Open Science" EARTO members require a "balanced" approach to open research data.

– **LERU** (www.leru.org)

Since its founding in 2002, the League of European Research Universities (LERU) has emerged as a prominent advocate for the promotion of basic research at European universities. LERU strongly believes that basic research plays an essential role in the innovation process and significantly contributes to the progress of society.

LERU aims at furthering the understanding and knowledge of politicians, policy makers and opinion leaders about the role and activities of research-intensive universities. Drawing on the impressive academic potential and expertise of its network, LERU has a strong and significant impact on research policy in Europe.

LERU is dedicated to continuing this through direct communication and discussions within the policy community, sound proposals for further improvement and progress, and carefully considered position papers on fundamental issues such as academic careers and the role of universities.

LERU documentation on OA to publications and research data

- **LERU Roadmap towards Open Access:** http://www.leru.org/files/publications/LERU_AP8_Open_Access.pdf
- **LERU Roadmap for Research Data**: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf
- Global Research Council (GRC)-**Review of Implementation of the GRC Action Plan towards OA to Publications**: http://www.jsps.go.jp/information/data/GRC_Open_Access_Review_of_Implementation_-_Summary_Report.pdf

– **Business Europe** (www.businesseurope.eu)

BUSINESSEUROPE is the leading advocate for growth and competitiveness at European level, standing up for companies across the continent and campaigning on the issues that most influence their performance. A recognised social partner, we speak for all-sized enterprises in 34 European countries whose national business federations are our direct members.

The organisation is headquartered in Brussels at the heart of the EU institutions. We work on behalf of our member federations to ensure that the voice of business is heard in European policy-making. We interact regularly with the European Parliament, Commission and Council as well as other stakeholders in the policy community. We also represent European business in the international arena, ensuring that Europe remains globally competitive.

They made a position paper on Open Access already in 2012. see https://www.businesseurope.eu/sites/buseur/files/media/imported/2012-01049-E.pdf .

- **Selected national ORD initiatives**

ORD is increasingly mentioned in research fora's and administrations in and outside the EU. However, it is only very few countries who have developed an official ORD policy and often only as a result of a bigger openness policy and only on the rarest occasion as an actual separate research policy. The remaining specific initiatives are voluntary and the majority of the efforts in creating an infrastructure continue to be on a very early pilot stage.

- Europe

  In Europe several countries are in the process of establishing an e-infrastructure for sharing public data including research data. Among the leading countries in the OECD's monitoring database are Finland (Social Science Data Archive), the Netherlands (open and voluntary cross-disciplinary database), Norway (voluntary national database), Denmark (Open Access Barometer under development) and Great Britain (center focusing on securing optimal use of open research data).

Links:

Finland's social science data archive

The Netherlands open data archive – DANS

Norway's open data portal

Great Britain's open data institute

Denmark Open Access Barometer: http://www.forskningsdatabase.dk

- Australia

  Australia has positioned itself as one of the leading countries outside the EU with regard to OA. Among other activities, the country has been very much involved in RDA. Furthermore, since 2013 Australia's national research council (ARC) has had an OA policy that imposes all research results that have been produced with the support of the ARC to be made public available within a period of 12 months.

  Link:

  Australia's Open Access Policy.

- Canada

  Canada has launched a general political initiative regarding *"open source government"*. Through this, publicly generated data are made available for the public. In addition, three big Canadian research councils have come together on a common OA policy including a special requirement vis-à-vis ORD (yet, this only applies to the Canadian Institute of Health Research (CIHR)).

  Links:

  Open Data portal

  Tri-Agency Open Access policy

- USA

  USA has launched a movement towards a more open source government with an "Open data policy" and a corresponding project under the title "Project open data". Both initiatives have been launched in order to make data generated through public funds more accessible for researchers as well as for the civil society. In addition, the Office of Science and Technology Policy (OSTP) sent out a memo in February 2013 imposing every American research councils and agencies with a yearly distribution level over a 100$ to produce a plan for the implementation of OA and ORD in their respective systems.

  Links:

  Project Open Data

  US – Open data policy

  Memorandum – Increasing Access to the Results of Federally Funded Scientific Research