

Open Research Data Pilot: barriers to implementation and instruments for supporting research

RITA BLANOS, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS

CATERINA FANARA, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS

Tutor: GIUSEPPE SCELLATO, Politecnico di Torino

Tutor esterno: PAOLA GARGIULO, CINECA

Master in Management of Research, Innovation and Technology
Master MIT IV edizione

FEBBRAIO 2016 – LUGLIO 2017

CONTENTS

| | |
|---|----|
| EXECUTIVE SUMMARY | 3 |
| 1. INTRODUCTION AND PURPOSES | 4 |
| 2. OPEN ACCESS | 6 |
| 2.1 Peer-reviewed scientific research articles | 7 |
| 2.2 Open access to research data..... | 8 |
| 2.3 Legal issues on OA..... | 9 |
| 2.4 Ethical issues on OA..... | 9 |
| 3. OPEN RESEARCH DATA PILOT..... | 10 |
| 3.1 Data Management Plan..... | 11 |
| 3.2 Data repository..... | 12 |
| 3.3 Licenses | 13 |
| 3.4 Opting out – partially or entirely | 14 |
| 3.5 Overview about the ORD up-take..... | 15 |
| 4. PROJECTS SUPPORTING OPEN ACCESS | 16 |
| 5. BARRIERS AND ENABLERS TO THE IMPLEMENTATION OF THE ORD | 19 |
| 5.1 Barriers and resistances to implementation | 19 |
| 5.2 Advantages and enablers | 20 |
| 6. METHODOLOGY | 21 |
| 7. TOOLS SUPPORTING THE ORD PILOT | 25 |
| 7.1 Guidelines and OA resources..... | 25 |
| 7.1.1 EC guidelines and documentation | 26 |
| 7.1.2 Other resources | 27 |
| 7.2 Support kits..... | 27 |
| 7.3 Help Desks..... | 28 |
| 7.4 Tools for Data Management Plans..... | 29 |
| 7.5 Finding a research data repository | 30 |
| 7.6 Licensing tools and legal guidelines and resources | 31 |
| 7.7 Costing Data Management..... | 33 |
| 7.8 On-line courses | 33 |
| 8. CONCLUSIONS AND RECOMMENDATIONS | 35 |
| 8.1 Specific recommendations for H2020 applicants | 37 |
| 8.2 Other general recommendations | 37 |
| 9. FURTHER REFERENCES | 38 |
| ANNEX 1: QUESTIONNAIRE A | 39 |
| ANNEX 2: QUESTIONNAIRE B | 40 |
| ANNEX 3: ORGANIZATION PROFILES..... | 41 |

EXECUTIVE SUMMARY

According to the EC policy, *Open Access generally refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable.*

The Europe 2020 Strategy for a smart, sustainable and inclusive economy fosters knowledge and innovation as the means that lead to growth and, in this context, Open Access (OA) can play a very important role. In fact, making research results (publications and data) accessible, accelerates dissemination of knowledge and scientific discovery, and enables new forms of data-intensive research.

The central aim of the present work is to highlight the grand challenges in opening access to Research Data and to identify main barriers and resistance to its implementation.

The study features a literature review of policy documents on OA, analysis of the current state of art and study of reports and projects. Structured interviews with relevant contact persons in different research performing organisations were later conducted, providing important feedback about the experience in participating in the Open Research Data (ORD) pilot, with special reference to the problems they faced and the solutions as well as the tools they experimented.

In addition the main resources on OA (guidelines, toolkits, licensing and costing tools, helpdesks, etc.), with specific reference to open research to data and data management issues, were analysed and listed and later grouped by topic. The list is not exhaustive, although it represents a helpful starting point for researchers and relevant stakeholders within research institutions and universities when approaching OA issues, with specific reference to the participation in the pilot.

On the basis of this work, we have identified specific recommendations to be addressed to research performing organisations with the aim of helping them in reviewing their approach and practices on OA to research data and data management, and in taking the necessary steps to rise the exciting challenge introduced in the H2020 programme with launch of the ORD pilot.

1. INTRODUCTION AND PURPOSES

Among the three goals of the EU research and innovation policy set by Commissioner Moedas, *Open Science (OS)* represents a **new approach to the scientific process based on cooperative work and new ways of spreading knowledge by using digital technologies and new collaborative tools**¹. Open Science represents a systemic change to the way science and research is done. It goes even beyond by **extending the principles of openness to the whole research cycle**, fostering sharing and collaboration as early as possible and refocusing science *from a 'publish or perish' perspective to a knowledge-sharing perspective*.

Assumed the paradigm shift embodied by the Open Science approach and the new Open Access policy and rules given by the European Commission (EC), the (potential) beneficiaries of Horizon 2020 (H2020) funds are facing **new challenges and opportunities** as well.

Concurrently, the Open Research Data (ORD) pilot and its extension to all thematic areas of H2020 bring about new complexities for applicants to the Programme that are not sufficiently equipped with the necessary knowledge and skills to be confident in a data-driven environment and tackle the context.

In this general frame, the main purpose of the present work is **to provide guidance to researchers in meeting the H2020 program new requirements** that now calls for the participation in the ORD pilot, as a default, for all the new proposals. Special attention will be given to research institutes and universities, as the primary venues in which science and data are produced.

The study of the open access landscape and of the features of the ORD pilot, together with the analysis of the relevant obstacles that can influence the participation, is part of this work. It also aims at identifying the appropriate **resources, tools and support services to facilitate researchers and data originators in the effective participation in the pilot**, with the ambition to contribute in increasing their research efficiency and production of high quality data with potential for long-term use.

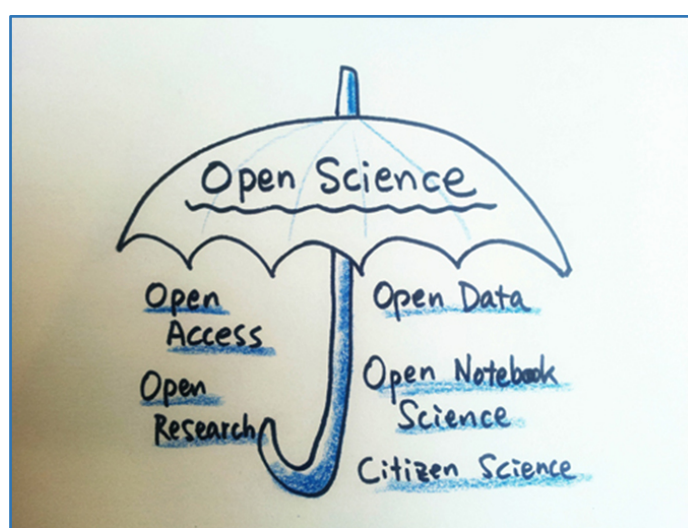


Figure 1- *The Open Science umbrella* (Image credit: Flickr user 지우 황 CC BY 2.0).

¹ European Commission (2016b). *Open innovation, Open Science, open to the world. A vision for Europe*. Brussels: European Commission, Directorate-General for Research and Innovation. ISBN: 978-92-79-57346-0 DOI: 10.2777/061652

*Open Science is also about making sure that science serves innovation and growth*² *Open access to publicly-funded research results and the possibility of knowledge sharing will encourage re-use of research output and accelerate the implementation of ideas for innovative products and services*³. The European Commission strongly believe that not only scientists will benefit from data sharing, but also citizens, either directly or indirectly through the development of products and services released from open data, and businesses as well, thanks to the marketing of new products and the creation of new jobs. Figures on the return on investments are given in the Research Data Alliance Europe Report (2014).

*Open Science is frequently defined as an ‘umbrella term’ that involves the various movements aiming at removing barriers for sharing any kind of output, resources, methods or tools, at any stage of the research process*⁴ (Fig. 1).

The European Research Council (ERC) is among the major international research performing and funding organisations playing a key role in the spread of the “open” approach. As a fundamental part of its mission, ERC supports the principle of OA to the published output of research and also promotes the basic principle of open access to research data.

Under the EU Research and Innovation funding programme Horizon 2020, a big step towards **Open Access (OA) to publications** was taken as well. All research projects receiving H2020 funding have the obligation to give open access to any peer reviewed journal articles they publish related to H2020 funding.

More recently, the Commission recognised that research data are as important as publications. Few years ago in 2012 it announced that it would experiment with open access to research data and **launched in 2014 the Pilot on Open Research Data (ORD pilot)** in H2020. *At first limited to some selected areas, as of the Work Programme 2017 the ORD pilot has been extended to cover all thematic areas of Horizon 2020 (“extended” ORD Pilot), making open research data the default setting*⁵. This move will boost competitiveness by accelerating innovation and collaboration, improving transparency, and avoiding duplication of efforts.

The pilot is currently under monitoring with a view to developing the European Commission policy on Open Research Data in future framework programmes. As clearly underlined by Neelie Kroes, former Vice-President of the European Commission for the Digital Agenda: *“We know that sharing and re-using research data holds huge potential for science, society and the economy. This Pilot is an opportunity to see how different disciplines share data in practice and to understand remaining obstacles.”* Commissioner Máire Geoghegan-Quinn said further: *“This pilot is part of our commitment to openness in Horizon 2020. I look forward to seeing the first results, which will be used to help set the course for the future.”*

While fostering the good management of the research data to optimise access, discoverability and sharing for use, re-use and impact, *the ORD pilot takes also into account the need to balance openness and protection of*

² Günther Oettinger Blog Post: “Open science for a knowledge and data-driven economy” (2015)

³ Carlos Moedas, “Open science for a knowledge and data-driven economy”, “Opening up to an ERA of Innovation” Conference. Brussels, 22 June 2015

⁴ FOSTER Project, “What is Open Science? Introduction”, <https://www.fosteropenscience.eu/content/what-open-science-introduction>

⁵ European Commission Fact Sheet, “Horizon 2020: Work Programme update supports competitiveness through open science”, [http://europa.eu/rapid/press-release MEMO-16-2604 en.htm](http://europa.eu/rapid/press-release_MEMO-16-2604_en.htm)

scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions⁶. The general Commission's approach to the pilot is '*as open as possible, as closed as necessary*' but opt-outs are possible at any stage of the project in a series of cases but the reasons have to be given and clearly addressed.

2. OPEN ACCESS

Although there are no legally binding definitions of 'access' or 'open access' in this context, the most imposing one is the definition appeared in the Budapest Declaration, a key political document that is the result of the Budapest Open Access Initiative in 2002. The Declaration represents a milestone of the Open Access movement and set forth the concept of "open access" as the *free and unrestricted online availability* of all the scientific and scholarly journal literature (peer-reviewed articles and any unreviewed preprints as well) *without financial, legal, or technical barriers*⁷. Later the definition of OA given within the Berlin Declaration (2003) promoted "*the Internet as a functional instrument for a global scientific knowledge base*"⁸.

While the right to read, download and print are basic elements, the right to copy, distribute, search, link, crawl and mine are also included as novel features under both definitions.

According to the EC policy, *Open Access generally refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable*⁹. It is important to know that in the context of research and innovation, 'scientific information' means:

1. peer-reviewed scientific research articles (published in scholarly journals) and
2. research data (all data underlying publications and raw or curated data).

According the Europe 2020 Strategy for a smart, sustainable and inclusive economy, knowledge and innovation are the means that lead to growth. In this context, OA can play a very important role:

- ✓ Research results (publications and data), and in particular those results coming from publicly-funded research, circulate more rapidly and widely, using digital media. This accelerates dissemination of knowledge and scientific discovery, and enables new forms of data-intensive research.
- ✓ The OA approach also allows the research findings to be systematically taken up by European business and industry and reach the market faster, thus accelerating innovation and leading to faster growth.
- ✓ In addition, OA stimulates competition among traditional scientific publishers and encourage the development of new not-for-profit scientific publishers, distributed academic networks and research libraries.

⁶ DG Research & Innovation, "Open Access", <http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess>

⁷ <http://www.budapestopenaccessinitiative.org>

⁸ <https://openaccess.mpg.de/Berlin-Declaration>

⁹ H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

2.1 Peer-reviewed scientific research articles

The European Commission has been developing policy and measures on OA since 2006. In 2008, it was launched a pilot on OA covering 20% of the FP7 budget in seven research areas. The pilot proved effective and promoted OA to publications through the implementation of the 'Gold' and 'Green' routs. At the same time, EC also funded a specific project to support the implementation of OA in Europe, by providing an infrastructure of technical and human resources (i.e. OpenAIRE project - Open Access Infrastructure for Research in Europe).

OA approach has been also extended to Horizon 2020. Within this framework programme, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results, which means that any scientific publications can be read online, downloaded and printed by everybody. Besides journal articles, grant beneficiaries in H2020 are also encouraged to provide open access to other types of scientific publications: monographs, books, conference proceedings, grey literature (research materials produced outside the traditional academic publishing and distribution channels, e.g. reports, working papers, evaluations...).

As is the case in FP7, H2020 mandates include two main and non-mutually exclusive routes towards open access to publications.

The Commission believes that both routes are valid and complementary approaches for open access to be effective, fair, affordable, competitive and sustainable for researchers and innovative businesses:

- A. *“Self-archiving ('green' open access): the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.*
- B. *Open access publishing ('gold' open access): an article is immediately published in open access mode. In this case, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors. These costs, often referred to as Article Processing Charges (APCs) are usually borne by the researcher's university or research institute or the agency funding the research.¹⁰ In European funded project, APCs are eligible for reimbursement during the duration of the project.*

Among the EC guidelines on OA, given in the framework of H2020, it is worth noting that:

- open access publications go through the same peer review process as non-open access publications;
- open access is not an obligation but becomes an issue only if publication is chosen as a means of dissemination and it is up to researchers whether they want to publish some results or not;
- open access does not interfere with the commercial exploitation of research results, since the decision on whether to commercially exploit results (e.g. through patents or otherwise) is made before the decision to publish (open access or not).

¹⁰ Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Options in open access and dissemination of data in H2020 are explained in figure 2. More information on the different aspects of knowledge dissemination routes are given in the European IPR Helpdesk factsheet "Publishing vs. patenting" ¹¹.

The workshop titled "Alternative Open Access Publishing Models: Exploring New Territories in Scholarly Communication" (EC DG-CONNECT, October 2015) ¹², was aimed to collect information about the "alternatives" to Green and Gold OA. Different OA publishing models, of variable ages and some of proven success and sustainability, were presented during the workshop demonstrating that there is no shortage of ideas on how to achieve or move towards open access. The Commission therefore considered exploring the next steps to take, aware that the move to open access is an open process that should hear from many stakeholders (researchers, funders, publishers, research users and librarians) and that it is therefore hoping to continue collecting views and possibly provide start-up capitals to help alternative OA publishing to become established and sustainable.

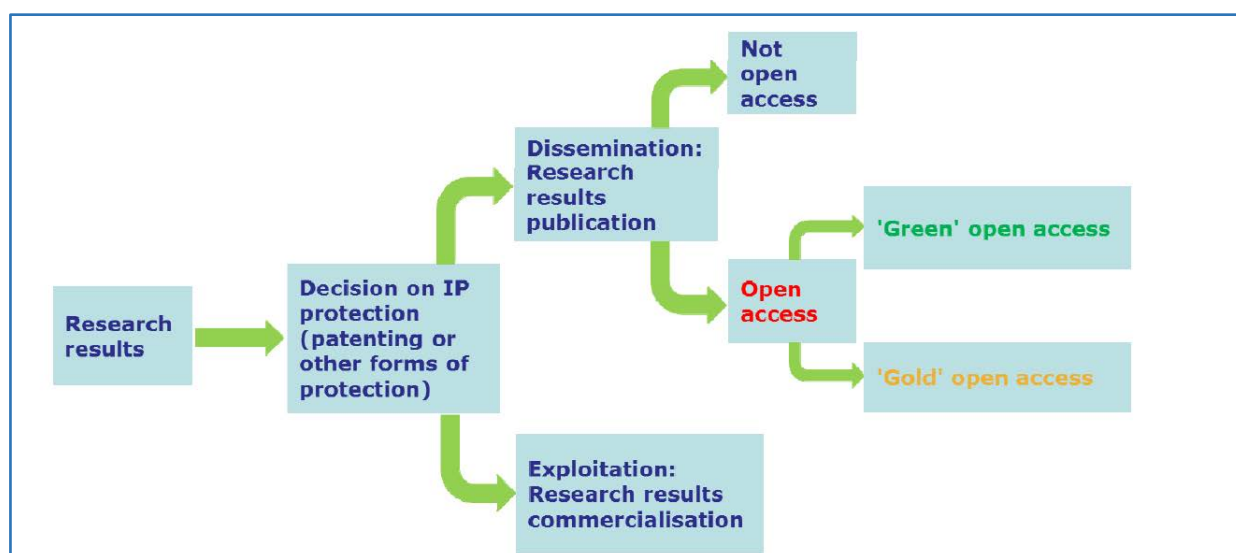


Figure 3 - Options in open access and dissemination of data in Horizon2020 (source: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

2.2 Open access to research data

OA to research data refers to the right to access and reuse research data in digital form. Terms and conditions within the H2020 programme are set out in the Grant Agreement (Art. 29 Dissemination of results - Open Access - Visibility of EU funding) ¹³.

In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

¹¹ <https://www.iprhelphdesk.eu/sites/default/files/newsdocuments/Fact-Sheet-Publishing-v-Patenting.pdf>

¹² https://ec.europa.eu/futurium/en/system/files/ged/oa_report.pdf

¹³ H2020 Programme AGA - Annotated Model Grant Agreement http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

2.3 Legal issues on OA

Numerous legal issues must be considered when implementing OA strategies or operating OA repositories and journals. Legal issues in OA includes intellectual property rights, data protection obligations and legislation. Many of the legal obligations to which stakeholders are subject are sometimes in conflict, specifically, intellectual property, privacy, data protection and open access mandates create complex and contradictory obligations which stakeholders must navigate in creative ways.

For example, even though some publishers allow their authors to self-archive their articles in institutional repositories or on their own personal websites, conditions and restrictions are frequently imposed to authors that are often obliged to observe an extended embargo period between the publication date and the date on which the document is made openly accessible online. Often this period exceeds the H2020 embargo requirements, that is 6 months for Science, Technology, Engineering and Mathematics (STEM) and 12 months the Social Sciences and Humanities (SSH).

When self-archiving is not allowed, authors can supplement their standard publishing agreements with contract addenda which permit to provide open access to their work in parallel with publication. Such addenda can also be used to support embargo period time-lapse.

Besides, repository managers must pay particular attention to the liability risks involved in the operation of a repository.

Moreover, data protection aspects play a very important role when it comes to OA to research data. A strict regime should apply to the collection, processing, and storage of personal data, which, as a rule, makes transmission conditional upon the previous, voluntary, and informed consent of the data subject. These issues must be always taken into account during the preparation phase of a project – that is, when designing the way in which data are to be collected.

2.4 Ethical issues on OA

Open access to research data can raise several ethical concerns that are much more important in medical research and publication as they directly affect the suffering humanity.

Publishing data without restrictions may in some cases conflict with established principles of ethical research, including respect for the autonomy of individuals, justice and beneficence. *Various disciplines have formalised such principles in codes of ethics, to properly inform about the nature of the research and how to treat data confidentially. Failing to meet such ethical standards may not only cause harm to research participants, it can also be detrimental to the scientific enterprise or society*¹⁴.

Some of the ethical concerns that sharing data can generate are unintended secondary uses and misappropriation, dual use, violations of privacy and confidentiality, unequal distribution of research results, commercialisation, restriction of scientific freedom.

¹⁴ RECODE Project “Policy RECommendations for Open access to research Data in Europe”, http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf

3. OPEN RESEARCH DATA PILOT

The recognition that research data are as important as publications led the EC in launch the **Pilot on Open Research Data (ORD pilot)** in 2014. At first limited to some selected areas, from the start of the 2017 Work Programme, the ORD pilot has been extended to cover all thematic areas of Horizon 2020, *making open research data the default setting*¹⁵. However, since not all instruments are suitable for research data sharing, some instruments have been excluded from the pilot. They are:

- "co-fund" and "prizes" instruments;
- "ERC proof of concept" grants;
- "ERA-Nets" that do not produce data;
- SME instrument, phase 1.

The ORD pilot applies to the digital research data generated in the project and distinguished in the following two types:

- ✓ the 'underlying data', that are data needed to validate the research results, including their associated metadata, and presented in scientific publications - which should be deposited as soon as possible;
- ✓ 'other data', including their associated metadata, that they want to share in the Data Management Plan - which should be deposited within the deadlines laid down in the Plan.

The main requirements of the ORD pilot for H2020 applicants are specified in the Model Grant Agreement (Art. 29.3)¹⁶. They can be summarized as follows:

- develop (and keep updated) a **Data Management Plan**;
- deposit data in a research data **repository**;
- make sure third parties can **freely access, mine, exploit, reproduce and disseminate** it;
- make clear (via the repository) what **tools** (e.g. specialised software or software code, algorithms, analysis protocols, etc.) will be needed to use the raw data to validate research results (and make these instruments available to the beneficiaries, where possible).

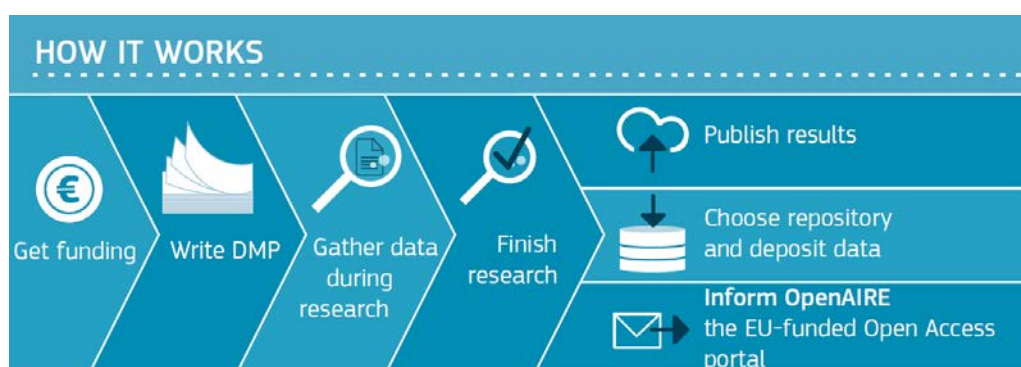


Figure 4 – Open Research Data in Horizon 2020 (source: opendata-infographic_072016.pdf by OpenAIRE)

¹⁵ European Commission Fact Sheet, "Horizon 2020: Work Programme update supports competitiveness through open science", [http://europa.eu/rapid/press-release MEMO-16-2604_en.htm](http://europa.eu/rapid/press-release_MEMO-16-2604_en.htm)

¹⁶ H2020 Programme AGA - Annotated Model Grant Agreement http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

Costs related to the implementation of the Open Research Data pilot (e.g. costs for providing open access, related research data management costs, data curation and data storage costs) may be eligible (see Article 6.2.D.3 of the Grant Agreement).¹⁷

3.1 Data Management Plan

A Data Management Plan (DMP) is a structured document providing the framework for research data stewardship. It should describe how a research activity will handle its data, both during the project and after its completion, and explain how the data generated will be collected and managed in the present and prepared for preservation and their re-use in the future.

A DMP is required for all projects participating in the “extended” ORD pilot, unless they opt out of the pilot. However, projects that eventually opt out are still encouraged by EC to submit a DMP on a voluntary basis. Developing a DMP can be a critical element in the execution of a research activity. The DMP does not have to be submitted at the proposal stage but as an early deliverable within the first six months of the project. However, the DMP is not a fixed document but evolves during the lifetime of the project. It then normally needs to be revised as the project progresses and when significant changes occur. The Plan can be updated even after the project has finished, in order to be calibrated to the data generated and to the uses identified by the consortium of partners.

DMP represent a key element of a good data management. It should describe how the many aspects of data analysis, data management, metadata generation, and data preservation will be dealt with during the project. It should also present which data will be stored, preserved and made available for re-use. The Plan should also include elements about the ownership of the data collections, including intellectual property rights if relevant, and clarify responsibilities and roles in the data management processes among the relevant parties in the research.

EC provides data management guidelines for applicants¹⁸ to reference as they write their DMPs, including a template called “FAIR DMP template” that is composed of a set of questions that the researchers should answer with a level of detail appropriate to the project. The EC template is not intended as a strict technical implementation of the FAIR principles¹⁹ but it is rather inspired by FAIR as a general concept. According to the “EC guidelines on FAIR Data Management in Horizon 2020”, the structure of a DMP should firstly address to FAIR principles (“*make data Findable, Accessible, Interoperable and Re-usable*”) (fig. 4) and cover all the following sections:

1. Data Summary
2. FAIR data
 - 2.1 Making data Findable, including provisions for metadata

¹⁷ H2020 Programme AGA - Annotated Model Grant Agreement
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

¹⁸ Guidelines on FAIR Data Management in Horizon 2020
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

¹⁹ Force11 Discussion Forum - <https://www.force11.org/fairprinciples>

- 2.2 Making data openly Accessible
- 2.3 Making data Interoperable
- 2.4 Increase data Re-use
- 3. Allocation of resources
- 4. Data security
- 5. Ethical aspects
- 6. Other issues

Developing a DMP is not only needed to comply with the EC recommendations but has undeniable (short- and long-term) benefits for researchers and institutions. Among those there are:

- definition of the framework and resources to support research data, which translates to better time management and lower costs during the project;
- avoiding unnecessary duplication (e.g. re-collecting or re-working data);
- continuity if project staff leave or new researchers join;
- allowing for the validation of the research results, as data underlying publications is maintained;
- ensuring the research activity is more visible, citable and has greater impact;
- increase in collaborations and advancements in research that come from the sharing of well organised, well documented and preserved data.

The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Figure 5 – The FAIR Guiding Principles (Wilkinson et al., 2016)

3.2 Data repository

Selecting a data repository and deposit research data is one of the practical steps to be taken when participating in the ORD pilot. A data repository is a digital archive collecting, preserving and displaying datasets, related documentation and their metadata in the long term. In such context, terms like “preservation” and “curation”

are used rather than “archiving” or “storage”: long-term accessibility implies expertise and services to convert data to new formats and to add value to the data, for instance by new functionality to query the data.

It is advisable to contact the repository when preparing the first version of the DMP. Repository managers may offer guidelines for sustainable data formats and metadata standards, as well as support for dealing with sensitive data and licensing. However, when data preservation standards or norms exist in the participant’s discipline, these should be followed.

The permanent access to research data is a challenge for all stakeholders in the scientific community. The new awareness that the long-term preservation and the principles of OA to research data can offer broad opportunities for the scientific community is leading more and more research centres and universities in building research data repositories that allow permanent access to data sets in a trustworthy environment. Due to disciplinary requirements, the landscape of data repositories is very heterogeneous. Thus, it is difficult for researchers, funding bodies, publishers and scholarly institutions to select appropriate repositories for storage and consultation of research data.

According to the of the OpenAIRE recommendations ²⁰, when submitting data it is suggested to (in order of preference):

1. *Use an external data archive or repository already established for your research domain to preserve the data according to recognised standards in your discipline.*
2. *If available, use an institutional research data repository, or your research group’s established data management facilities.*
3. *Use a cost-free data repository such as Zenodo.*
4. *Search for other research data repositories (see <http://re3data.org/>).*

In any case, preference should be given to trusted and “certified” repositories that are relevant to the scientific communities and have obtained accreditation or certification to appropriate standards. This ensures the quality of data repositories and the effectiveness of the quality assurance process.

3.3 Licenses

In order to permit the widest possible re-use, appropriate licenses should also be applied to research data when deposited. Licences are the most effective ways of communicating permissions to potential re-users of data. In this context, *a licence is a legal instrument for a rights holder to permit a second party to do things that would otherwise infringe on the rights held. It is important to note that only the rights holder (or someone with a right or licence to act on their behalf) can grant a licence. It is therefore imperative that the intellectual property rights (IPR) pertaining to the data are established before any licensing takes place.* In any cases, *licences can be terminated only by expiry of the licensor’s IPR or, for a particular licensee, through breach of terms.* ²¹

²⁰ OpenAIRE Research Data Management Briefing Paper (April 2017)
<https://www.openaire.eu/public-documents?id=808&task=document.viewdoc>

²¹ Alex Ball, How to License Research Data (Digital Curation Centre, 2014)

Since the data should be “open”, the EC's policies point towards using Creative Commons Licenses²². Creative Commons licences give the creators of the works control over how they may be used, than simply declaring them public domain or reserving all rights.

Within the Pilot, the following licenses are suggested to be attached to the data deposited:

- Creative Commons Zero (CC-0)²³: it is for dedicating the works to the public domain since all copyright interests are waived. *It works as an irrevocable, royalty-free and unconditional licence for anyone to use the work for any purpose. The rights waived include database rights, so CC0 is suitable for use with data.*¹⁶
- Creative Commons Attribution (CC-BY)²⁴: it permits to anyone to *share (copy and redistribute the material in any medium or format) and adapt it (remix, transform, and build upon the material for any purpose, even commercially)*. The creator must *give appropriate credit, provide a link to the license, and indicate if changes were made*.

3.4 Opting out – partially or entirely

Participating in the ORD pilot does not necessarily mean opening up all research data. The EC principle about data is “*as open as possible, as closed as necessary*”. Projects can therefore opt out at any stage (before or after the signature of the grant agreement) but reasons have to be clearly given. Equally, H2020 proposals are not evaluated more positively because they have agreed to share their research data and are not penalised if they opt-out as well.

Opts out are possible in a series of cases that include²⁵:

- a) participation is incompatible with the obligation to protect results (to be commercially or industrially exploited)
- b) participation is incompatible with the need for confidentiality (concerning security issues)
- c) participation is incompatible with rules on protecting personal data
- d) participation would jeopardise the project objectives
- e) the project will not generate / collect any research data
- f) other legitimate reasons.

A beneficiary in multi-beneficiary project can keep its data closed if specific arrangements are made in the consortium agreement and those are conforming with the reasons for opting out.

²² <http://creativecommons.org/>

²³ <https://creativecommons.org/publicdomain/zero/1.0/>

²⁴ <https://creativecommons.org/licenses/by/3.0/>

²⁵ Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

3.5 Overview about the ORD up-take

As result of the EC monitoring, the Directorate-General for Research and Innovation provided an overview of the up-take of the pilot in the H2020 calls for the two periods: from March 2014 to July 2015 ²⁶ and from August 2015 to August 2016 ²⁷.

Considering the numbers of proposals and finalised grant agreements for both periods ²⁸, table 1 below provides an overview of the overall participation, opt-outs and opt-in rates in the ORD pilot so far.

| Category | Mar 2014- Jul 2015 dataset | Aug 2015-Aug 2016 dataset | 2014-2016 Average |
|--|----------------------------|---------------------------|-------------------|
| Participation rate (core areas ²⁹) | 65,4% | 64,26% | 64,83% |
| Opt-outs (core areas ²⁹) | 34,6% | 35,74% | 35,17% |
| Opt-ins (all other areas of H2020) | 11,9% | 17,05% | 14,47% |

Table 1 – Participation, opt-out and opt-in rates and average for the ORD pilot (March 2014 – August 2016) (source: “Open access: the uptake of the Open Research Data pilot in Horizon 2020. Explanatory note to the updated 2016 dataset”, 15/12/2016)

As can be seen, for participation rate and opt-outs, the figures for both periods are very close. An average of 64,83% of projects in the core areas participated in the pilot and an average of 35,17% of proposal/project in the core areas opt-out. Conversely, for opt-ins an increase of 5,15 % can be observed in the second period.

The survey also gives information about reasons for opt-outing (at proposal stage) and the figures. It appears that the most frequent opt-out reasons are (in descending order): IPR concerns, privacy concerns, no research data generated, other legitimate reasons, main project objective jeopardised, and national security related confidentiality. Graphic 1 below shows the results for both periods and provides the total average. Compared to the first year's sample, there is a marked increase for IPR concerns in the second one. “Privacy concerns” remain the second most important reasons for opting out, followed by "no research data generated". It is also worth noting that as for the voluntary participation in calls not covered by the pilot, the highest numbers of opt-ins for the period 2015-2016 was observed in the following calls:

- Blue Growth: Unlocking the potential of Seas and Oceans
- Earth Observation-2015-LEIT SPACE
- Energy efficient buildings.

It should be noted that in some calls – such as the European Researchers' Night - an opt-out is practically inevitable, since this action itself is not concerned with generating research data. This means that 100% participation in the pilot is not possible or even desirable (since not all projects generate data).

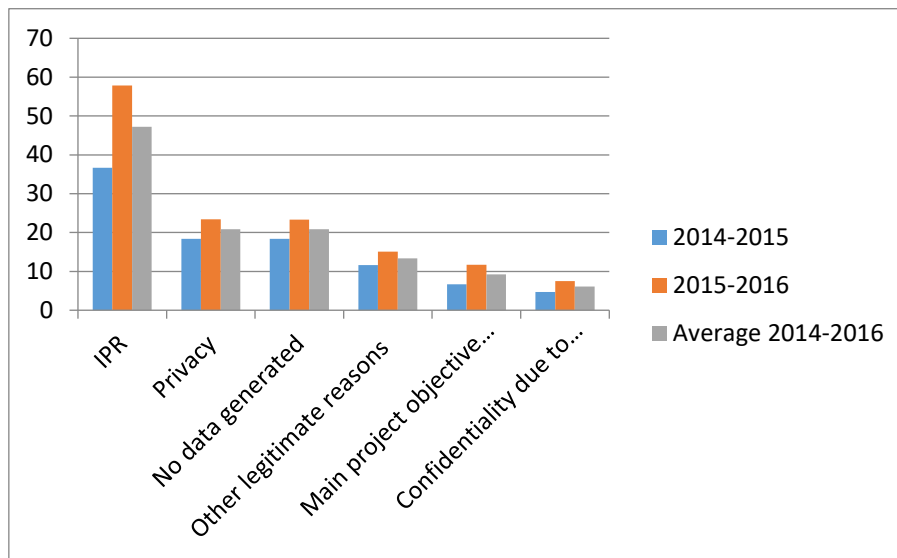
²⁶ 2014-2015 dataset available at <https://data.europa.eu/euodp/en/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020/resource/ee828747-5930-48ef-af3f-a51a647d870e>

²⁷ 2015-2016 dataset available at: the uptake of the Open Research Data pilot in Horizon 2020. Explanatory note to the updated 2016 dataset, <https://data.europa.eu/euodp/en/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020/resource/7fb899dd-6859-4361-9c30-70b66773592e>

²⁸ The following instruments are excluded from the survey: SME instrument, Cofund, and Prizes, ERC grants.

²⁹ For the complete list of the “core areas” of the Horizon 2020 Work Programme 2014-15 see: “OA factsheet” at https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Open_Access.pdf

Absolute numbers (rather than percentages) shows that, in both considered periods, the call with the most participants is the Marie Curie Individual Fellowship with an average of 325 projects participating in the pilot. Even though data showed that nearly two thirds of the participants stayed in the ORD pilot and one third opted out, the “extended” ORD pilot now covers scientific areas where data sharing is not so widespread or might prove more difficult. For this reasons, EC expects that there may be an increase in opt-outs. However reliable information on this will only be available in the coming years.



Graphic 1 - **Reasons for opting out** (source: “Open access: the uptake of the Open Research Data pilot in Horizon 2020. Explanatory note to the updated 2016 dataset”, 15/12/2016)

4. PROJECTS SUPPORTING OPEN ACCESS

Open access means giving free access to scientific publications and data, that can be used by other researchers for scientific purposes; however, making research data open is a time consuming activity that requires specific rules to be followed. A good management of research results is essential to ensure that they will be preserved, accessible and reused by other researchers.

Many projects founded by the European Commission, assist stakeholders on open access issues by providing guidelines on how to comply with national or institutional policies, indications on where to deposit and how to licence research results. Researchers can also find training materials and toolkits that supports the implementation of data management plans. They have also the opportunity to get assistance on OA from several help desks that provide specific support for conforming to the EC Open Access policies.

EDISON (2015-2017)

<http://edison-project.eu/>

The EDISON project focuses on activities to establish the new profession of 'Data Scientist', following the emergence of Data Science technologies (also referred to as Data Intensive or Big Data technologies) which changes the way research is done, how scientists think and how the research data are used and shared. This

includes definition of the required skills, competences framework/profile, corresponding Body of Knowledge and model curriculum. It has developed a sustainability/business model to ensure a sustainable increase of Data Scientists, graduated from universities and trained by other professional education and training institutions in Europe.

EOSCpilot - European Open Science Cloud for Research pilot project

<https://eoscpilot.eu/about-eoscpilot>

The EOSCpilot project will support the first phase in the development of the European Open Science Cloud (EOSC) that aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market. EOSC will enable trusted access to services and will offer to researchers and science and technology professionals a virtual environment (globally interoperable and accessible) to store, share and re-use their data across disciplines and borders.

EUDAT (2011-2015)

<https://www.eudat.eu>

EUDAT was a three-year project that delivered a Collaborative Data Infrastructure (CDI) with the capacity and capability for meeting future researchers' needs in a sustainable way. Its design reflect a comprehensive picture of the data service requirements of the research communities in Europe and beyond. This will become increasingly important over the next decade as we face the challenges of massive expansion in the volume of data being generated and to be preserved and accessed too. EUDAT also provides the opportunity for data-sharing between disciplines and cross-fertilisation of ideas.

FOSTER

<https://www.fosteropenscience.eu/>

FOSTER (2014-2016)

FOSTER was a coordination initiative that aimed at supporting different stakeholders, especially young researchers, in adopting Open Access in the context of the European Research Area (ERA) and in complying with the open access policies and rules of participation set out for Horizon 2020. FOSTER established a European-wide training programme on open access and open data, consolidating training activities at downstream level and reaching diverse disciplinary communities and countries in the ERA. Each type of stakeholder was provided with a range of relevant training toolkits, practical advice, support and help in engaging, dynamic and outcome-oriented way.

FOSTER PLUS (2017-2019)

The FOSTER PLUS project will focus on promoting the practical implementation of Open Science, with activities targeting academic staff, young scientists and policy-makers in particular. Building on the existing

FOSTER portal and training materials, FOSTER PLUS will develop more advanced level and discipline-specific materials that build capacity for the practical adoption of Open Science and promote a change in culture. FOSTER PLUS will develop a multi-module Open Science Toolkit, covering key topics such as responsible research and innovation, research data management, software carpentry, text and data mining, reproducible research and open peer review.

LEARN (2015-2017)

<http://learn-rdm.eu>

The aim of the project was to take the LERU Roadmap for Research Data, produced by the League of European Research Universities (LERU), and develop it in order to build a coordinated e-infrastructure across Europe and beyond. LEARN delivered a model of Research Data Management (RDM) policy, a Toolkit of Best Practice for Research Data Management and an Executive Briefing in five core languages so as to ensure wide outreach.

OpenAIRE

<https://www.openaire.eu/>

OpenAIRE (2009-2011)

The aim of the project was to deliver "an electronic infrastructure and supporting mechanisms for the identification, deposition, access, and monitoring of FP7 and ERC funded articles", where the main supporting mechanism was establishment and operation of the European Helpdesk System. Additionally, it offered "a special repository for articles that can be stored neither in institutional nor in subject-based/thematic repositories" and created a new portal to make all deposited articles and data freely accessible worldwide.

OpenAIRE Plus (2012-2014)

OpenAIREplus created a 2nd-Generation Open Access Infrastructure by significantly expanding in several directions the outcomes of the OpenAIRE project. OpenAIREplus developed an open access, participatory infrastructure for scientific information that significantly expanded its base of harvested publications to also include all OA publications. Deposited articles became openly accessible through an enhanced version of the OpenAIRE portal.

OpenAIRE 2020 (2015-2018)

OpenAIRE2020 represents a pivotal phase in the long-term effort to implement and strengthen the impact of the OA policies of EC, building on the achievements of the previous OpenAIRE projects. OpenAIRE2020 will expand and leverage its focus from the agents and resources of scholarly communication to workflows and processes, from publications to data, software, and other research outputs, and the links between them, and strengthen the relationship of European OA infrastructures with other regions of the world, in particular Latin

America and the U.S. Through these efforts, OpenAIRE2020 will truly support and accelerate Open Science and Scholarship, of which Open Access is of fundamental importance.

PASTEUR4OA (2014-2016)

<http://www.pasteur4oa.eu/>

The aim of the project was to encourage the development of matching policies on open access and open data in the European Union, according to the European Commission Recommendation on “Access to and preservation of scientific information” and in view of maximizing alignment with the H2020 policy on access to the research funded by the Commission. The project helped to develop and reinforce open access strategies and policies at the national level and facilitate their coordination among all Member States by creating a network of centres of expertise that developed a coordinated and collaborative programme of activities in support of policymaking at the national level under the direction of project partners.

RECODE (2013-2015)

<http://recodeproject.eu/>

The aim of RECODE project was to leverage existing networks, communities and projects to address challenges within the open access and data, dissemination and preservation, sector. The sector includes several different networks, initiatives, projects and communities that are fragmented by discipline, geography, stakeholder category (publishers, academics, repositories, etc.) as well as other boundaries. Many of these organizations are addressing the barriers to open access to research data, such as stakeholder fragmentation, technical and infrastructural issues, ethical and legal issues, and state and institutional policy fragmentation. However, these organizations are often working in isolation or with limited contact with one another. RECODE provided a space for European stakeholders in the open access and data dissemination and preservation sector to work together to find common solutions for these issues. The RECODE project conducted studies of good practice and exchange good practice principles and formulated recommendations for open access to research data targeted at different stakeholders and policy-makers in support of the Commission’s policies.

5. BARRIERS AND ENABLERS TO THE IMPLEMENTATION OF THE ORD

5.1 Barriers and resistances to implementation

Many researchers are not used to share data of their research and disseminating publications outside of their team due to a number of reasons. Most of these reasons are related with a lack of education/information about the OA issues or, often, of an erroneous data sharing culture among researchers and data managers.

The ERAC Report “Opinion on Open Research Data”³⁰ highlights the most common arguments that researchers use to justify their behaviour. Interesting concerns also come out from the NordForsk policy paper

³⁰ ERAC Task Force on Open Access to Research Data, “Opinion on Open Research Data” Report, ERAC 1202/16

“Open Access to Research Data – Status, Issues and Outlook”³¹. Here below a short list is presented that includes both personal opinion and resistances of researchers and data originators and actual barriers, even though the border between the two is not always detectable:

- ✓ Fear that reanalysis of datasets could **reveal errors** or produce alternative conclusions;
- ✓ Feeling that data management is a particularly **time/money consuming activity** that distracts researchers from R&D activities;
- ✓ **Lack of data infrastructures and of standards;**
- ✓ **Anxiety of competition and of being “scooped”** (and therefore to have reduced publication opportunities);
- ✓ **Fear of losing the competitive advantage** when sharing data and then to miss the chance to attract students and researchers;
- ✓ **Lack of dedicated funding** to prepare data and documentation for sharing;
- ✓ **Absence of professional rewards** for data sharing;
- ✓ **Ethical and legal constraints;**
- ✓ Sharing of data not allowed due to **non-disclosure agreements** with private partners;
- ✓ Reluctance in sharing **data that are considered private;**
- ✓ Assumption that **discipline specific data cannot be reused** in different research contexts;
- ✓ **Lack of evidence about the deriving benefits of data sharing** for the researcher and the research institute.

5.2 Advantages and enablers

Although, it is recognised that open research to data can offer many **opportunities** at scientific, economic and societal levels, as a fundamental part of the future of Open Science, many researchers and data managers are not really aware of the **advantages** and real benefits coming from a broader access to scientific data and publications.

Open Access to scientific data mainly helps to:

- ✓ Have **greater impact and visibility** of research;
- ✓ Encourage **collaboration between data users and data creators;**
- ✓ **Avoid duplication of efforts** (better efficiency and reduction of costs);
- ✓ Enable **improved quality research;**
- ✓ **Speed up innovation** (faster progress to market that means **faster growth**);
- ✓ Improve transparency of the scientific process and the wider **involvement of citizens and society.**

Furthermore, *participating in the ORD Pilot means to contribute in co-shaping European policy on opening up research data and be part of the new era of Open Science*³².

³¹ NordForsk Policy Paper 1. “Open Access to Research Data – Status, Issues and Outlook” (2016)

³² OpenAIRE FactSheets “Open Research Data Pilot in Horizon 2020”

Along with the (perceived or real) barriers to data sharing and re-use, the NordForsk Policy Paper ³³ also provide a lists of enablers for data sharing, resulting from researchers interviews, that is reported here below:

- data sharing expectations of funders and journals;
- peer expectations and sharing practices in the research community;
- availability of data repositories and standards;
- desire to showcase data quality;
- researchers' data management skills;
- organisational support;
- acknowledgement received for data sharing;
- data publication and metrics.

This analysis is useful since it provides indications of where real intervene might be needed and gives hints on how to further improve the sharing and re-use of research results.

6. METHODOLOGY

In order to understand the difficulties faced by researchers and data managers when approaching OA to Research Data and data management, structured interviews have been used to collect opinions, remarks from representatives of different research and academic institutions, participating in the ORD pilot. The main objective was to have feedback about their experience in implementing the pilot, with special reference to the problems they faced and the solutions and the tools they experimented in data management issues.

It is well recognised that a structured approach to data management is mandatory for performing state of art research and that a sound data management has real advantages and benefits. Managing research data throughout its lifecycle ensures its long-term value and prevents data from falling into digital obsolescence. A good and rigorous data management also greatly facilitates the re-use of research data (with the consequence of increasing the impact of research results) and thus saving precious research funds and avoiding unnecessary duplication of work. Proper data management is also a key prerequisite for effective data sharing throughout the scientific community.

Under these premises, the EU policy on Open Research focused on the Data Management Plan (DMP) as a central element of the ORD pilot, as the best way of ensuring that a serious and organised approach to data management is followed.

In the formulation of a DMP a data management checklist can be the leading guide. This is why we chose the "ETH-EPFL Data Management Checklist" as a helpful starting point in our study to get in touch with different parties dealing with scientific research in Europe.

This checklist, created in close collaboration between the Digital Curation Office at ETH-Bibliothek and the

³³ NordForsk Policy Paper 1, "Open Access to Research Data – Status, Issues and Outlook" (2016)

Research Data Team at EPFL Library ³⁴, is one of the deliverables of the project on Research Data Life-Cycle Management in swissuniversities' program on scientific information. It is based on existing international guidelines on research data lifecycle, but it is flexible enough to be applied to requirements from different funding organisations. The different sections of the list cover general planning and the phases of the data life-cycle ³⁵, from data collection and creation to data sharing and long-term management. Other specific sections cover documentation and metadata, file formats, storage, ethical and intellectual property issues as well.

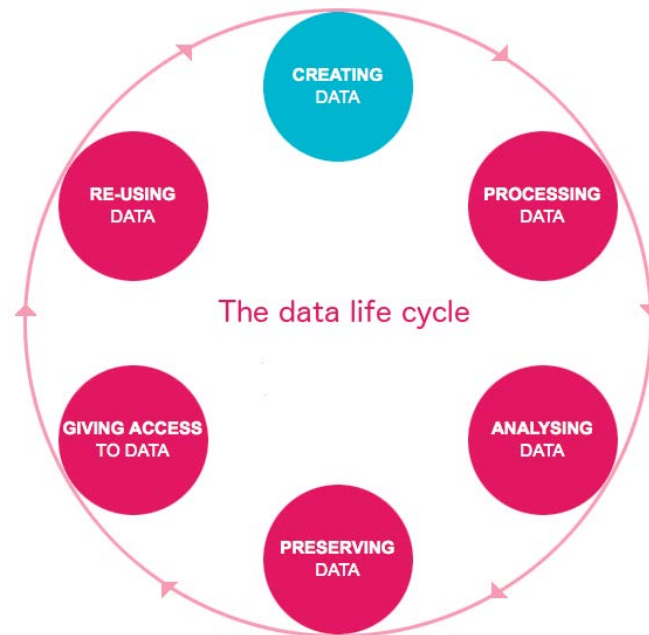


Figure 6 - Data life cycle (according to UK Data Archives)

The “ETH-EPFL Data Management Checklist” was used to get ideas and contents in preparing a short “questionnaire” (Questionnaire A, see Annex 1) that was sent (through emails) to selected relevant stakeholders (within national/international institutions) directly involved in one or more of the activities listed in table 2. It was specifically asked in which section (or phase of the data life-cycle) of the “ETH-EPFL Data Management Checklist” they were able give their input or provide services.

The other questionnaire (Questionnaire B, see Annex 2) was addressed to coordinators of H2020 research project participating in the ORD pilot, in order to get feedback on their experience in meeting the pilot requirements and dealing with data management issues, and to identify the problems they faced and the solutions/tools founded.

| | Activities | Questionnaire |
|-----------|---|----------------------|
| a) | Partnership in H2020 research project participating in the ORD pilot | A |
| b) | Data management | A |
| c) | Management of data repositories | A |
| d) | Support on OA | A |
| e) | Coordination of H2020 research project participating in the ORD pilot | B |

Table 2 – List of activities performed by the contacted stakeholders with reference to the ORD pilot

³⁴ <http://www.library.ethz.ch/en/About-us/News-archiv/News-2016/Checklist-Data-Management-for-research-projects>

³⁵ www.data-archive.ac.uk/create-manage/life-cycle

In many cases, the questionnaires have been integrated with in-depth questions and remote calls.

Among those contacted, the following parties gave feedback about their experience within the pilot and provided further insights (table 3).

| ORGANIZATION | CONTACT | ROLE | ACTIVITY |
|---|--|---|-----------------|
| APRE Agenzia per la Promozione della Ricerca Europea | M. DI ROSA | EU Horizon 2020 – National Contact Point SC5 and SC2 | d) |
| Consorzio Interuniversitario CINECA | C. CACCIARI | Distributed Computing Expert | a) b) c) |
| DANS – Data Archiving and Networked Services | M. GROOTVELD | Senior Policy Officer | a) b) c) |
| DCC – Digital Curation Center | M. DONNELLY | Senior Institutional Support Office, External Reviewer for DMPs in H2020 projects | b) d) |
| Directorate-General for Research and Innovation | OPEN ACCESS TEAM | Support on OA issues | d) |
| GFZ German Research Centre for Geosciences | K. ELGER | Data Scientist | b) c) |
| OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale | A. CRISE | Senior Scientist | a) |
| OpenAIRE NOAD (National Open Access Desk) | P. GARGIULO | OA projects, Training & Support Coordinator | d) |
| CNR Consiglio Nazionale delle Ricerche - IMAA | L. MONA D. CASTELLI | Coordinators and data managers of H2020 projects participating in the ORD pilot | b) c) e) |
| Università di Bologna, Alma Mater Studiorum | E. CICOGNANI D. FRASCARI/G. ZANAROLI E. MOLLONA T. GALLINA TOSCHI | | e) |
| Università di Roma Sapienza | C. SCUDERI | | e) |

Table 3 – Affiliation and roles of the respondents.

Profiles introducing each organization and its main fields of interest are reported in Annex 3.

The answers to the questionnaires and the follow-ups have been grouped by topic and are presented in the table below. When available, suggested solutions/supporting actions are indicated as well.

| PROBLEMS FACED IN PARTICIPATING IN THE ORD PILOT | | SUGGESTED SOLUTIONS and/or SUPPORTING ACTIONS |
|---|--|--|
| Data management | Confusion among applicants about “data management” and “Open Access” | Establishment of training courses differentiated for researchers and support staff , addressing the Open Access and Open Data issues and clarifying the requirements of the pilot and about how to manage publications and research data. |
| | Wrong assumptions among applicants on “data management” and ORD pilot requirements | |
| | The activities required by the pilot are considered too time-consuming | Development of specific technical skills and (when possible) internal resources devoted to data management |
| | Complexities in fulfilling the EC requirements on data management within the ORD pilot | |
| | More difficulties for beginners in data management | Improvement of the team DM technical skills and/or engagement of partners expert in DM |
| | Lack of discipline-specific and widely adopted practices for managing data | Look for synergic relations with experienced institutes |
| | Difficulties for scientists to adhere to standards when collecting their data and metadata | Improve the timely involvement of dedicate data managers |
| | Dealing with sensitive data and licensing | Work together with the DM/IPR experts and figure out the appropriate licence for specific datasets |
| | Significant amount of person/months to be dedicated to OA issues | Early and seized planning of resources to be devoted to the activities related to the ORD pilot |
| | Onerous work in managing data collected , in making them FAIR and guaranteeing the accessibility to data after the end of the project | Identify sustainable and shared solutions that guarantee accessibility, interoperability and reusability of data across time |
| Data policy | Differences in data policies, licences and legislations among partners | Make those information explicit outside organizations can help in understanding them and foster harmonizing procedures |
| | Strict data management policy requested by some data provider not motivated by any objective reason | Education and training of researchers about the meaning of "openness" and the benefits of opening up research data |
| | Cautionary attitude of data providers (some requesting an embargo period) | |
| | Possible conflicts arising between the data owner and the repository provider. | Use of comprehensive agreement between the data owner and the repository provider and clear attribution of responsibilities from the beginning. Careful definition of policies about data integrity and update. |

| | | |
|-----------------------------------|---|--|
| Data Management Plan (DMP) | The wrong belief that drafting a DMP is unnecessary and it is a waste of time | Training and advocacy to researchers on the advantages of a good data management as an essential element of research best practices |
| | Poor standardisation in DMP's structure and content | Use of a well thought-out standardised DMP template that facilitates the smooth data journey from in-project storage to long-term preservation |
| | Limited incentives (few real penalties for poor DMP, and few rewards for good DMP) | In the European context, the DMP is still quite new and the researchers should get used to the idea of creating and maintaining DMPs. The implementation of a reward/penalty policy by EC should help. |

Table 4 – Synthesis of the feedback provided (grouped by topic) and suggested solutions/supporting actions.

It can be observed that responses by stakeholders confirm from one side the criticalities already reported by literature (See chapter 5. Barriers and enablers to the implementation of the ORD pilot) and from the other side highlight a set of problems related to specific aspects for the implementation of the pilot.

7. TOOLS SUPPORTING THE ORD PILOT

The participation in the ORD pilot is a demanding activity that entails the knowledge of the H2020 requirements on the pilot and EC policy on open research data, as well as specific technical skills on data management. As specified in the Model Grant Agreement (Article 29.3), a Data Management Plan is required for all projects participating in the pilot.

EC provides guidelines and other resources that are mainly delivered by EU-funded projects supporting OA for scientific publications and research data and providing infrastructures and services to the research community. Other resources on data management and OA issues are given by other European institutes operating in the field of Open Science. Those instruments are also suited for application on the ORD pilot context.

Within the present work, the main resources and tools supporting open research to data have been analysed and are here presented, grouped by topic. The list is not exhaustive, although it represents a helpful starting point for researchers and relevant stakeholders within research institution and universities when grapple with the OA issues, with specific reference to the participation in the ORD pilot.

7.1 Guidelines and OA resources

The purpose of these tools is to provide explanation for rules and policies on OA that are applicable to beneficiaries in Horizon 2020 funded projects, as well as to give good practices and recommendations, addressed to both researchers and librarians, regarding data management planning that can help research data to be findable, accessible, interoperable and reusable (FAIR). Fact sheet, written as “frequently asked questions” (FAQ) documents, provides answers to queries received from Horizon 2020 applicants.

7.1.1 EC guidelines and documentation

Horizon 2020

<http://ec.europa.eu/programmes/horizon2020/>

DG Research and Innovation - Open Science

<http://ec.europa.eu/research/openscience/index.cfm>

Participant Portal H2020 Online Manual

http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Guidelines on FAIR Data Management in Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

European Research Council (ERC) Guidelines on Implementation of Open Access to Scientific Publications and Research Data

https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

European IPR Helpdesk - Open Access to scientific publications and research data in Horizon 2020: Frequently Asked Questions (FAQs)

<https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Fact-Sheet-Open-Access-to-Publications-and-Data-in-H2020-FAQ.pdf>

EC Library and e-Resources Centre - Open Access Resources

http://ec.europa.eu/libraries/links_en.htm

7.1.2 Other resources

Research Data Management Briefing paper

<https://www.openaire.eu/briefpaper-rdm-infonoads>

The extensive briefing paper gives an overview of Research Data Management with practical sections about data management planning, and archiving the research data for reuse. It also gives background sections about making research data reusable, and an overview of stakeholders who are involved in data management.

Book “Managing and Sharing Research Data. A Guide to Good Practice”

<https://uk.sagepub.com/en-gb/eur/managing-and-sharing-research-data/book240297#description>

Written by experts from the UK Data Archive with over 20 years of experience, this book gives post-graduate students, researchers and research support staff the data management skills required in today changing research environment.

The book features guidance on: how to plan your research using a data management checklist, how to format and organize data, how to store and transfer data, research ethics and privacy in data sharing and intellectual property rights. It also explores topics related to data strategies for collaborative research, data publishing and citing, making use of other people’s research data, etc.

RDA 23 Things: Libraries for Research Data

<https://www.rd-alliance.org/23-things-libraries-research-data-rdas-libraries-research-data-interest-group.html>

The Research Data Alliance (RDA) provides a short overview of practical, free, online resources and tools addressed to librarians that helps incorporating research data management into their practice within research institutions.

Science-Metrix

<http://www.science-metrix.com/en/publications/our-work>

It is an independent research evaluation firm, internationally recognized leader in the assessment of science and technology (S&T) using bibliometric methods, and specialized in the evaluation of science-based programs and initiatives.

Among others, Science-Metrix makes available scientific reports and conference presentations on OA publication strategies, state of the art of Open Data and evolutions of policies on OA, commissioned by government, educational, non-profit or private organizations in and outside Europe.

7.2 Support kits

The following are toolkits, created by two of the most important projects that supports OA, that provide guidelines, factsheets, slides, case studies and best practices regarding data management.

OpenAIRE Plus Support kit for Open Research Data pilot

<https://www.openaire.eu/d4-3-support-kit-for-open-research-data-pilot>

The support kit is made up of a set of guidelines, webpages, briefing paper, factsheets, webinar recordings, a postcard, slides, video, and FAQs. The kit is addressed to a range of stakeholders (project officers, project managers, EC grantees and researchers, projects coordinators, research administrators, data repository managers and data librarians) to assist them in understanding and complying with the requirements for the Horizon 2020 Open Research Data pilot.

LEARN Toolkit of Best Practice for Research Data Management

<http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf?pdf=RDMToolkit>

Freely available to download, this RDM Toolkit is a precious resource for anyone helping their institution to better manage research data. It is a collection of case studies illustrating best practices in relation to research data management. It includes: 23 best-practice case studies from institutions around the world, 8 main sections, on topics such as policy and leadership, open data, advocacy and costs, one model RDM policy (by the University of Vienna) and an overview of 20 RDM policies across Europe.

7.3 Help Desks

Researchers, research institutions and policy makers dealing with OA (to publications and data) issues, have the opportunity to get assistance from some Help Desks that provide specific support for compliance with the EC Open Access policies including Open Research Data pilot.

OpenAIRE NOADS

<https://www.openaire.eu/contact-noads>

The National Open Access Desks connect researchers, research institutions, and policy makers at a national level on the one end, and the OpenAIRE project services on the other. The focus of the National Open Access Desks activities is on support for compliance with the EC Open Access policies. National Open Access Desks can help finding the appropriate repository for each country, and can answer questions concerning Open Access, OpenAIRE, copyright issues, any special national rules and regulations concerning Open Access, and so on. They will redirect questions if necessary, especially when national issues, like copyright, are involved.

IPR helpdesk

<https://www.iprhelpdesk.eu/>

The European IPR Helpdesk offers free of charge, first-line support on Intellectual Property (IP) and Intellectual Property Rights (IPR) matters to beneficiaries of EU funded research projects and EU SMEs. Offering a broad range of informative material, a Helpline service for direct IP support and on-site and online training, the European IPR Helpdesk's main goal is to support IP capacity building along the full scale of IP practices.

EC Directorate-General for Research and Innovation – RTD Open Access team

<https://data.europa.eu/euodp/data/publisher/rtd>

The Directorate-General for Research and Innovation defines and implements European Research and Innovation (R&I) policy with a view to achieving the goals of the Europe 2020 strategy and the Innovation Union flagship initiative. The RTD OA team provides input and assistance to public on the OA issues including ORD and is responsible to monitor the pilot and provide an overview of its uptake.

7.4 Tools for Data Management Plans

EC requires from beneficiaries of Horizon 2020 funded projects to develop and implement Data Management and Sharing Plans (DMPs). Those plans typically state what data will be created and how during the project, and outline the plans for data sharing and preservation, noting what is appropriate given the nature of the data and any restrictions that may need to be applied.

Horizon 2020 FAIR Data Management Plan (DMP) template (Annex 1 to Guidelines on FAIR Data Management in Horizon 2020)

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

ERC Data Management Plan Template

http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-erc-tpl-oa-data-mgt-plan_en.docx

DCC resources for DMP

The DCC ³⁶ has analysed UK funders' policies and developed various tools on data management in response to the different topics here below. Those resources are also suited for application on the ORD pilot:

- ***DMPonline***

A flexible web-based tool to assist users to create personalized plans according to their context or research funder. The tool also aids researchers by providing examples of guidance and best practice via 'crowdsourced' links to DCC resources and external advice.

- ***Funders' data plan requirements***

Summary of funders' expectations for data management and sharing plans

- ***Checklist for a Data Management Plan***

A list of questions and guidance that researchers may find useful when writing data management and sharing plans.

³⁶ Digital Curation Centre (DCC) <http://www.dcc.ac.uk/resources/data-management-plans>

- ***DMP checklist leaflet***
A foldout summary of the Checklist (hard copies available upon request).
- ***FAQ on Data Management Plans***
A short list of key questions pertaining to Data Management Plans.
- ***How to Develop a Data Management and Sharing Plan***
A guide that outlines the process of developing a data management and sharing plan. It provides practical guidance to researchers and support staff in meeting the expectation of UK funders.
- ***Guidance and examples***
Advice to help you write your data management and sharing plan.

7.5 Finding a research data repository

Projects participating in the ORD Pilot should deposit in a research data repository and take measures to enable third parties to access, mine, exploit, reproduce and disseminate, free of charge for any user. There are many resources available on line that can support researchers to identify the most suitable data repository.

Zenodo

<https://zenodo.org/>

The OpenAIRE project, in the vanguard of the open access and open data movements in Europe was commissioned by EC to support their Open Data policy by providing a catch-all repository for the EC funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability and Zenodo was launched in May 2013. Zenodo is a multi-disciplinary repository that can be used for the long-tail of research data to ensure that everyone can join in Open Science.

Re3data

www.re3data.org

Re3data is a global registry of research data repositories from a diverse range of academic disciplines. It provides information on repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. This DataCite service is referenced by multiple publishers in their editorial policies as the best tool to identify the most appropriate data repository and recommended in the European Commission's Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.

Where to keep research data - DCC Checklist for Evaluating Data Repositories

<http://www.dcc.ac.uk/sites/default/files/documents/publications/Where%20to%20keep%20research%20data.pdf>

The checklist help researchers make informed choices about where to deposit data. It is also relevant to managers with responsibility for defining policy on Research Data Management (RDM).

B2SHARE

<https://www.eudat.eu/services/b2share>

It is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store and share small-scale research data from diverse contexts.

7.6 Licensing tools and legal guidelines and resources

The aim of these tools is to provide support, by giving concrete recommendations to researcher, on key legal issues related to open research data, including data protection, intellectual property rights, data sharing, licensing and re-use.

DCC “How-to Guides & Checklists”

<http://www.dcc.ac.uk/resources/how-guides>

The DCC website section “How-to Guides” provides working-level knowledge of curation topics, background concepts and practical steps aiming to help people (in research or support roles) implement data management capabilities in their organisation. The “Checklists” section also provide working-level knowledge, but cover less of the relevant background. Each aims to ensure practitioners have addressed the full scope of a challenging curation topic, and provides further sources of guidance. Among them, here we mention the following:

DCC “Checklist on Legal Aspects of RDM”

<http://www.dcc.ac.uk/resources/how-guides/rdm-law>

This checklist is intended to help researchers and support staff involved in developing or delivering support for research data management (RDM). The document summarizes the following common elements of the main challenges for RDM support:

1. Protection of Personal Data
2. Freedom of Information (FOI) and Environmental Information (EIR)
3. Intellectual Property Rights (IPR) in Data and Databases
4. Data Sharing, Licensing and Re-use
5. Legal Considerations of Cloud Service Provision

DCC “How to License Research Data”

http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf

This guide is addressed to principal investigators, researchers responsible for the data and those who provide access to them through a data centre, repository or archive. The guide helps in deciding how to apply a licence to research data and which licence would be most suitable. Though it concentrates on the UK context, some aspects apply internationally. The document covers the main challenges for RDM support services, and the

services and sources of legal information available to them:

1. Why license research data?
2. Licensing concepts
3. Prepared licenses
4. Bespoke licenses
5. Standard licenses
6. Multiple licensing
7. Mechanisms for licensing data
8. Licensing related information

DCC “Curation Reference Manual”

<http://www.dcc.ac.uk/resources/curation-reference-manual>

It is an ongoing, community-driven project, which involves members of the DCC community and contains advice, in-depth information and criticism on current digital curation techniques and best practice.

EUDAT licensing tool and legal guide

<https://eudat.eu/services/userdoc/license-selector>

<http://wiz.eudat.eu/#/app/home>

The EUDAT tool includes a built-in license wizard that helps in licencing data suggesting suitable license for ones research data. The license selector tool allows a user to attach a specific license to their data set or software package in order to protect it from certain usage or abuse. This document is targeted at all users that want to add a license to their data set, but don't know which license suits best to their data set distribution and rights requirements. The EUDAT Legal Guide has been developed by the EUDAT consortium through the help of a legal expert.

Amnesia OpenAIRE

<http://amnesia.imis.athena-innovation.gr:8080/amnesia/>

AMNSIA is a service recently developed by OpenAIRE that allows data curators to anonymize their data. Research data – especially in fields like medicine, biotechnology and the social sciences – often contain personal data. This means that many datasets, in their raw form, cannot be made available on an Open Access basis as is required by the Open Research Data pilot due to conflicts with rules on protection of personal data. Hence, incompatibility with data protection regulations is one of the major reasons for opting out of the ORD pilot. The best way to fulfil the requirements of the pilot and data protection rules at the same time is to anonymize personal (research) data before making them openly available. Anonymized data are no longer personal data, which means that data protection rules are no longer applicable.

Open Data Commons

<https://opendatacommons.org/>

Open Data Commons provides a set of legal tools and open data licences that enable organisations and individuals to make their data legally “open”, free for anyone to use, reuse, and redistribute.

7.7 Costing Data Management

Early planning of data management and sharing can significantly reduce the costs of these activities during the research, in terms of time and resources needed. There is not only one method for assessing costs, but a number of costing models exist to help in this task.

UK Data Archive - Data management costing tool

<http://www.data-archive.ac.uk/create-manage/planning-for-sharing/costing>

The UK Data Archive has developed a simple tool that can be used for costing the additional expenses – above standard research procedures and practices – that are needed to make research data shareable beyond the primary research team.

Curation Costs Exchange

<http://www.curationexchange.org/>

They give helps in costing curation activities and provide with the basic concepts used to assess curation costs. Different cost models are suggested to describe one’s organization activities. The Cost Comparison Tool also permits comparing digital curation costs with other organisations.

7.8 On-line courses

Projects supporting OA as well as universities and research centres provide on-line courses focused on open science and open data issues in order to help users understand how to manage the digital data collected throughout their research.

FOSTER

<https://www.fosteropenscience.eu/>

The FOSTER portal supports e-learning, blended learning, self-learning and dissemination of training materials and provide on-line courses focused on open science and open data issues.

MANTRA

<http://datalib.edina.ac.uk/mantra/>

Backed by Edina and Data Library at The University of Edinburgh, MANTRA is a free, online non-assessed course with guidelines to help users understand and reflect on how to manage the digital data collected

throughout research. It has been crafted for the use of post-graduate students, early career researchers and information professionals.

Research Data Management and Sharing

<https://www.coursera.org/learn/data-management>

Created by the University of North Carolina and the University of Edinburgh and available at Coursera (a provider platform for on-line learning), it is an on-line course providing learners with a basic introduction to research data management and sharing. The topics range from diversity of data and their management needs across the research data lifecycle to identification of the components of good data management plans and best practices for working with data (organization, documentation, storage and security of data).

RDNL Essentials 4 Data Support

<http://datasupport.researchdata.nl/en/about-the-course/>

Research Data Netherlands provide a course focused on “data supporter”, that is anyone supporting researchers in storing, managing, archiving and sharing research data (library staff, IT staff, data librarians and specialists and anyone with duties involving data management).

There are three different ways to take the course: online only (without registration), online only with user profile, full course (online and face-to-face with certificate).

DataONE

<https://www.dataone.org/education-modules>

The Data Observation Network for Earth offers data management education modules addressed to young researchers in the Earth Science. Topics covered includes data management, data sharing, data management planning, data entry and manipulation, data quality control and assurance, data protection, metadata, data citation and legal and policy issues on data.

8. CONCLUSIONS AND RECOMMENDATIONS

The analysis made so far shows that implementing OA and, more specifically, participating in the “extended” ORD pilot is a challenging but demanding and time-consuming activity for applicants to H2020 calls. Researchers and data managers within research institutes and academia are facing new complex tasks that require specific technical knowledge and involvement of additional and dedicated resources.

Although there is a general consensus and a certain awareness on the potential benefits of OA for science, the present study shows that the road towards the openness of research data is perceived differently by several representatives of the research community. Feedback and insights, collected from researchers and other relevant stakeholders, involved in diverse activities within the ORD pilot, disclose that there are still important resistances and real barriers to overcome. Based on the experiences they reported, main critical issues have been identified, grouped and are here detailed.

Cultural and knowledge issues

The lack of education and information about OA issues or, often, of an erroneous data sharing culture among researchers, data managers and other stakeholders cause confusion about the basic principles of OA and the EC requirements related to the ORD pilot, and foster wrong assumptions on the data management activity in general.

The different levels of experiences and knowledge of research staff among/within disciplines also increase the difficulties in carrying out the activities on data management required by EC in a successful and efficient way. To overcome these obstacles, researchers and technicians, new and established, should be provided with education and specific skills to strengthen their competencies in research data management and stewardship.

Structural issues

Participating in the ORD pilot is a demanding activity in terms of efforts and/or working time. The complexity of the EC requests regarding the pilot and the novelty of these issues need a timely and sized planning of the needed resources (human resources, person/months, costs, etc.) since the proposal stage.

Technical issues

Within the research community, it is recognised a lack of discipline-specific and widely adopted practices for managing data and difficulties for scientists to adhere to shared standards when collecting their data and metadata. With reference to the development of the Data Management Plan, there is evidence of a poor standardisation in the DMP structure and content and the need for specific skills in drafting and managing the Plan during the project lifetime.

Technological aspects are also perceived as a concern in implementing OA to research data. Dedicated research data repositories are not so widespread among research institutions and it is necessary to develop comprehensive and collaborative technical infrastructures and services for managing and sharing data, and to guarantee a long-term preservation.

Legal and ethical issues

Several legal issues must be considered when implementing OA strategies, including intellectual property

rights, data protection obligations, licensing and legislation. Open access to research data can also raise several ethical concerns especially when dealing with personal and sensitive data (such as medical data). The complexity of dealing with legal and ethical issues should not be seen as a barrier but as an opportunity to find sound solutions. Synergies between data managers and IPR experts are crucial in managing the various aspects linked to data protection and to the choice of licences for specific data sets.

Regulations and policy issues

Policies on data management are often lacking or not sufficiently defined or implemented within research institutes and universities. In addition, significant differences in data policies and data management processes can be found among diverse institutions. It also happens that cautionary attitudes and strict DM policies requested by some data providers are not truly motivated by any objective reason. Divergent approaches to open research to data in different scientific fields also emerge as critical in the development of appropriate policies.

Motivating issues

Resistances to the OA implementation are also amplified by a lacking awareness of the benefits of open access and data sharing. Besides an adequate system of incentives and rewards for researchers and data managers is missing and FAIR data is not yet a criterion used for the evaluation of the researcher's career and of projects.

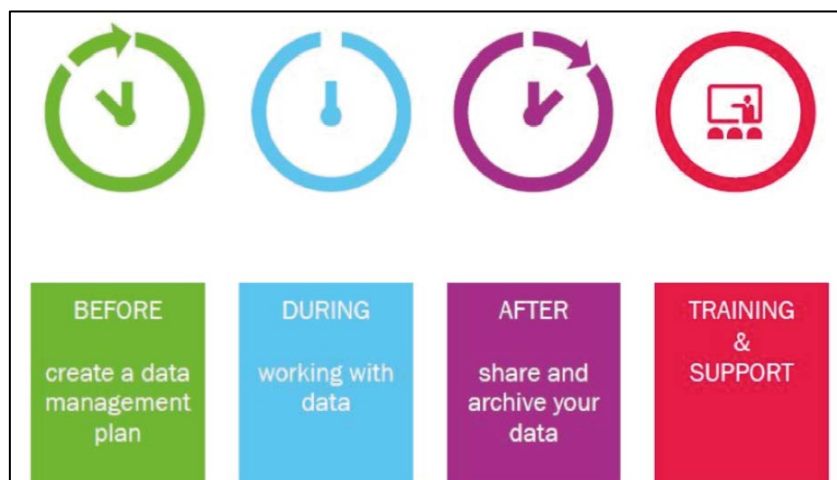


Figure 7 – Data lifecycle and training needs
(source: NordForsk Policy Paper 1 “Open Access to Research Data – Status, Issues and Outlook”, 2016)

Taking into account these critical issues, a list of recommendations is here proposed and specifically addressed to research institutes and university, with the aim to help facing the new complexity originating from the participation in the “extended” ORD pilot. Some general recommendations are also provided to the relevant stakeholders to help them contributing to the efficient implementation of the OA approach.

8.1 Specific recommendations for H2020 applicants

- Gain a deep insight into the EC guidelines and policy on open research data in order improve the understanding of the ORD pilot requirements and rules.
- Participate in specific training sessions for researchers and data managers in order to strengthen one's competences and skills in dealing with data, metadata and stewardship.
- Organize advocacy programmes dedicated to the staff supporting researchers in data management activities (research offices, IT departments, academic and research units, data libraries, etc.).
- Plan in advance (since the proposal stage) the commitment of the human and financial resources that are needed for the data management and stewardship (in terms of staff, person/months, costs, etc.).
- Build or make use of dedicated research data repositories that facilitate researchers to archive their data in appropriate places.
- Develop comprehensive and collaborative technical infrastructures and services for managing and sharing FAIR data and their long-term preservation.
- Turn to counselling services on ethical and legal issues (data protection obligations, intellectual property rights, data sharing, licensing and re-use) that can originate from openness to research data.
- Implement a clear policy on data management within the research institute/academia.

8.2 Other general recommendations

- Encourage education and information about OA issues in order to foster the data sharing culture among researchers and other relevant stakeholders.
- Work together with other institutions in order to share infrastructures, knowledge and best practices.
- Strengthen the awareness among the research community about benefits of open access to data and publications.
- Promote an adequate system of incentives and rewards for researchers and data managers who are responsible for producing, sharing and managing high-quality data.
- Develop relevant research data policies and procedures within institutions where the Open Access is the default.
- Consider the essential linkage between research data Policy, Technology and Support. A coordinated and parallel approach of these three aspects should be taken into account for a conscious and successful use of research data.

It is a shared idea that, in the digital age, Research Data can drive innovation and stimulate new discoveries, to the great benefit of Society. Being aware of the unsolved problems, all stakeholders in the research workflow have a role to play and act in order to overcome the major gaps and make Science more efficient and productive.

9. FURTHER REFERENCES

- European Open Science Cloud (EOSC) Summit, Input Paper Sessions 1, 2 , 3 and 4 (2017) <http://ec.europa.eu/research/index.cfm?pg=events&eventcode=44D86060-FBA1-1BD1-9355822B162BBOEE>
- Johnson et al. (2017) OpenAIRE Report, Towards a Competitive and Sustainable OA Market in Europe – A Study of the Open Access Market and Policy Environment
- Schmidt B, Gemeinholzer B, Treloar A (2016) Open Data in Global Environmental Research: The Belmont Forum’s Open Data Survey. PLoS ONE 11(1): e0146695. doi:10.1371/journal.pone.0146695
- Research Councils UK et al. (2016), Concordat on Open Research Data
- Wilkinson et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship Scientific Data 3, Article n. 160018 <http://dx.doi.org/10.1038/sdata.2016.18>
- Guedj D & Ramjoué C, Biomed Data J. 2015; 1(1): 11-14 <http://dx.doi.org/10.11610/bmdj.01102>
- OpenAIREplus Executive report (2014) <https://zenodo.org/record/15464#.WVDsjWjyhPY>
- Research Data Alliance, The Data Harvest: how sharing research can yield knowledge, jobs and growth (2014) <https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf>
- Fava I & Gargiulo P (2013) What do Italian Researchers think about Open Research Data? http://eprints.rclis.org/29671/1/What%20do%20Italian%20Researchers%20think%20about%20Open%20Research%20Data_rev-charts.pdf
- LERU Research Data Working Group, LERU Roadmap for Research Data (2013) http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf
- “Commission launches pilot to open up publicly funded research data” (2013) EC press-release IP-13-1257_en.htm
- Communication from the Commission, Towards better access to scientific information: Boosting the benefits of public investments in research (2012)
- “Scientific data: open access to research results will boost Europe's innovation capacity” (2012) EC press-release IP-12-790
- Communication from the Commission, Europe 2020 A strategy for smart, sustainable and inclusive growth (2010) <http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf>
- Ivo Grigorov, a video for the FOSTER Project “What is open science” <https://youtu.be/7Kric3x7zr0>

ANNEX 1: QUESTIONNAIRE A

**MASTER DEGREE PROGRAM IN
“MANAGEMENT OF RESEARCH, INNOVATION AND TECHNOLOGY”
MASTER MIT IV ED.
MIP Politecnico di Milano Graduate School of Business**

Project work

**"Open Research Data Pilot: barriers to implementation and instruments for supporting research"
(R. BLANOS, C. FANARA)**

- 1) IS YOUR ORGANIZATION A PARTNER IN ANY H2020 PROJECTS PARTICIPATING IN THE OPEN RESEARCH DATA PILOT?
IF YES:
IN WHICH PROJECTS AND IN WHICH ROLE?

IF NOT:
DOES YOUR ORGANIZATION HAVE EXPERIENCE AS DATA MANAGER/REPOSITORY (OR WITH OTHER ROLES) IN PROJECTS PARTICIPATING IN THE OPEN RESEARCH DATA PILOT?
- 2) UPON YOUR EXPERIENCE, WHICH ARE THE MAIN PROBLEMS ENCOUNTERED IN ADHERING TO THE OPEN RESEARCH DATA PILOT?
- 3) REFERRING TO THE ETH'S "DATA MANAGEMENT CHECK LIST" (<https://www.library.ethz.ch/de/Media/Files/Data-management-checklist>), IN WHICH POINT OF THE LIST CAN YOUR ORGANIZATION GIVE ITS CONTRIBUTION?
- 4) DOES YOUR INSTITUTION OFFER REPOSITORY SERVICES?
IF YES:
WHICH GUIDELINES FOR SUSTAINABLE DATA FORMATS AND METADATA STANDARDS DO YOU PROVIDE?
- 5) HOW DO YOU SUPPORT RESEARCHERS IN DEALING WITH SENSITIVE DATA AND LICENSING?
- 6) AS DATA MANAGER, HOW DO YOU HANDLE PROBLEMS RELATED TO: INTEGRITY, DATA UPGRADE AND ETHIC ISSUES FOR SENSITIVE DATA (E.G. ETHICAL DATA)?
- 7) WHICH ARE THE POSSIBLE PROCEDURES FOR CONTROLLING DATA ACCESS IN A REPOSITORY?

ANNEX 2: QUESTIONNAIRE B

**MASTER DEGREE PROGRAM IN
“MANAGEMENT OF RESEARCH, INNOVATION AND TECHNOLOGY”
MASTER MIT IV ED.
MIP Politecnico di Milano Graduate School of Business**

Project work
"Open Research Data Pilot: barriers to implementation and instruments for supporting research"
(R. BLANOS, C. FANARA)

GIVEN YOUR EXPERIENCES AS COORDINATOR OF A H2020 PROJECT PARTICIPATING IN THE OPEN RESEARCH DATA PILOT, ARE YOU DIRECTLY INVOLVED IN THE DATA MANAGEMENT ACTIVITY?

IF YES,

- 1) HOW DO YOU MANAGE THE PARTICIPATION IN THE PILOT AND THE FULFILMENT OF THE H2020 REQUIREMENTS?
- 2) WHICH TOOLS, EXPERTISE OR GUIDELINES DID YOU FIND USEFUL?
- 3) WHICH ARE THE MAIN PROBLEMS AND CRITICAL ISSUES FACED IN TAKING PART IN THE PILOT?

IF NOT,

COULD YOU PLEASE GIVE US THE CONTACTS OF THE PARTNER WHO IS RESPONSIBLE FOR THIS TASK IN YOUR PROJECT?

ANNEX 3: ORGANIZATION PROFILES



ACTIVITIES: As a non-profit research organization, APRE, in close collaboration with the Ministry of Education, University and Research (MIUR), provides its members (as well as businesses, government agencies, and private individuals) information, support and assistance for participation in national and European programmes and collaborative initiatives in the field of Research, Technological Development and Innovation (RTDI) and in the transfer of research results.

CONTACT PERSON: **Matteo Di Rosa**, EU Horizon 2020 - National Contact Point SC5 and SC2.



ACTIVITIES: Cineca is a nonprofit Consortium, made up of 70 Italian universities, 8 Italian Research Institutions and the Italian Ministry of Education. Today it is the largest Italian computing center, one of the most important worldwide. With more seven hundred employees, it operates in the technological transfer sector through high performance scientific computing, the management and development of networks and web based services, and the development of complex information systems for treating large amounts of data. It develops advanced Information Technology applications and services, acting like a trait-d'union between the academic world, the sphere of pure research and the world of industry and Public Administration

CONTACT PERSON: **Claudio Cacciari**, Degree in Physics. Experience in software integration, Service Oriented Architecture (SOA) tools, Grid and Cloud environments. Collaboration to some FP7 European infrastructural projects, such as PRACE for the computing and EUDAT for the data management, and EMI for Grid middleware integration.

ACTIVITIES: DANS encourages researchers to make their digital research data and related outputs Findable, Accessible, Interoperable and Reusable.

DANS's activities are centred around 3 core services: data archiving, data reusing and training & consultancy. Sharing and reusing research data promotes science. Datasets collected for a certain research purpose can also contain answers to research questions from very different research in the same or in another research discipline. DAN's core services are: DataverseNL for short-term data management, EASY for long-term archiving, and NARCIS, the national portal for research information. By participating in (inter)national projects, networks and research, DANS contributes to continued innovation of the global scientific data infrastructure. Open if possible, protected where necessary. DANS is an institute of the Dutch Academy KNAW (Royal Netherlands Academy of Arts and Sciences) and funding organisation NOW (Netherlands Organisation for Scientific Research)

CONTACT PERSON: Marjan Grootveld, Senior policy officer.

DCC - DIGITAL CURATION CENTRE

<http://www.dcc.ac.uk/>

ACTIVITIES: The Digital Curation Centre (DCC) is an internationally-recognized center of expertise in digital curation with a focus on building capability and skills for research data management. The DCC provides expert advice and practical help to research organizations wanting to store, manage, protect and share digital research data. They provide access to a range of resources including our popular How-to Guides, case studies and online services. Their training programmes aim to equip researchers and data custodians with the skills they need to manage and share data effectively.

They also provide consultancy and support with issues such as policy development and data management planning.

Since mid-2011 they have worked closely with research organizations to provide tailored support for service development, from making the case for research data management through the assessment of needs to the design and implementation of policy, support, infrastructure and services.

CONTACT PERSON: Martin Donnelly, Senior Institutional Support Office, external reviewer for data management plans.



DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION

<http://ec.europa.eu/research/index.cfm?pg=dg>

ACTIVITIES: The Directorate-General for Research and Innovation defines and implements European Research and Innovation (R&I) policy with a view to achieving the goals of the Europe 2020 strategy and its key flagship initiative, the Innovation Union.

To do so, the DG contributes to the European Semester by analysing national R&I policies, by assessing their strengths and weaknesses, and by formulating country specific recommendations where necessary. It monitors and contributes to the realisation of the Innovation Union flagship initiative and the completion of the European Research Area. It funds excellent Research and Innovation through Framework Programmes taking a strategic programming approach.

CONTACT: RTD Open Access Team



HELMHOLTZ-CENTRE POTSDAM - GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES (GFZ)

<http://www.gfz-potsdam.de/en/home/>

ACTIVITIES: The GFZ German Research Centre for Geosciences is the national research centre for Earth sciences, research at the GFZ focuses on the geosphere within the highly complex System Earth with its further subsystems, its interacting subcycles, and its wide network of cause-and-effect chains. This we do in a close interdisciplinary collaboration with the related scientific disciplines physics, mathematics, chemistry, and biology as well as with the engineering sciences disciplines of rock mechanics, engineering hydrology and seismology. Their mission is to assess and understand relevant physical, chemical, and biological processes within the geosphere and to predict future developments. They integrate methods of Earth-observation with laboratory- and field-experiments, as well as with modelling approaches.

CONTACT PERSON: Kirsten Elger, Library and Information Services



**ISTITUTO NAZIONALE DI OCEANOGRAFIA E DI GEOFISICA
SPERIMENTALE (OGS)**

<http://www.ogs.trieste.it/en>

ACTIVITIES: The National Institute of Oceanography and Applied Geophysics - OGS - is a public research Institute which acts internationally in the fields of Earth and Marine Sciences, Oceanography, Geophysics and Seismology. The Institute aims at safeguarding and enhancing the environmental and natural resources and focuses its efforts on evaluating and preventing geological, environmental and climatic risks, and spreading the scientific culture and knowledge.

OGS has four locations in the Friuli Venezia Giulia Region (North-Eastern Italy) and it is structured under four main Departments:

- Oceanography - OCE
- Geophysics - GEO
- Seismological Research - CRS
- Research Infrastructures - IRI

With its strategic infrastructures of excellence (such as the oceanographic research vessel OGS Explora), OGS makes its own expertise available for research related to environment and climate, biodiversity and ecosystem functionality and to the study of seismicity, hydrodynamic and geodynamic phenomena having an impact on both environment and population.

DEPARTMENT INVOLVED: Department of Oceanography - OCE

The mission of the Department of Oceanography - OCE is to deepen and to advance the knowledge in the field of oceanography with a multidisciplinary approach in order to answer urgent societal and environmental-related challenges. The activity plan is developed around four main research themes. Within the OCE, two centers of excellence and three research infrastructure (ECCSEL, PRACE, Euro-Argo), are important elements of OGS scientific and thriving core.

CONTACT PERSON: Alessandro Crise, Research Director with degree in Physics. Partner of major European projects leading WP/tasks connected with ocean observing systems, operational oceanography and coupled physical-biogeochemical modelling of Mediterranean Sea. His expertise includes hydrodynamic and biogeochemical modelling, operational oceanography, and science-to-policy interface.



OPENAIRE NATIONAL OPEN ACCESS DESK - ITALY

<https://www.openaire.eu/contact-noads>

ACTIVITIES: The National Open Access Desks connect researchers, research institutions, and policy makers at a national level on the one end, and the OpenAIRE project services on the other. The focus of the National Open Access Desks activities is on support for compliance with the EC Open Access policies.

National Open Access Desks can help find the appropriate repository in each country, and can answer questions concerning Open Access, OpenAIRE, copyright issues, any special national rules and regulations concerning Open Access, and so on. They will redirect questions if necessary, especially when national issues, like copyright, are involved

CONTACT PERSON: Paola Gargiulo, OA projects, Training & Support Coordinator.



NATIONAL RESEARCH COUNCIL (CNR)

<https://www.cnr.it/en>

ACTIVITIES: The National Research Council (CNR) is the largest public research institution in Italy, the only one under the Research Ministry performing multidisciplinary activities.

CNR's mission is to perform research in its own Institutes, to promote innovation and competitiveness of the national industrial system, to promote the internationalization of the national research system, to provide technologies and solutions to emerging public and private needs, to advice Government and other public bodies, and to contribute to the qualification of human resources.

DEPARTMENTS INVOLVED

1) IMAA atmospheric observatory

The CNR-IMAA atmospheric observatory (CIAO) represents a well-established ground-based remote-sensing observatory for the study of weather and climate. The observatory consists of a combination of advanced systems able to provide high quality long-term observations of aerosol and cloud properties. Since 2000, systematic observations of aerosol, water vapour and clouds have been collected and the acquisition of new active and passive microwave profilers has strengthened the equipment required for performing accurate aerosol and cloud observations. Currently, CIAO represents the largest ground-based remote-sensing station in the Mediterranean Basin and is one of the first atmospheric observatories in Europe.

The main scientific objective is the long-term measurement for the climatology of aerosol and cloud properties. The observation strategy is mainly organized in order to provide quality assured measurements for satellite validation and model evaluation and to fully exploit the synergy and integration of the active and passive sensors for the improvement of the atmospheric profiling.

CONTACT PERSON: **Lucia Mona**, researcher at the CNR-IMAA. Responsible for the EARLINET (European Aerosol Research Lidar NETwork) database and its link to the ACTRIS data portal within the ACTRIS2 (H2020) project.

PROJECTS DESCRIPTION:

EARLINET, the European Aerosol Research Lidar Network, was established in 2000 as a research project with the goal of creating a quantitative, comprehensive, and statistically significant database for the horizontal, vertical, and temporal distribution of aerosols on a continental scale. Since then EARLINET has continued to provide the most extensive collection of ground-based data for the aerosol vertical distribution over Europe.

ACTRIS is the European Research Infrastructure for the observation of aerosol, clouds, and trace gases. The Research infrastructure is composed of observing stations, exploratory platforms, instruments calibration centres, and a data centre and provides access to a large number of services, implemented through its virtual, physical and remote access programme, comprising wide and open access to high-quality information, tools and services through the ACTRIS Data Centre, hands-on access to a large number of ground-based observing platforms and advanced instrumentation, and physical and remote access to thematic calibration centres for instrument testing and development, quality assurance, training and best practices, and international collaboration.

2) Institute of Information Science and Technologies (ISTI)

The CNR-ISTI main mission is to producing scientific excellence and playing an active role in technology transfer. Part of the largest research institution in Italy (the National Research Council, or CNR), CNR-ISTI is organised in 16 technology centres or labs, each of them pursuing a well-defined set of scientific objectives. State-of-the-art technology development and training are other activities performed by the centres.

CONTACT PERSON: **Donatella Castelli**, Senior Researcher. BlueBRIDGE Project Coordinator

PROJECT DESCRIPTION

BlueBRIDGE's overall objective is to support capacity building in interdisciplinary research communities actively involved in increasing scientific knowledge about resource overexploitation, degraded

environment and ecosystem with the aim of providing a more solid ground for informed advice to competent authorities and to enlarge the spectrum of growth opportunities as addressed by the Blue Growth Societal Challenge. BlueBRIDGE capitalizes on past investments and uses the proven D4Science infrastructure that counts over 1400 users, integrates more than 50 repositories, executes around 20,000 models & algorithms per month and provides access to over a billion records in repositories worldwide, with 99,7% service availability. BlueBRIDGE aims to develop innovative services in the following areas:

- Blue Assessment - services for stock assessment and for the generation of unique identifiers for global stocks;
- Blue Economy - services supporting the analysis of socio-economic performance in aquaculture;
- Blue Environment - spatial planning services to identify aquaculture and fisheries infrastructures from satellite imagery;
- Blue Skills - on-line training services and capacity building on existing training modules for fisheries scientists and other practitioners.



UNIVERSITY OF BOLOGNA, ALMA MATER STUDIORUM

<http://www.unibo.it/en/homepage>

ACTIVITIES: Faced with the relentless evolution of our society, driven by ongoing paradigm shifts, the University is called upon not only to respond to externally determined needs but also to pre-empt, from within, the key questions around which to plan its training and research activities. So, in order to remain true to its centuries-old mission, the University is presenting a very sound, coherent and realistic Strategic Plan, and showing the potentials that will enable it to tackle the challenges of the coming years.

New knowledge and learning, straddling the humanities, science and technology, are the focus of the research activities of the University, as it engages in an intensive exchange of ideas with all the private stakeholders who have its same concerns and share with it the ethical and independent character of research.

DEPARTMENTS INVOLVED

1) Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM)

<http://www.dicam.unibo.it/en>

The Department of Civil, Chemical, Environmental and Materials Engineering (DICAM) is the result of a large aggregation of research skills and laboratory facilities of diverse but interrelated engineering branches traditionally developed at the University of Bologna. The department integrates the scientific expertise of various research groups, such as Structural, Transport, Hydraulic, Survey and Territory

Engineering, Applied Chemistry and Materials Science, Chemical, Mining, Petroleum and Environmental Engineering.

CONTACT PERSON: Dario Frascari and Giulio Zanaroli: Assistant Professors. Team leaders of H2020 MADFORWATER project (Development And application of integrated technological and management solutions FOR wasteWATER treatment and efficient reuse in agriculture tailored to the needs of Mediterranean African Countries).

PROJECT DESCRIPTION

MADFORWATER. The aim of this project is to develop a set of integrated technological and management solutions to enhance wastewater treatment, reuse for irrigation and water efficiency in agriculture in three MACs (Tunisia, Morocco and Egypt). MADFORWATER will develop and adapt to three main hydrological basins in the selected MACs technologies for the production of irrigation-quality water from drainage canals, municipal, agro-industrial and industrial wastewaters, and technologies for water efficiency and reuse in agriculture, initially validated at laboratory scale. Selected technologies will be further adapted and validated in four field pilot plants of integrated wastewater reuse. Integrated strategies for wastewater treatment and reuse targeted to the selected basins will be developed, and guidelines for the development of integrated water management strategies in other basins of the three target MACs will be produced, considering climate change, population increase and economic growth scenarios. The social and technical suitability of the developed technologies and non-technological instruments in relation to the local context will be evaluated with the participation of MAC stakeholders and partners. Guidelines on economic instruments and policies for the effective implementation of the proposed water management solutions in the target MACs will be developed.

2) Department of Computer Science and Engineering (DISI)

<http://www.cse.unibo.it/en>

DISI is the single reference point for Computer Science and Engineering education, research and technology transfer at the University of Bologna. Its goal is to contribute to the advancement of society through education, research and technology transfer. This goal is achieved by offering a broad spectrum of expertise ranging from theoretical computer science to software, hardware and application design and development.

CONTACT PERSON: Edoardo Mollona, Full Professor. Team leader of the H2020 PERCEIVE project (Perception and Evaluation of Regional and Cohesion policies by Europeans and Identification with the Values of Europe).

PROJECT DESCRIPTION

PERCEIVE (Perception and Evaluation of Regional and Cohesion policies by Europeans and Identification with the Values of Europe) project aims at both mapping and explaining inter- and intra-regional variations in: a) the experiences and results of cohesion policy implementation, b) citizens' awareness and appreciation of EU efforts for delivering cohesion and c) European identities and citizens' identification with the EU. In doing so, PERCEIVE will develop a comprehensive theory of "cohesion in diversity", and use this theory to create a better understanding of the channels through which European policies contribute to create both different local understandings of the EU and different levels of European identification across profoundly different European regions.

3) Department of Psychology (PSI)

<http://www.psicologia.unibo.it/it/dipartimento>

The main research activities cover different areas of psychology: general and cognitive psychology, cognitive neuroscience and neuropsychology, psychobiology and psychophysics, psychology of emotions, work and organizations, psychology of health, clinical and community psychology and psychological evaluation.

CONTACT PERSON: Elvira Cicognani, Full professor. Coordinator of Horizon 2020 project CATCH-EyoU (Constructing AcTive CitizensHip with European Youth)

PROJECT DESCRIPTION.

CATCH-EyoU. Through the joint contribution of different disciplines (Psychology, Political Science, Sociology, Media and Communications, Education) CATCH-EyoU has the aim to identify the factors, located at different levels (psychological, developmental, macro social and contextual) influencing the different forms of youth active engagement in Europe. Through different studies, qualitative, quantitative, and an active citizenship intervention in schools, the project will provide a multifaceted understanding of the different factors influencing the perspectives of young people on Europe and of the ways in which young people engage in society. The project offers to policy makers new instruments and "conceptual lenses" to better understand young generation, how they approach public authorities and how they engage materially and symbolically in order to participate in the construction of the societies they inhabit and shape the governmental regimes under which they live.

4) Department of Agricultural and Food Sciences (DISTAL)

<http://www.distal.unibo.it/it>

DISTAL was founded in 2012 by the merger of the Departments of Economy and Agricultural Engineering, Food Science, Food Protection and Enhancement, to foster the integration and strengthening of existing skills in the three structures.

CONTACT PERSON: Tullia Gallina Toschi, Full professor. Coordinator of Horizon 2020 projects, OLEUM and PLOTINA.

PROJECTS DESCRIPTION

OLEUM

The OLEUM project will develop new and improved existing analytical methods for detecting fraud and ensuring quality of olive oil, and to improve technology sharing by establishing of a wide community of laboratories and institutions involved in quality control. Improvements in the quality, safety and authenticity of olive oils will boost consumer confidence and ultimately enhance the competitiveness of the EU olive oil market.

PLOTINA (Promoting gender baLance and inclusiOn in research innovaTioN and trAining). The overall objective of PLOTINA is to enable the development, implementation and assessment of self-tailored Gender Equality Plans (GEPs) with innovative and sustainable strategies for the Research Performing Organizations (RPOs) involved. This objective will be achieved by:

- Stimulating a gender-aware culture change
- Promoting career-development of both female and male researchers to prevent the waste of talent, particularly for women
- Ensuring diversification of views and methodologies (in this case by taking into account the gender/sex dimension and analysis) in research and teaching.



SAPIENZA UNIVERSITY, ROME

<http://en.uniroma1.it/>

ACTIVITIES: Sapienza University, which was founded in 1303 by Pope Boniface VIII in Rome, is one of the oldest universities in the world and a top performer in international university rankings. The future of Sapienza is firmly rooted in its rich past and the academic and research excellence of its community. Sapienza's mission is to catalyse the development of a knowledge society by promoting research, education and international cooperation.

Sapienza has 11 Faculties, 63 Departments and various research centres that drive high levels of excellence in archaeology, physics and astrophysics, as well as humanities and cultural heritage, environmental studies, nano-technology, cellular and gene therapy, design and aerospace.

DEPARTMENTS INVOLVED

Department of the Earth Sciences (DES)

<http://www.dst.uniroma1.it/en/>

The goal of DES is to offer high level education and to produce and guarantee high quality scientific research. Earth Sciences embrace a large number of disciplines concerning three main topics: Georesources and Energy, Environment, Risk. These topics are treated with the aim to provide the basic knowledge of the evolution of planet Earth and in particular: structure of matter at the atomic scale, Earth's interior and continental dynamics, causes of volcanic eruptions and earthquakes, earth-surface processes, water resources protection and management, life evolution, global changes.

CONTACT PERSON: Marco M. Scuderi, Ph.D. Marie-Curie Fellowship with the project FEAT (Fluid pressure in EArthquake Triggering).

PROJECT DESCRIPTION.

FEAT. The project aims to document the evolution of fault friction parameters as a function of fluid overpressure using a world-class rock deformation apparatus. The laboratory experiments are build on the characterization of fault zone structure, fluid flow, and deformation processes, which is reconstructed from careful field evaluations of ancient faults that represent exhumed analogues of seismically active structures. An important part of this work is the interaction with the energy industry to investigate the role of fluids in induced seismicity.