ERA in Action on open access to research data: key findings from the surveys and the experts' discussion

Context

On 17 July 2012 the European Commission (EC) published the communication **'Towards better access to scientific information: Boosting the benefits of public investments in research'**. Within this communication the EC made explicit reference to Open Access (OA) to research data. The EC has publicly communicated its plans to set up **a pilot scheme** on OA to and re-use of research data generated by projects **in selected areas of Horizon 2020**. The EC will also encourage, where appropriate, the publication of software codes used to produce or process the data.

In designing and implementing the pilot actions the EC will take into account possible constraints on making research data openly accessible, which may pertain to privacy, national security of data, confidentiality and know-how and knowledge brought into projects as inputs. Generally, the pilot scheme will not apply to projects whose primary aim would be contradicted by making research data accessible. However, the EC indicates that it is still looking for ideas how to overcome obstacles to the implementation of OA to research data. In addition, the EC has set a target for member states, namely that "by 2014, policies for OA to scientific articles and data will have been established in all member states at all relevant levels". Hence, the EC invites research agencies and funding bodies to develop policies on OA.

The Informal Group of Research and Technological Development Liaison Offices (IGLO)[1] has produced a survey on the state of implementation of OA in Europe and collected data from fifteen individual experts at national funding agencies, universities and libraries across Europe prior to a collective discussion in Brussels, in which 13 experts from 9 countries participated, along with representatives from major stakeholder organisations and industry invited as observers. The situation varies across member states and countries associated with Horizon 2020. While some organisations already have a strategy, policies or mature infrastructures for OA to data in place, others are still in the early strategy formulation phase. Nevertheless, all respondents and participants in the event have identified OA to research data as a priority area to follow in the years to come. OA to data is necessary in order to produce a competitive advantage for Europe and its institutions, to innovate and to safeguard good scientific practice, to address major societal challenges, and to account for the public investment in research.

This paper presents the main findings of both the survey and the discussions in the form of main challenges and key recommendations for the implementation of OA to research data in Horizon 2020.

---

[1] *IGLO is an informal association of Brussels-based non-profit R&D Liaison Offices. The aim of IGLO is to facilitate and enhance the interaction, information exchange and co-operation between Members of IGLO, their national research systems and the European institutions on issues related to EU RTD, in particular, the Framework Programme.*

The main issues identified by the experts can be classified in the following six challenges:

1. Cultural challenges

Many organisations were concerned with **the low data sharing culture** among active researchers in some fields. The first group that should be convinced about the benefits of OA to research data are the researchers themselves. As infrastructures for data sharing are established in many countries, the major challenge is to create a culture of sharing. A key concern is convincing the researchers that sharing data is in their own best interest and enhances their reputation. As a result, not sharing data should generally be seen as a well-underpinned exception and considered as a competitive disadvantage.

The benefits of sharing **vary from discipline to discipline** and have to become clear to all researchers and their institutions. The same applies to the notion of "openness", which can refer both to view and re-use and should be clearly and reasonably defined with regard to stages of the research process, the types of data and intended uses. As the efforts for data curation occur at the level of researcher or institution and vary with regard to purpose (verification/replication of data versus re-use by research and society), the necessary degree of openness should be carefully considered. As a principle all publicly funded data should be openly accessible, but specific barriers can be in place, such as confidentiality and privacy issues. Others have to be removed, such as impossible anonymisation of data, proprietary software to use the data and the lack of metadata. The definition of reasonable embargo periods may help to create an atmosphere of trust also among researchers.

2. Organisational challenges

The **lack of clarity** on the responsibilities and competences for policies on OA to data further hinders the implementation of OA at national and institutional level. This uncertainty results in a lack of coordination and communication between the different stakeholders at European and national levels. This trend is strengthened by the existence of different data management levels at organisations. The question on the authority deciding which data is to be made public or not has raised some concerns as well. Alliances between key stakeholders have to be created, also at the level of individual institutions. It has been observed that research data management also requires a broad buy-in from all stakeholders on national and institutional levels. The involvement of all relevant partners in the ecosystem already at the beginning of the research process is key.

The delimitation between useless and useful data is still to be defined but should be driven by the research communities.

The experts expressed concerns about unnecessary **administrative burdens** on research and innovation institutions resulting from an inadequate and too rigid implementation of OA. The question was raised as well whether OA to data is indeed desirable for all disciplines. Existing strategies for the implementation of OA explicitly call for different definitions and mechanisms across disciplines.

3. Legal challenges

The most important legal challenge that needs to be addressed is the problem of definition and the legal uncertainty this implies. The term 'research data' is very vague and does not have the same meaning for different disciplines. Moreover, the terminology around OA to research data needs to be further refined in order to allow an efficient and useful exchange between researchers.

The **lack of common legal regulations and standards** at European and national level also hinders an efficient implementation of OA to research data. Finally, the question of the use

and reuse of data still need further discussion, especially with regards to licensing, Intellectual Property Rights (IPR) and the Data Protection Regulation and the Database Directive. A challenge is the fact that research data do not necessarily fall under IPR protection, whereas the software needed to access and use it might do so. Therefore, open access has to be enabled on different technical and organisational layers of the infrastructure, while the use of open software should be encouraged.

4. Financial challenges
Various experts pointed out the financial challenges and the lack of clarity on who will support the costs of an open data policy. While there are initiatives at European level, they are still under development. Institutions need to know clearly how much financial support will be available from the European Union. The potential infrastructure costs could be huge. Some experts proposed to come with (financial) incentives for researchers to open up their data sets to others.

5. Ethical challenges
The sensitivity of data has been addressed by many. There is a general fear on the side of research institutions for data misuse for commercialisation purposes. Privacy issues and data protection also need to be sorted out, especially for health research and clinical trials. The new Data Protection Regulation should not impede research.

6. Technical challenges
Finally, technical issues need to be solved. Some institutions lack the capacity to archive data and make them available to others. Solutions must be created at European level in order to allow efficient storage and simple access to research data. The question on the duration of data storage needs to be addressed as well. Standards and formats should be in place that allow for the interdisciplinary re-use of data.

Recommendations
There was an overall positive reaction to the idea of a pilot on Open Access to data in Horizon 2020 among the participants to the session. The pilot would give a chance to fine-tune expectations, regulations, workflows and processes with regard to data management and curation. The evolution of open data fits in with the idea of change in scholarly communication through open science. Nevertheless, the following points need special attention when it comes to develop and implement a policy for Open Access to research data**:**

1. The concept of research data varies across disciplines. It is therefore recommended that **funders apply a generic definition** (e.g. "research data are factual records, which may take the form of numbers, symbols, text, images or sounds, that are commonly accepted in the research community as necessary to validate research findings"[2]), while research communities develop different and precise terminologies for their disciplines, based on traditions and formats. The European Commission could play an enabling and facilitating role by establishing adequate framework conditions, by creating a level playing field for all disciplines and by promoting open access to research data while clearly delineating possible restrictions. The trustworthiness of the process is key, and it is important to **allow for a bottom-up development** and sensibly mix top-down and bottom-up approaches.

2. Opening research data should be the result of a step-by-step approach. The concept of "intelligent openness" is key here. Data has to be accessible, but also assessable, intelligible, re-useable and referable. Open Access to publicly funded data should not

---

[2] Robello, R, Griffith University, ANDS Guide, 2012.

prevent research. For that reason, the goal of data sharing should be clear, since data curation for access requires less adjustment efforts than for re-use in an interdisciplinary and societal context. A process with different steps for opening up data is recommended. For practical reasons, **research data underlying publications should be made open at first.**

3. In some disciplines there can be a low data sharing culture. Therefore it is advised to **create a culture among researchers** where data sharing is supported and commonplace rather than the exception. In climate change and environment, for example, data sharing is widespread. Researchers collect and share standardised and interoperable data. These examples can be taken as best practices. Champions of data sharing should be supported across disciplines. Moreover, there should **be incentives** for researchers to share their data. Incentives could take the form of:
    a. Including **Data Management Plans** (DMP) in research proposals, which could also help tackle legal, ethical and technical challenges.
    b. Linking reward systems to **career development** by, for example, encouraging the citation of data sets and base scientific evaluation not only on journal publications but also on the creation and sharing of open research data sets.
    c. **Developing training activities and tools** for researchers to help them develop skills, to raise awareness and to ease the data sharing process.
    d. **Disciplines need to be supported and encouraged to develop adequate codes of conduct** for sharing and re-use of their data.

Open Access to research data is still a relative new theme that deserves positive attention. Therefore, **awareness campaigns on the benefits of open access to research data** would be welcome, to show governments, the research and business community the added value of making research data publicly available.

4. A **mapping of infrastructures and repositories** should be carried out to be aware of where research data are stored. The European Commission should thus **facilitate the discoverability and re-usability** of research data. With regard to technical standards, interoperability of metadata and access to software should be encouraged.

5. Where necessary, agreements should be established containing specifications about data ownership, access and use. The situation with regard to rights should be as clear as possible considering access and re-use of data. If licenses are to be used, they should be as open as possible. With regard to other data relevant for research and society, different restrictions might apply, taking into account secrecy, privacy regulations and commercial interests. These **restrictions should be clarified** and included as exemption to the general rule of openness in EC regulations. To satisfy legitimate research innovation or other interests, an appropriately short embargo period on the opening of research data could be envisaged.

6. Open access to research data should not create additional **administrative burdens**. Developing tools at European level helping researchers and providing advice could make a difference. We recommend careful consideration of the effects of new policies and funding flows on the workflows for data production, management and curation as well as on the efforts of stakeholders involved. It is important to support and strengthen the collaboration between all research data stakeholders in the ecosystem.

7. It should be clarified **who will cover the costs for infrastructures and who supports the costs of an Open Data policy**. Opening data must not become an additional burden for researchers, let alone a financial one. We recommend to ear-mark funds in research

projects for data management. At the European level, reimbursement of costs linked with data management has to be made possible. Concerning the overall data infrastructure, it is important to **support governments/organisations** in setting up institutions that can archive data. At the moment there is a lack of institutions that can archive data and make them available. Infrastructure should be built on different levels, including a central European one.

8. We recommend to include a proper definition of **goals and objectives** of open access to data that takes into account the complexity of re-use. We also advise a **careful evaluation** and monitoring of the Horizon 2020 pilot to research data, taking into account experiences with the OA to publication pilot as well as experiences with patent regulations.

Please note that this document does not constitute any formal opinion nor publication of IGLO.