ODS

Open Digital Science

(SMART 2014/0007)

Final Study Report

by

E. Prem, F.S.Sanz, M. Lindorfer, D. Lampert, J. Irran

eutema GmbH (Austria)

eu|te|ma
TECHNOLOGY MANAGEMENT

in co-operation with ZSI and Universidad de Zaragoza

ZSI

Universidad Zaragoza
1542

Vienna, May 2016

Abstract

Open Digital Science (ODS) and **Open Science** describe new and open practices in science, research and innovation that make extensive use of digital technologies. The use of digital technologies facilitates openness regarding data, methods, results, actors or publications with an emphasis on scalability of the approach in terms of data, access or computation. The vision underlying this study is to explore whether radically different scientific practices based on digital technologies are emerging, what they consist of, and how they are changing the relation of science and society.

The impact of Open Science practices is most evident in scientific publication (Open Access). A new generation of researchers uses digital tools in practically all steps of the scientific workflow, from research funding to critical discourse. This has led to concepts such as Citizen Science, Open Innovation, Open Methodology and Open Notebook Science, for which good practice examples are described. New ways of assessing scientific merit (altmetrics) and of funding (e.g. crowd sourcing) are also emerging. Six futuristic scenarios developed in this study depict potential impact of new open science practices. The study concludes with a set of indicators to measure open science implementation and to create an Open Science Observatory.

About the authors

Erich Prem received the graduate engineering and the Ph.D. degrees in computer science from the Technical University of Vienna, Wien, Austria, and the MBA degree in general management from Donau University, Krems, Austria. He is currently the CEO of eutema GmbH, Vienna, Austria. He is also a Lecturer at the Vienna University of Technology. He was a Researcher at the Austrian Research Institute for Artificial Intelligence, and a Guest Researcher at the Massachusetts Institute of Technology. He has a scientific background in artificial intelligence and epistemology. For the last 15 years, he has been involved in research and innovation management and policy with a focus on information and communications technology (ICT) research strategy. He is the author or co-author of more than 70 scientific articles. Dr. Prem is an experienced coordinator of many international projects in ICT research and technology management.

Fermin Serrano Sanz is researcher of the Institute for Biocomputation and Complex Systems Physics of the University of Zaragoza, and executive director of the Ibercivis Foundation. With a background on computer science and innovation management, Fermin currently deploys participatory projects of citizen science promoting public engagement. Since 2006 Fermin has collaborated with more than 40 research groups reaching almost 50.000 volunteers. Fermin is the coordinator of the Socientize project which delivered the White Paper on Citizen Science for Europe. In the past, Fermin has worked in other research centres such as CSIC and CERN, and currently he collaborates with other institutions including the European Commission and Ars Electronica.

Martina Lindorfer holds a MA in international development and is employed at the Centre for Social Innovation (ZSI) since 2008. Currently she is project manager of different FP7 and H2020 projects and main contact person for cooperation with Latin America and the Caribbean. She has long standing experience in science policy analysis and in issues related to Science, Technology & Society (STS). Her methodological competences include impact assessment, indicator development and qualitative methods of empirical social research. In the last years she has also been teaching at the University of Vienna.

Dietmar Lampert has a background in computer science and international development studies. He has joined the Centre for Social Innovation (ZSI) in 2007 as researcher to strengthen the team responsible for evaluating CIR-CE, a national Austrian programme designed to support innovation-based co-operation between Austrian industry and partners from Central and Eastern Europe. The evaluation was based on observing developments over time (ongoing; interim, terminal, and ex-post to assess the sustainability) and utilised methods from social network analysis. Since then, Dietmar has been employing a wide range of quantitative and qualitative methods in a number of projects and evaluation endeavours, both on the programme and on the project level. He has been involved in stakeholder consultations since 2009, be it straight-forward citizen consultations or complex multi-year endeavours that employed a variety of foresight methods. His present research interests include foresight, bibliometrics, altmetrics, patent analysis, impact monitoring and evaluation, visualisations and visual data exploration, which led him to co-forming a scientometrics team at ZSI.

Jörg Irran is a graduate engineer (with honours) in informatics (Technical University Vienna). He is an experienced computing professional with a strong background in electronics, system administration, robotics and artificial intelligence. He was a Researcher at the Austrian Research Institute for Artificial Intelligence and participated in several EU funded projects. In the recent past he contributed to studies carried out by the EC such as on the future of Artemis and ENIAC and on Next Generation Computing. He was involved in the management of large microelectronics projects funded by ENIAC / ECSEL and in multi KET pilot production assessment and demonstration activities.

TABLE OF CONTENT

1 Executive Summary

Open Digital Science (ODS) or, more commonly, **Open Science,** describes new and open practices in science, research and innovation that make extensive use of digital technologies. The use of digital technologies facilitates openness regarding data, methods, results, actors or publications with an emphasis on scalability of the approach in terms of data, access or computation. The vision underlying this study is to explore whether radically different scientific practices based on digital technologies are emerging, what they consist of, and how they are changing the relation of science and society.

Objectives

The main task of this study was to explore the extent to which new scientific practices based on digital technologies are emerging, what they consist of, and how they are changing the relation of science and society. The study identifies players, analyses the impact of emerging new practices on science and society, and provides suggestions on how to guide and monitor the implementation of such a vision. It identifies quantitative and qualitative indicators and metrics to assess the uptake and impact of open science and proposes indicators on how to set up a permanent observatory. The study aims to analyse in particular the consequences of open and digital science in a wider societal and policy context. The three formal objectives of the study are

- Validation and fine-tuning of the open digital science vision
- Developing metrics of open science uptake and impact
- Describing best-practices in new emerging scientific practices

Methodology

The study results are based on desk research on the open science concept, on current practices and organisations and on previous work on metrics. An analysis of current trends was performed to gain an understanding of expected developments of open science. Twenty-six expert interviews provided insights into the open science concept, current practices and trends. Intermediate results were presented at the study website www.opendigitalscience.eu where a discussion forum also served to collect feedback from the scientific community. Workshops were held at scientific conferences and events in Graz, Lisbon, and Brussels. These workshops included discussions on the six scenarios and provided important refinements. Further workshops with the study's Advisory Board created important input and feedback. For the metrics task, an online questionnaire was created and sent to experts in the field. Preliminary results and conclusions were also published online to involve community feedback.

The Open (Digital) Science Concept

The terms of this study originally focused on the notion of '*open digital science' (ODS).* Although this concept emphasizes the role that digital technologies play without any doubt in the current transformation of scientific processes, it is neither broadly accepted nor is it widely used. We therefore simply use *'digital science'* whenever possible.

ODS is embedded in a global context where science is faced with increased fragmentation, internationalization and a strong emphasis on innovation and application. Apart from the open data movement, there are new trends in opening various points in the scientific workflow to a broader public. This has led to concepts such as citizen science, open innovation, and open access. But it also points to new ways of assessing scientific quality and merit (altmetrics) and new ways of funding research (e.g. crowd sourcing).

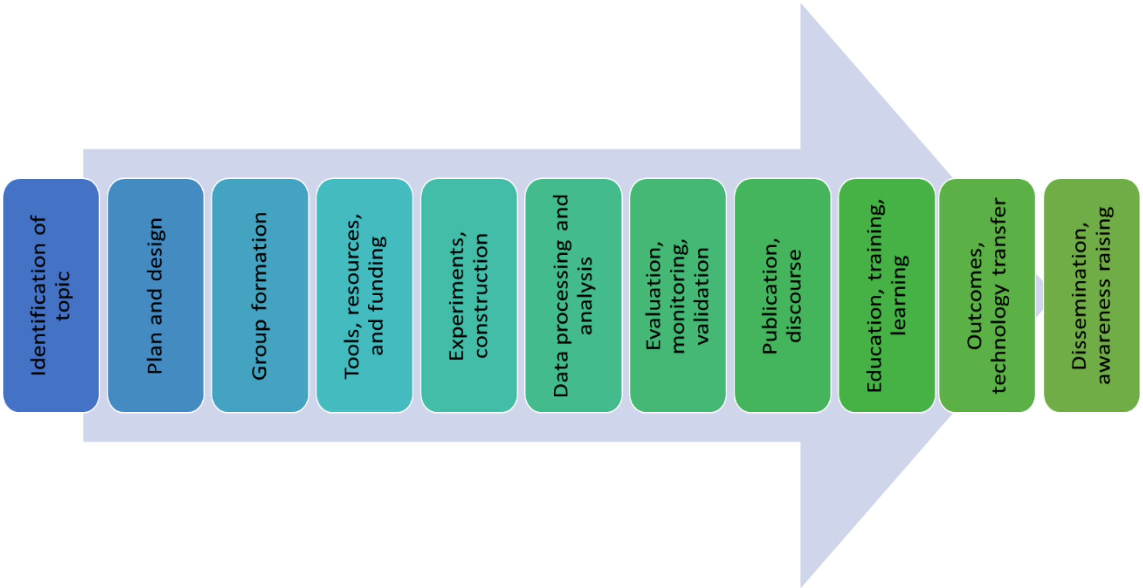


Figure 1: New and open scientific practices address every step in the scientific research process

ODS should be distinguished from other terms such as open science (which includes also being non-digital), digital science and science 2.0 (where the emphasis is on digital tools), e-science (which often focuses on high-performance computation and other e-infrastructure), citizen science (which engages the general public in research) and open access (which means online and usually free access to publications). The increasing importance of ODS and the related fields is evidenced by a growth of open data and open access sources as well as an increase in the use of the related terms by scientists world-wide. Being digital in nature, ODS shares important characteristics with information and communication technologies (ICT). This includes features of group-forming networks, zero marginal costs effects and the power of formal modelling and simulation, but also the bi-directionality of communication and the ability to work over long distances.

Drivers of ODS are not only new information and communication technologies, but also the general growth in digital data driven in turn by new and cheaper sensor technologies as well as cheaper and more powerful ICT. Increased networking of scientists with electronic tools and the general pressure to publish and to commercialize knowledge have also been cited as important drivers of ODS. Finally, changes in the self-image of researchers and the maturing of the 'digital native' are driving the future of the area.

ODS in Europe and worldwide

Analysing open science practices is a huge challenge as the field is evolving rapidly and a large number of projects, initiatives, and organisations are active and cover nearly all steps of the scientific workflow. The following table provides examples for each process step to exemplify current trends:

Process step	ICT usage	Examples
Topic selection, funding	Open peer review, crowd funding of research, online problem data bases and open innovation systems	Generic crowdfunding platforms supporting science: Kickstarter[1], Indiegogo, Goteo, RocketHub. Specific crowdfunding platforms for science: - https://experiment.com - https://ilovescience.es (ES) - https://walacea.com (UK) - https://fundscience.org.au/ (AU)
Data collection	Big data, new sensor systems, automated data collection from ‘internet-of-things’, laboratory robots, interaction with citizens	Safecast is a global project to empower people with data, primarily by mapping radiation levels and building a sensor network, enabling people to contribute and freely use the data collected. http://blog.safecast.org/
Data analysis, generation of hypotheses and theories	Artificial Intelligence-based and statistical methods of knowledge discovery, data mining, interactive and visual data analysis; new computing infra-structure (including shared, distributed computing), interaction with citizens	Cognition systems - http://www.ibm.com/smarterplanet/us/en/ibmwatson/ is an artificially intelligent computer system capable of answering questions posed in natural language Wikisurveys - http://www.allourideas.org/ dedicated to creating new ways of collecting social data
Cooperation, discussion, evaluation and critical reflection	New electronic forms of discussions, collection of micro-knowledge, interaction with citizens and artists, new metrics, reputation and re-commender systems	Open-source electronics prototyping platform based on flexible, easy-to-use hardware and software. It is intended for artists, designers, hobbyists and anyone interested in creating interactive objects or environments - arduino.cc is an open-source hardware openly available, allowing the Arduino boards to be manufactured by anyone easy to use and re use. It is leading to a community of sharing based on openness beyond open source software and hardware.
Publication	Open access publi-cation, open data (data re-use), open source software (software re-use), open metho-dology, collaborative writing, new media	- The Directory of Open Access Journals (DOAJ) is a website that lists open access journals maintained by Infrastructure Services for Open Access. It contains records for 10,000 journals[2] - http://arxiv.org is a repository of e-prints, of scientific papers with more than a million papers by the end of 2014. It is not peer reviewed but moderated. In 2014 the submission rate was more than 8,000 per month - ResearchGate is a social networking site for researchers to share papers, ask and answer questions, and find collaborators with more than 7 million users

[1]https://www.kickstarter.com/discover/tags/science

[2]Adams, Caralee (5 March 2015). "Directory of Open Access Journals introduces new standards to help community address quality concerns".

Process step	ICT usage	Examples
Other aspects	Open methodology, open educational resources	Public Labs[3] is a community of practice using inexpensive DIY techniques to change how people see the world in environmental, social, and political terms.

Table 1: ICT use in Digital Science

As digital technologies in general and digital tools for science and research are still rapidly developing, their impact on open science is likely to continue to be strong. A trend towards an even broader roll-out of in principle connectable and new devices directly impacts on the options for researchers and scientists to create new experiments. The trend towards distributed computing and cloud services is also likely to support scientists in managing data, processing, open and interactive processes and speed in the creation, management, and evaluation of scientific research projects. New tools will facilitate the virtualisation of processing over large, distributed processing networks that are dynamically created and perhaps continually adapted due to changing requirements. The tendency to store data where it is produced or to rely on the clouds will persist. However, we are also seeing limits of data storage capabilities and data is already becoming processed where it is produced (e.g. satellite data).

Perhaps more interesting, it is likely that large software and cloud service providers will also participate in the small- and large-scale analysis of data. Their tools are likely to become very powerful, yet comparatively easy to use. Already today, players like Google provide solutions for the statistical analysis of data, for visualization and search in large data bases. These applications do not always fulfil all requirements for critical researchers, e.g. neither are the working principles fully clear nor are the data bases easily available. However, the tools are freely available, powerful, reliable, and they are often equipped with visually powerful graphics. This makes them difficult to resist, even for the critical scientist.

Impact and uptake: scenarios of an Open Science future

To better understand the potential impact of digital science, a set of six scenarios was developed. These scenarios should not be read as predictions of the future. They describe aspects of what the future may bring and what might be necessary to make this happen – in particular with an emphasis on desired characteristics of such a future – and were used to stimulate the discussion with experts. The six scenarios address different aspects of open science:

- *ScienceFlex* describes a world in which Citizen Science platforms support the continuous cooperation of professional researchers and citizen scientists. It includes the vision of seamlessly integrated Open Data and tools for building low-cost sensors for the use in Citizen Science work.

- *InnoSpeed* discusses how open science can drive innovation, in particular for SMEs. It shows how ICT can help to improve access to and practical use of scientific knowledge for technically competent innovators who are not scientists. It addresses technical challenges from multilingual system support to semantic search and interoperability.

[3] http://publiclab.org/

- *BlurredBounds* describes a not-too distant future where the boundaries between academia and industry become nearly invisible as organisations dynamically create projects and recruit staff for virtual global project teams

- *Digital Studies* elaborates on how digital and open science changes the educational work of universities. It discusses the anticipated split into a few highly prestigious international branch campuses and specialized niche suppliers of education. New forms of teaching and a cultural change towards recognizing alternative educational pathways help Maria to combine being a mother with opening new career paths.

- *Policy Dialogue* offers insights into the past choices of future policy makers. It depicts two policy makers in the future who look back to their decisions today and the choices they made about open science incentives and policies.

- *Ancient Now* discusses open science challenges from the perspective of the technological future. These challenges include the broad problem of trust when networking with people, legacy systems and interoperability as well as metrics. In the scenario, a teacher talks about the many problems for open science collaboration in a past where many technology components were already there, but could not be easily combined.

Metrics – new indicators to measure progress in and towards Open Science

Interviews and literature review suggest that we have clearly entered a re-evaluation process of how research performance should be assessed. It is still highly unclear though, what the destination of this journey will be. The consensus amongst experts with regards to performance indicators is that merely considering the quantitative dimension of research performance and popularity/quality values (e.g. numbers of citations) is outdated and needs to be improved with indicators which take into account new opportunities offered by open and digital science. A set of possible indicators for various aspects of digital science was developed and discussed. In addition, 34 experts responded in a questionnaire.

The proposed indicators were structured in measuring various research aspects namely: conceptualisation & data gathering/creation, data analysis, diffusion (publication), review and evaluation, reputation system & trust, open science skills and awareness; and science with society. The following table provides an overview of core indicators judged to be useful to monitor open science implementation:

Indicator	Cluster
% of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards	Data Gathering
accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects	Data Gathering
% of machine-readable data / metadata	Data Gathering
quality of metadata (versioning, volume, data format, description of fields, etc.)	Data Gathering

Indicator	Cluster
availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research)	Data Gathering
usability of simulation results (models, data, and code)	Data Gathering
(types of) open data services offered	Data Gathering
is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no))	Data Gathering
% of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)	Diffusion
% of publications with free licencing (public domain, attribution, all kinds of sharing)	Diffusion
% of peer reviews that include reproducibility and transparency as review criteria	Review
data communication recognised as criterion for career progression (yes/no)	Reputation System
% of research personnel / research disciplines skilled in OS	Skills
% of research personnel active in OS	Skills
% of curricula that include OS skills (also prior to higher education)	Skills
% of research personnel aware of standards (is there a standard (relevant to open science), how to adhere to it, etc.)	Skills
% of research personnel familiar with those standards	Skills
# of researchers having signed an open science pledge	Skills
# of research organisations having signed an open science pledge	Skills
openness in call for proposals (open proposals, open submissions, open review)	Science & Society
increase in % of citizens engaging in open science	Science & Society
circulating and communicating research results outside the academia is standard (yes/no)	Science & Society
provision of affordable sets of public interest data / metadata	Science & Society

Table 2: Overview of core indicators

Conclusion

The dynamic nature of both digital technologies and open science makes it necessary to continue in-depth discussions on how to shape the future of open research. This includes coordination of policies and in some areas even the discussion on general objectives for policy-making. Given its open nature, these discussions need to be designed as participatory processes and it is recommendable that online tools play an important role here.

It will also be necessary to strengthen regulatory frameworks to develop a coherent and more harmonized environment. This regards, for example, a more coherent approach to open publication, but also open data. It is important to recognize the currently different approaches that exist not just across agencies and funding schemes, but also in the different scientific disciplines.

There is an opportunity now to actively shape robust and sustainable science with society engagement and industry implications as new tools and platforms are emerging and while the interest in this area is strong.

2 Introduction

As early as in the late 70ies, French philosopher Jean-François Lyotard[4] diagnosed a fundamental alteration of the status of knowledge as societies entered the post-industrial age and knowledge became increasingly computerized. Lyotard predicted the rise of knowledge as an informational commodity that for him also characterizes the postmodern condition of the Human mind. Today we are facing a situation where radically different scientific practices based on digital technologies are changing the relation of science and society, and where a generation of researchers is emerging who uses digital technologies in just about every step of the scientific process to broadly and often freely disseminate results.

Just how strong this influence really is, is not evident yet. While the trends towards free online provision of papers, reports, data and software tools is strong, it is not yet practiced in all fields of science and research at the same level. While physics, computer science and humanities publish strongly online, biomedical research or chemistry are sometimes said to lag behind. While many young researchers vastly exploit the opportunities of digital technologies, more established scholars still prefer traditional models. And while there are experiments in areas such as crowd funding of research, open peer review, and open methodology, these are far from common practice today.

Thus the question arises where we really stand in the development of open science and its support with digital technologies. But before we can answer this question, we need to understand the characteristics of digitally enabled open science – and we need to develop a perspective on how to best evaluate and assess the status quo in this domain. The study presented here aims to contribute to such an improved understanding of open science with a clearer vision of the field of digitally supported open science (and research), ways of measuring the state-of-affairs of open science, and of pointing out current practices and developments.

Open science evolution

The demand for science to be open can be traced to the beginnings of scientific inquiry. Indeed it could be argued that the requirement for science to be verifiable or – more modern - reproducible connects it with publication and therefore with openness. While there obviously exists research that is pursued in secrecy, there is a point where its results and methods need to be opened, for example in education and training, but also to maintain science's capacity for self-correction.[5] Also, science and research are largely social phenomena characterized by exchange of ideas, critical reflection, and competitive discourse. All of this makes it hard to see how completely 'closed' science could persist anywhere over extended periods.

Still, the apparent hype of 'open science' is well-founded. It is rooted in the explosion of digital technologies used in the scientific process.[6] Features such as shorter access times, usefulness of digital publications and databases, ease and speed of digital communication contribute to the expansion of open science. As soon as digital technologies were broadly available to the scientific community, they were used for the exchange of ideas, data, and publications. Early components and predecessors of the internet such as email, gopher, and ftp facilitated the timely and shared access to data and publications, but also software, mostly in academia.

[4] J.F. Lyotard, The postmodern condition. Manchester University Press 1979 (1984).

[5] H. Pfeiffenberger, Evolution of Open Science in Europe and the Helmholtz Association. http://epic.awi.de/39044/1/2015-09-24-RDA-PaNSIG-OA-Pfeiffenberger.pdf

[6] Sometimes, the term 'Science 2.0' is used to refer to increased digital technology use in science.

The broad availability of low-cost internet connectivity, large memory spaces, easy-to-use tools, and the diffusion to the business world and the broad public has led to a new scale of this exchange. It is also driven by enormous amounts of data often created automatically and sometimes even analysed automatically by computerized tools today. In addition, practices of work in the scientific community have changed. It has been suggested that today only less than 50% of researchers still use paper, the others use electronic tools on their desktop, and a small fraction uses only connected electronic devices.[7] All of this naturally creates the need to store, manage, access, and maintain electronic research data along with publications.

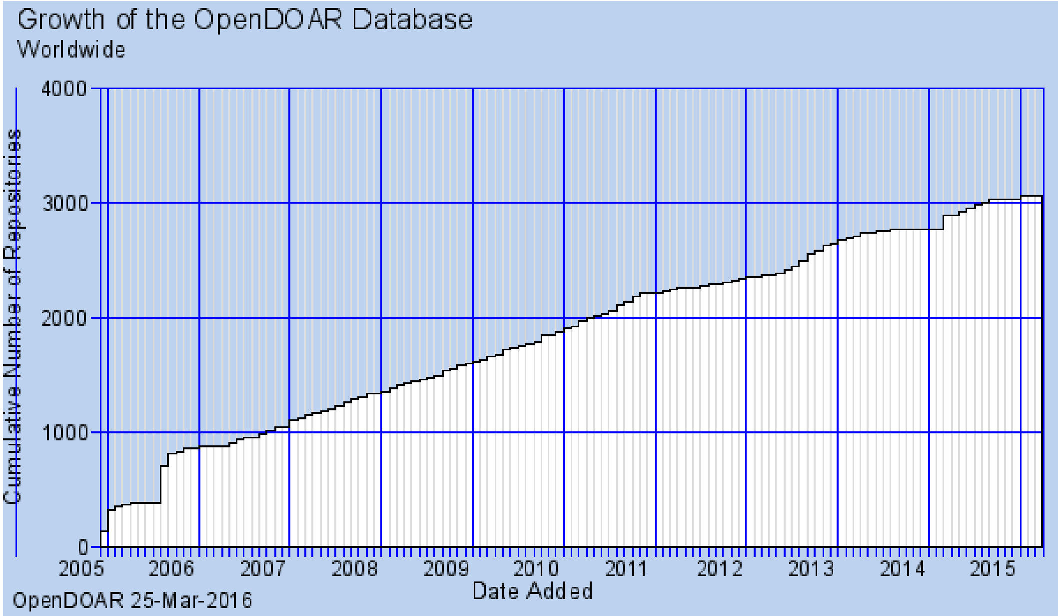


Figure 2: Growth of the OpenDOAR Database (DOAR is the Directory of Open Access Repositories, available at www.opendoar.org)

It is also important to understand that these trends not only impact on scientific work, but also on its commercial aspects – from publishing houses to scientific software houses and the large field of patenting. Publishing houses have reacted with different models for open publications – where ‘open’ does not necessarily imply ‘free’. But they also aim beyond just publishing papers with tailored services to manage research data including the provision of meta-data services. For a very long time, measuring publications and their impact was the most important way of indicating a researchers’ productivity and importance. This is now shifting to more and broader indicators including the publication of data, references made to research data etc. It is clear that this will change the power play in the publishing industry and it is not fully clear where this is going. Much will depend on the attitude and rules of universities, research organizations, public research agencies and other actors in the scientific research arena.

One trend that seems to be clear as of now is the increasing interest of a new generation of researchers in making the best use of digital technologies for research, publication, and dissemination of many aspects of their work. These aspects not just include research results and data, but also methods (‘open methods’), software (‘open software’), and even lab books[8] (‘open

[7] https://www.elsevier.com/connect/the-evolution-of-open-science-how-digitization-is-transforming-research

[8] Sanderson, K; Neylon, C (September 2008). "Data on display". Nature **455** (7211): 273. *doi:10.1038/455273a*

notebook science'). New (and perhaps more importantly old) generations of researchers now require additional training on digital tools for science, open science etc. New initiatives, including those of the European Commission[9] provide such training in the various elements. The challenges involved are not small, however, given that the whole area is still very much evolving and provides a moving target.

Finally, there are also concerns about the massive use of digital technologies in science that should be taken seriously. It has been argued that computation even threatens the scientific method as bug-free code is an impossibility.[10] Even using the same algorithms, programming languages, input data may produce different results depending on implementation details. Critics argue that open peer review creates skewed assessments; crowd-funding of research may result in underfunding important areas; and open publication leads to an avalanche of irrelevant papers online. These concerns – together with the potential advantages of digital tool use in research and open science – should provide sufficient reasons to motivate this study on open science, on digital science, and on how to measure it.

[9] e.g. FOSTER – facilitate open science training for European research. https://www.fosteropenscience.eu/

[10] L. Hatton, A. Giordani, Does computation threaten the scientific method. https://sciencenode.org/feature/does-computation-threaten-scientific-method.php

3 Objectives and methodology

The main task of this study was to explore the extent to which new scientific practices based on digital technologies are emerging, what they consist of, and how they are changing the relation of science and society. In particular, the study analysed and stimulated openness of scientific knowledge, by and for everyone to access, acquire, and benefit from. The study identifies main players, analyses the impact of the emerging new practices on science and on society, and provides suggestions on how to guide and monitor the implementation of such a vision. It identifies quantitative and qualitative indicators and a metrics to assess uptake and impact of open science and proposes indicators on how to set up observatory. The study aims to analyse in particular the consequences of open and digital science in a wider societal and policy context.

A note on terminology

The terms of this study originally focused on the notion of '*open digital science'.* Although this term emphasizes the role that digital technologies play without any doubt in the current transformation of scientific processes, it is neither broadly accepted nor is it widely used. In the course of this study it became clear that many experts found the term rather confusing. It was then decided to drop the emphasis on digital technologies from the term and rather operate with the more commonly known phrase of *'open science'.* However, the first part of the study still aims to analyse the specific relationship of digital technologies and science (and research). In the following, we will refer to ODS only when emphasizing this particular digital aspect, otherwise we use 'open science' for ease of understanding.

3.1 The objectives

The direct objectives underlying the ODS-project are to validate an ODS vision with precise definitions of the relevant terms. Furthermore metrics of ODS uptake and impact should be developed.

Objective 1 Validation and fine-tuning of the ODS vision

Objective 1 implies to first analyse the currently ongoing ICT-enabled transformed interactions in science, society and policy. This includes an analysis of worldwide trends, new practices, ongoing activities (funding programmes, institutional arrangements...). In going beyond a mere collection of items, the study should also analyse the role of technologies and infrastructures, and of course the data world.

ODS validation means to explore consequences in scientific, societal and policy context, for example in the area of innovation policies, social innovation, but also the impact of scientific knowledge on policy in relation to ODS. The tender takes a rather broad approach and includes aspects such as creativity and education. This is appropriate for an analysis of the ODS vision impact on science itself, but even more so on society. The ODS concept also touches on creativity, perhaps most clearly visible in creative, interactive, or participatory innovation. Other important areas of potential ODS concept impact are education, new forms of societal engagement and new forms of policy processes directly linking science, society and policy.

Objective 2 Metrics of ODS uptake and impact

The second main objective was to develop a set of indicators and impact assessment metrics, both quantitative and qualitative. The indicators should allow the EC and other policy makers to analyse the level of ODS uptake and to monitor ODS impact as an evolution over time. This will eventually lead to the creation of an observatory of ODS uptake and impact. To reach this goal, a broad engagement of stakeholders through an online community was required and part of the objective.

Examples for potential indicators are number of disciplines using ODS type of research practice, number of institutions committed to ODS type of activity, ODS related research programmes; examples for the latter are success stories of impact of ODS type of activities on innovation, policy, education etc. In our view, the presentation of indicators is an important feature as proper visualization can be a driver for all stakeholders.

Objective 3 Main players and best practices

The third core objective was to identify some main players and best practices in the EU and worldwide. This includes research programmes, public institutions, businesses, academia and research, social enterprises, creative forces, grassroots communities (and NGOs). This naturally overlaps with the data collection of Objective 1, but with a focus on best practices and policy issues.

The study also had a web presence including visualizations of communities, programmes, players, for example a map of institutions/funding bodies engaging in ODS, or scientific networks (using a crowd mapping approach where users can add themselves and suggest somebody else), cf. www.opendigitalscience.eu.

The online community supported this activity with contributions, feedback and refinement about practices, programmes and experiences.

3.2 Approach

The methodology was based on three distinct principles:

- DIGITAL: ODS are built on and enabled by digital technologies. We therefore need to understand the underlying technology characteristics and how they influence ODS. (E.g.: network effects, big data, openness, transparency, mobility.) These technologies are quickly evolving which implies that we need to base our understanding of ODS development on this evolution of digital technologies. It is important to understand technology evolution, but also the limits of information technology.
- SCENARIOS: The different aspects (e.g.: citizen science, open publishing, open data, big data, funding ODS, innovation technologies, etc.) of ODS are difficult to embrace in a single definition or description. We therefore facilitated the ODS vision validation using future ODS scenarios as an intermediary conceptual step. These scenarios about the future of ODS will help to communicate the ideas, discuss trends and their implications, understand issues and collect policy needs.
- OPEN: As ODS is all about openness and involvement of broad audiences, a wide-ranging stakeholder communication process was set-up. This involved the study website at www.opendigitalscience.eu and a range of presentations, discussions and workshops at conferences, e.g. ICT 2015 in Lisbon or the i-know conference in Graz.

Work was organized in the following sequence of steps:

1. Desk research on ODS concept, metrics, best-practices, programmes, policies, players - naturally building on existing work of all partners. This includes concrete success stories.
2. Conceptual analysis of ODS resulting in a clearly described refined current ODS vision
3. Analysis of digital science trends to better understand the technology characteristics inherited by ODS and to predict potential evolution of ODS in the near future from a technological point of view.
4. Interviews with experts from research, industry, policy and RTD management with the aim of getting input for all 3 objectives of the study
5. Proposing a set of initial metrics for discussion with experts
6. Describing best-practices and success stories
7. Development of future ODS scenarios (4-6) exemplifying ODS semantics, implications, trends and thus providing the necessary level of concreteness for detailed discussions with experts
8. Open online discussions on ODS scenarios, metrics and best-practices/success stories
9. Discussion in focus groups regarding the scenarios and including policy options
10. Description of a refined ODS scenario and vision, a sustainable metrics framework and best-practices for policies and success stories
11. Final validation workshop
12. Report finalization

Parallel activities over the whole process duration:

- EC communication in meetings and reports
- Public web presence on www.opendigitalscience.eu
- Consultation with the Advisory Group of the study

4 Open Digital Science

Open Digital Science (ODS) and **Open Science** describe new and open practices in science, research and innovation that make extensive use of digital technologies. The use of digital technologies facilitates openness regarding data, methods, results, actors or publications. Emphasis is on scalability of the approach in terms of data, access or computation.

ODS relies on the combined effects of technological development and cultural change towards collaboration and openness in research. This *digitally-enabled borderless transformation of science* and research allows new synergies and collaborations in knowledge sharing at large scales. As a result, new innovations, new questions, new solutions and new approaches arise from the involvement of researchers from all the disciplines with a wide range of contributors from all societal groups underpinning outcomes that benefit more efficiently both science and society.

In our opinion, there are two main points which are also well expressed in the term 'open and digital science':

- Digital: New scientific practices make massive use of digital information and communication technologies. While computers have played a significant role in science, research, and engineering ever since they were invented, there is now an unparalleled usage potential of digital technologies in all steps of the scientific process. Being digital opens an opportunity of scalability, i.e. to efficiently handle a large or growing amount of research by exploiting ICT tools.

- Open: New scientific results and practices have become much more accessible to other scientists and researchers across disciplines and also to people outside of the relatively small international science and research community – such as citizens or artists. Digital technologies facilitate the communication of scientific work and its results at very low costs, with minimum time delays and at high speed. Perhaps most importantly, digital technologies facilitate the self-guided, immediate and often unpaid access to interesting results, data, computer programmes etc.

4.1 The concept

The main question we face within this study is whether radically different scientific practices based on digital technologies are emerging, what they consist of, and how they are changing the relation of science and society. At European scale, we are observing a diverse set of initiatives promoting the uptake of ODS for research and innovation in the context of European Research Area[11] and Digital Agenda for Europe[12]. Just to mention some, we find Open Science, e-Infrastructures[13], Collective Awareness Platforms[14], Citizen Observatories[15], Global System Sciences[16], Digital Social Innovation[17], Responsible Research and Innovation[18], Smart Cities[19] and Citizen Science[20]. Moreover, each member state has different research funding channels related to open

[11] http://ec.europa.eu/research/era/index_en.htm
[12] http://ec.europa.eu/digital-agenda/
[13] https://ec.europa.eu/digital-agenda/en/e-infrastructures
[14] http://ec.europa.eu/digital-agenda/en/collective-awareness-platforms
[15] http://www.citizen-obs.eu/
[16] http://global-systems-science.eu/
[17] http://digitalsocial.eu/
[18] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation
[19] https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/citizen-science-and-smart-cities
[20] http://ec.europa.eu/digital-agenda/en/citizen-science

digital science, as well as some charitable organizations. Such a variety of drivers proves that ODS is gaining high relevance and interest in both the academic and innovation roadmaps.

This heterogeneous scenario leads to a recurrent concern which is the different understandings and values of “borderless science”. Open access facilitates research and makes the results more widely available and useful, playing a facilitator role for inter-, multi- and trans-disciplinary approaches leading to obtain new knowledge through the scientific method. The transformation of process and methodologies increase overlapping between science and research, bringing together new actors to explore new ways of working, innovating and knowing[21]. This paradigm shift is affecting also areas beyond science, finding emerging trends in areas so diverse such as humanities and arts, or economy and currencies, retailing among others. They have in common the impact in technologies and the human-centred approach[22].

The convergence of limitless access to broadband internet with new platforms and services is laying the foundations of new scales for data capture, store and computing power available[23]. These innovations require complementary attitudes where amateur contributors from the general public with specific micro knowledge or rare skills which can be used to enrich research[24]. This participation of people from outside established research organizations are shaping digital scenarios for organization and collective knowledge creation[25]. The nature of shared information is increasing notably in formats, from articles to methodologies, texts, data, images, audio, video, multimedia, and executable code[26]. Also, contributions from non-professionals are covering all aspects of scientific research, such as forming research questions, recording observations, analysing data, or using the resulting knowledge[27].

In this context, one key question for this study is how to measure different evolutions and dynamics of this cross-cutting transformation[28] where the entire world can play on the leading edge, and if these assessment tools can be applied to individuals or to larger organizations (institutions, companies, countries, societies). We even find some efforts addressing issues like monetary value of volunteer's unpaid participation[29]. However, different actors of the heterogeneous groups of stakeholders of ODS collaborate because of intrinsic, professional, educational or social motivations rather than only for mere financial motivations.

Knowledge is globalized when it is in principle globally available and accessible. The globalization of knowledge today has reached a new stage: it has transformed the economy of knowledge radically in ways that are comparable to the transformation in recent years of a monetary economy to a system in which local and global developments are coupled by almost instantaneous interactions. New potential for the globalization of knowledge has emerged, such as the global system of science and the World Wide Web, offering immediate worldwide access to the knowledge produced within this system. Due to the increased mobility of people and things, research hubs and the human resources of science have become global assets. The migration of scientific knowledge is no longer characterized by the trajectories of individuals or by the dynamics of fellow traveling, but rather by global social pattern.[30]

[21] http://www.ict-art-connect.eu/
[22] http://www.insiteproject.org/about/a-manifesto/
[23] https://www.eifonline.org/DigitalWorld2030
[24] http://socientize.eu/sites/default/files/white-paper.pdf
[25] http://caps2020.eu/wp-content/uploads/2014/07/CAPS_Handbook.pdf
[26] https://mitpress.mit.edu/sites/default/files/9780262517638_Open_Access_PDF_Version.pdf
[27] Citizen Science and Policy: A European Perspective by Muki Haklay
[28] https://ec.europa.eu/dgs/connect/en/content/open-digital-science
[29] http://www.pnas.org/content/112/3/679.full
[30] The Globalization of Knowledge in History Jürgen Renn licensed under a Creative Commons Attribution-Non Commercial-Share Alike 3.0 Germany (cc by-nc-sa 3.0) Licence.

```
Message-ID: <6484@cernvax.cern.ch>
Date: 6 Aug 91 14:56:20 GMT
From: timbl@info.cern.ch (Tim Berners-Lee)
Newsgroups: alt.hypertext
Subject: Re: Qualifiers on Hypertext links...

In article <64...@cernvax.cern.ch> I promised to post a short summary
of the WorldWideWeb project. Mail me with any queries.

                WorldWideWeb - Executive Summary

The WWW project merges the techniques of information retrieval and
hypertext to make an easy but powerful global information system.

The project started with the philosophy that much academic information
should be freely available to anyone. It aims to allow information
sharing within internationally dispersed teams, and the dissemination
of information by support groups.
```

Figure 3: From "short summary of the WorldWideWeb project"
(http://www.w3.org/People/Berners-Lee/1991/08/art-6487.txt)

ODS is embedded in today's context of scientific practices which are undergoing significant changes:

Increased knowledge fragmentation and internationalization

Technological developments and new discoveries allow the appearance of cascades of approaches. To address new challenges, more and more specialization of professionals is required. Big areas of knowledge like natural sciences or social sciences in fact hide multiple scales of disciplines, categories and sub categories with specific issues and tools. Economic developments allow countries to deploy site specific scientific agendas and internet-based infrastructures. As a result, the amount of scientific production is exploding world-wide. The Web of Science, which is the largest citation database, has more than one billion cited references from the year 1900 to the present[31]. The evolution of research topics is growing quickly in the last decade.

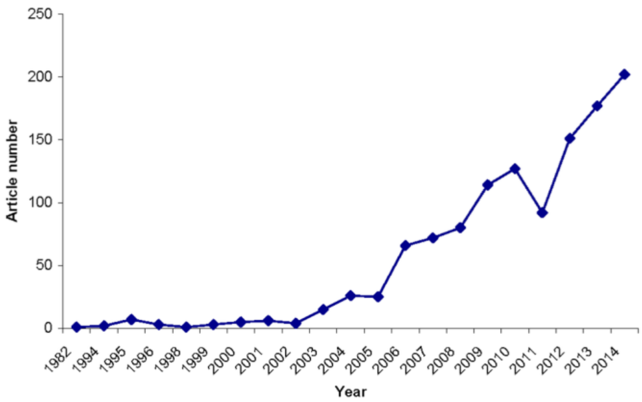


Figure 4: Number of scientific papers in the Web of Science including Open Access in the topic

[31] http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science.html

We are facing an unprecedented internationalization of research which is likely to induce sustained changes in the practices of science and research. Today, there are global archives with resources for scientific research (such as protein structural data or data on biodiversity) where data is provided, shared and reused by scientists all over the world[32]. Regarding co-authored articles, China have increased from 9.000 in 1998 until more than 74.000 in 2011[33]. The United States continues to play a central role in science networks but scientific excellence spreads. We find the top-30 high-impact research institutions spread over 14 different locations.

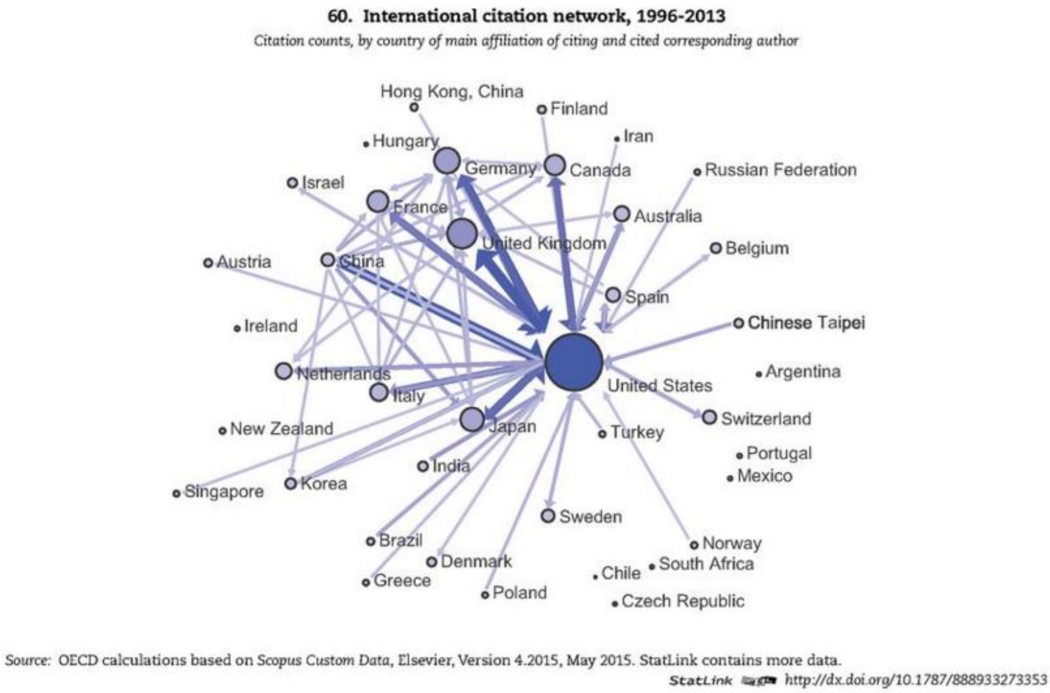


Figure 5: International citation network, 1996 - 2013

Application, data and impact orientation

Several authors point to the increasing application orientation in research and in particular in research management and research funding. This is both driver and driven by the commodification and internationalization of science and the capitalisation of knowledge[34].

As an example, the EU's Horizon 2020 programme reflects the policy priorities of the Europe 2020 strategy and addresses major concerns shared by citizens in Europe and elsewhere. A challenge-based approach aims to bring together resources and knowledge across different fields, technologies and disciplines, including social sciences and the humanities. This will cover

---

[32] The European Commission also performed a public consultation on Open Research Data: http://ec.europa.eu/research/science-society/document_library/pdf_06/report_2013-07-open_research_data-consultation.pdf

[33] OECD Science, Technology and Industry Scoreboard 2013 http://www.oecd-ilibrary.org/sites/sti_scoreboard-2013-en/01/04/index.html?itemId=/content/chapter/sti_scoreboard-2013-73-en

[34] A. Baskaran, R. Boden, Globalization and the Commodification of Science. In: M. Muchie, Li Xing, Globalisation, Inequalities and the Commodifciation of Life and Well-being. Adonis & Abbey, London, 2006.

activities from research to market with a new focus on innovation-related activities, such as piloting, demonstration, test-beds, and support for public procurement and market uptake. It will include establishing links with the activities of the European Innovation Partnerships. Funding will focus on the following challenges: Health, demographic change and wellbeing; Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the bio-economy; Secure, clean and efficient energy; Smart, green and integrated transport; Climate action, environment, resource efficiency and raw materials; Europe in a changing world - inclusive, innovative and reflective societies; Secure societies - protecting freedom and security of Europe and its citizens.

European Commissioners Günther H. Oettinger and Carlos Moedas published a joint blog post on the occasion of the conference Opening up to an ERA of Innovation[35] highlighting that Open Science is also about making sure that science serves innovation and growth. Developments in the context of the Digital Single Market[36] shall contribute to the common effort to make the EU more competitive and maintain excellence in science.

Complementarily, examples such as the development of Linux or the on-line encyclopaedia Wikipedia show how developers share knowledge, rules, sources and outcomes and thus quickly mature initial design efforts into first working prototypes and later fully elaborated systems.

4.1.1 Existing definitions of concepts included under the ODS umbrella

There is no broad agreement on these terms, although there is a fairly high degree of overlap in explanations used.[37]

- **Science** builds and organizes knowledge in the form of testable explanations and predictions. It is associated with the scientific method itself which consists of systematic observation, measurement, and experimentation, and the formulation, testing, and modification of hypotheses.
  In this study, science is used in its broadest sense, covering natural science, physics, mathematics, medicine as well as engineering, technical sciences and social sciences, and humanities. (A more adequate study title might thus be ODS-R: Open Digital Science and Research).

- **Open science**[38] is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society. By applying open source methods to science, open science is conducted in a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available with conditions that allow reuse, redistribution and reproduction of research. The rise of the internet has significantly lowered the cost and time required to publish or obtain data enabling transparent and transdisciplinary research.

[35] http://ec.europa.eu/commission/2014-2019/oettinger/blog/open-science-knowledge-and-data-driven-economy_en

[36] http://ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf

[37] Two glossaries with many of the terms mentioned here can be found at http://blogs.egu.eu/network/palaeoblog/2015/02/03/welcome-to-the-open-glossary/ and http://www.righttoresearch.org/resources/OpenResearchGlossary/

[38] https://en.wikipedia.org/wiki/Open_science

The European-funded project FOSTER Facilitate Open Science Training for European Research[39] has developed an open science taxonomy as an attempt to map the open science field, as depicted in the following Figure:

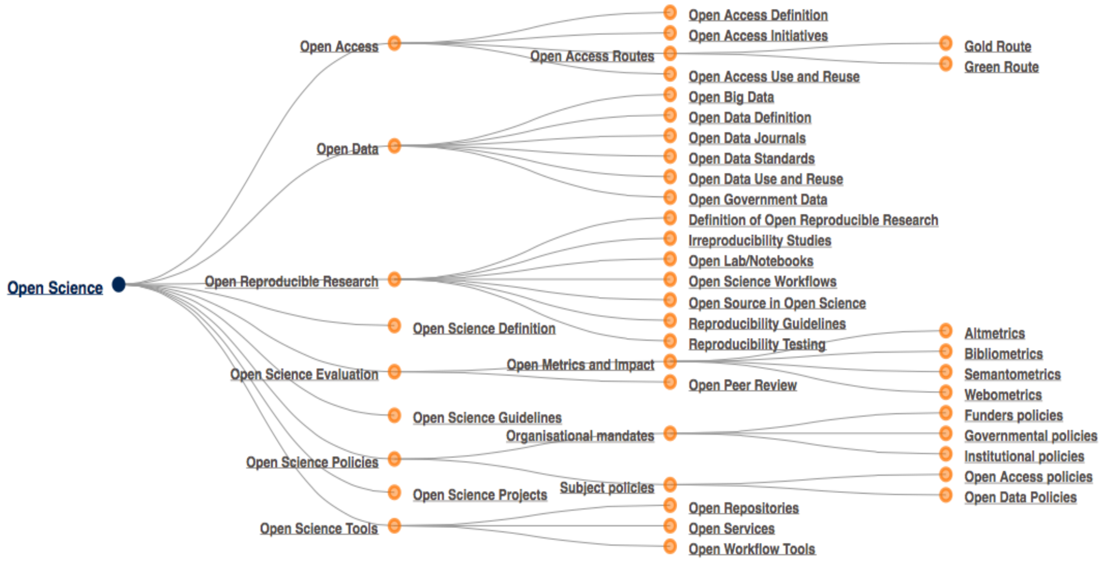


Figure 6: Open science taxonomy
(https://www.fosteropenscience.eu)

- **Digital science[40]** means a radical transformation of the nature of science and innovation due to the integration of ICT in the research process - new tools for scientific collaboration, experiments and analysis- and the internet culture of openness and sharing. This change is about the way research is carried out, disseminated, deployed and transformed by digital tools, networks and media. Digital science makes it possible not only to perform research more efficiently but to transform science changing the way scientific discoveries can take place and enabling faster and wider diffusion of scientific knowledge.

  At the same time, Digital science enables emergence of new scientific practices, disciplines and paradigms to respond to the new challenges through global distributed collaborations where citizens and society participate as contributors and direct beneficiaries of scientific knowledge. It is more open, more global and collaborative, more creative, and closer to society. It relies on the combined effects of technological development and cultural change towards collaboration and openness in research.

[39] https://www.fosteropenscience.eu
[40] Digital science in H2020, Concept Paper

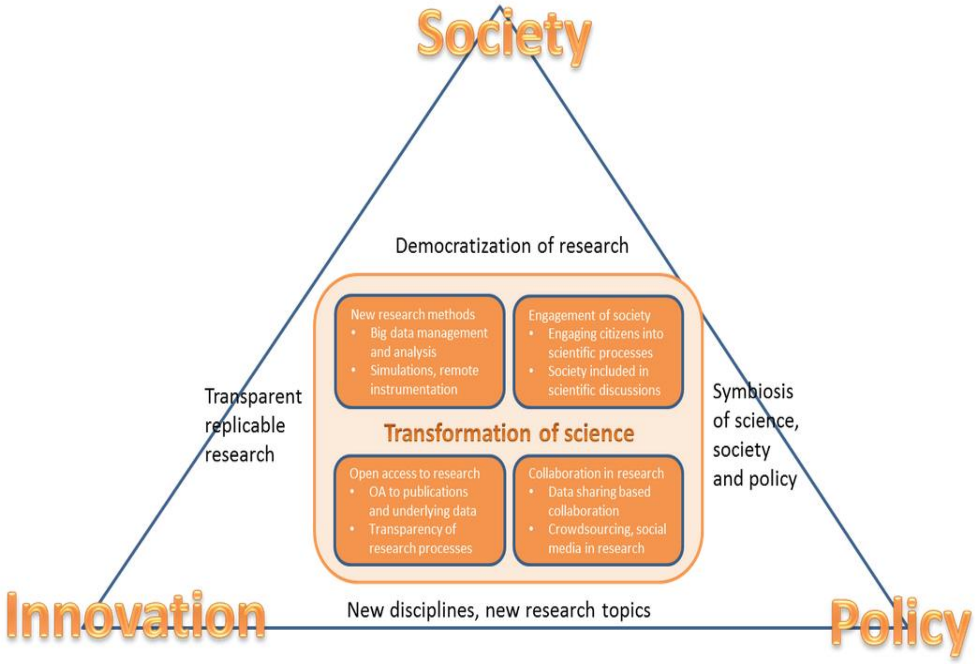


Figure 7: Vision for Digital Science
(http://ec.europa.eu/digital-agenda/en/news/digital-science-horizon-2020)

- **e-Science** means science increasingly carried out through distributed global collaborations enabled by the internet, using very large data collections, terascale computing resources and high performance visualization. Sometimes understood as "big science" it stresses the exploitation of expensive high-capacity/high-performance ICT facilities to address fundamental questions. As a result of the usage of the powerful new data, communication and computing possibilities science is being transformed by enabling virtual experiments which were not possible before. Data is generated on a large scale through images, sensors, simulations, logging of online human activities and all this can be stored for later processing. In this way it enables new research e.g. in social sciences and human behaviour.[41]

    - e-Infrastructure group42 is formed by official delegations of ministries of science from various European countries to coordinate activities with international initiatives outside of Europe. Its vision for the future is an open e-Infrastructure enabling flexible cooperation and optimal use of all electronically available resources.

    - eScience commons uses the metaphor of the Commons for the e-Infrastructure resources, which among others refer to networking, computing, storage, data and software, along with digital tools and collaboration opportunities. The e-Infrastructure Commons is the political, technological, and administrative framework for an easy and cost-effective shared use of distributed electronic resources across Europe.

[41] http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/geg-report.pdf
[42] http://e-irg.eu

> The project EGI43 is establishes the Data Commons through an Open Science Cloud under its Open Science Commons44 to freely share open research data, tools, applications, virtual laboratories and the related knowhow that the researchers

e-Science is not necessarily fully open as it mainly focus in the scientific community. However, e-IRG also foresees wide access and its recommendation also refer to data aspects that have to be fully taken into account and the data policy, including the data sharing rules, and the data life cycle which should contain sufficient information on the data (metadata) to enable global usage within the discipline, across disciplines, and in new research settings that could possibly not be envisaged at the time of creation of the data[45]. The following Table presents an overview of the services delivered by e-infrastructure providers and their access modes:

e-Infrastructure	GÉANT	EGI	PRACE	HELIX NEBULA	EUDAT	ZENODO	OpenAIRE
Services							
Network	x						
HTC		x					
HPC			x				
Cloud	x	x		x			
Data		x			x	x	x
Access Modes							
excellence driven		x	x				
market driven	x	x		x	x		
wide access	x	x			x	x	x

Table 3 Example of European e-Infrastructures and their providers, and which access mode is used[46]

- **Open source research[47]** is research conducted in the spirit of free and open source software. The central theme of open research is to make the methodology freely available via the internet, along with any data or results extracted or derived from them. Ideas are also shared openly and public discussion are promoted rather than private emails. This permits a massively distributed collaboration as anyone may participate at any level of the project.

  Projects that provide open data but do not offer open collaboration are often referred to as "open access" rather than open research. Providing open data is a necessary but not

[43] http://www.egi.eu
[44] https://documents.egi.eu/public/RetrieveFile?docid=2575&version=5&filename=OpenScienceCloud-EGI-v1.pdf
[45] http://e-irg.eu/documents/10920/277005/Best+Practices+for+the+use+of+e-Infrastructures+by+large-scale+research+infrastructures.pdf
[46] Source Best Practices for the use of e-Infrastructures by large-scale research infrastructures - http://e-irg.eu
[47] http://openwetware.org/wiki/Open_Source_Research

a sufficient condition for open research, because although the data may be used by anyone, there is no requirement for subsequent research to take place openly.

- **Citizen science** refers to the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources. Participants provide experimental data and facilities for researchers, raise new questions and co-create a new scientific culture. While adding value, volunteers acquire new learning and skills, and deeper understanding of the scientific work in an appealing way. As a result of this open, networked and trans-disciplinary scenario, science-society-policy interactions are improved leading to a more democratic research, based on evidence-informed decision making as is scientific research conducted, in whole or in part, by amateur or non-professional scientists.[48] Scientific work is sometimes crowdsourced and undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions[49]. Citizen science is growing in scale and scope as an important form of public participation in the scientific enterprise. It therefore receives increasing attention from policy makers at local, national, and international levels.[50]

  Some citizen science projects like Galaxyzoo or fold.it have delivered significant research outcomes, but these projects are distinct from those in which participants are able to influence the overall direction of the research, or in which participants are expected to have creative input into the science behind the project. However, the average value of contributions received per project in Zooniverse is over $200,000 the first 180 days.[51]

- **Science 2.0** describes the on-going evolution in the modus operandi of doing research and organising science resulting predominantly from a bottom-up process driven by the increasing number of researchers operating in a globally networked digital systems and driven by the globalisation of the scientific community, as well as the increasing societal demand to address the Grand Challenges of our times.[52] Science 2.0 refers to the usage of tools like wikis, blogs and video journals to share findings, raw data and "nascent theories" online. Science 2.0 combines hypothesis-based inquiries with social science methods, partially for the purpose of improving those new networks.

- **Open Access** refers to online, free of cost access to peer reviewed scientific content with limited copyright and licensing restrictions. According to the European Commission open access to scientific research data enhances data quality, reduces the need for duplication of research, speeds up scientific progress and helps to combat scientific fraud.[53] There are two routes to open access; the open access journals and the open access repositories or open archives. The Gold route to open access is delivered via publishing an article in a journal. The journal may be an open access journal (pure open access), or a subscription based journal (hybrid open access) that offers an open access option. The

[48] Green Paper on Citizen Science for Europe, Socientize.eu 2014
[49] Oxford English Dictionary. Retrieved 13 September 2014.
[50] Muki Haklay, Citizen Science and Policy: A European Perspective. Washington, DC: Woodrow Wilson International Center for Scholars, 2015.
[51] Crowd science user contribution patterns and their implications, Henry Sauermann, 679–684, doi: 10.1073/pnas.1408907112
[52]https://ec.europa.eu/research/consultations/science-2.0/background.pdf
[53]European Commission, Recommendation on access to and preservation of scientific information, C(2012) 4890 final, Brussels, 17 July 2012. http://ec.europa.eu/research/sciencesociety/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

Green route to open access is delivered via self-archiving an output into a repository (there are two types of repositories, institutional and subject repositories).

- Three leading statements established Open Access as a movement; Budapest Open Access Initiative[54]. Bethesda Statement on Open Access Publishing[55] and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities[56].

- OpenAIRE is the Open Access Infrastructure for Research in Europe aiming is to gather the metadata of research output to enable EU researchers, businesses and citizens to have free and open access to EU-funded research papers. It includes over 11.5 million open access documents from over 600 data providers. Next pictures show an evolution of the number of publications and OA adopted by publishers.

The following data gives an impression of the current status of open access. It lists publications by data source (top 20 only):

	Datasource	count of number
1	Europe PubMed Central	3205411
2	DOAJ-Articles	1821602
3	arXiv.org e-Print Archive	1051435
4	Research Papers in Economics	651996
5	CyberLeninka - Russian open access scientific library	514563
6	Hyper Article en Ligne	367096
7	INRIA a CCSD electronic archive server	344713
8	DSpace @ Cambridge	186838
9	Wageningen Yield	165902
10	TÜBİTAK ULAKBİM DergiPark	145377
11	Hindawi Publishing Corporation	133821
12	Hrčak - Portal of scientific journals of Croatia	125036
13	Project Euclid	119248
14	igitur archive - Utrecht University Repository	114188
15	Repository TU/e	104831
16	GREDOS	96640
17	Theseus	91536
18	EconStor	85987
19	National Repository of Grey Literature	78803
20	Other	69295

Table 4: Open access publications by data source

[54]http://www.budapestopenaccessinitiative.org
[55]http://legacy.earlham.edu/~peters/fos/bethesda.htm
[56]http://openaccess.mpg.de/Berlin-Declaration

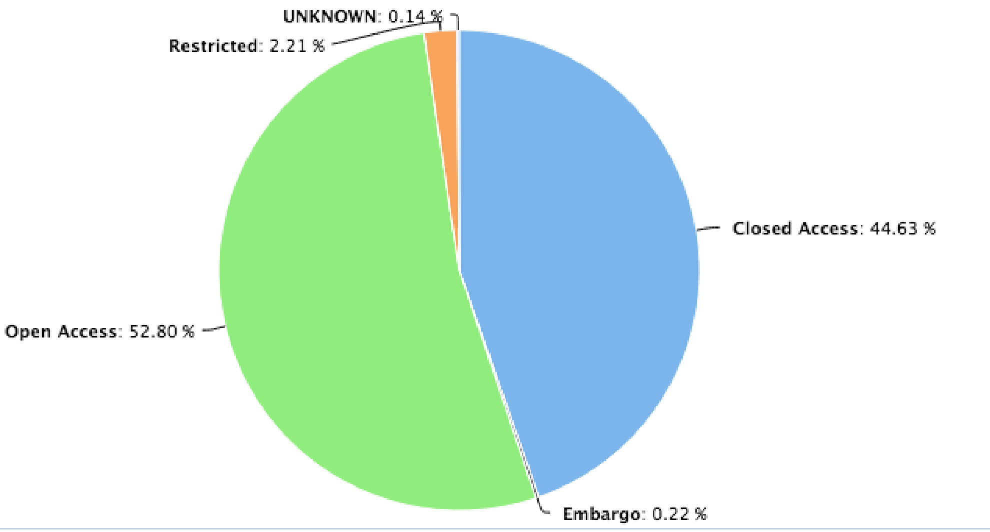


Figure 8: FP7 publications breakdown by access mode (OpenAIRE https://www.openaire.eu - captured September 2015)

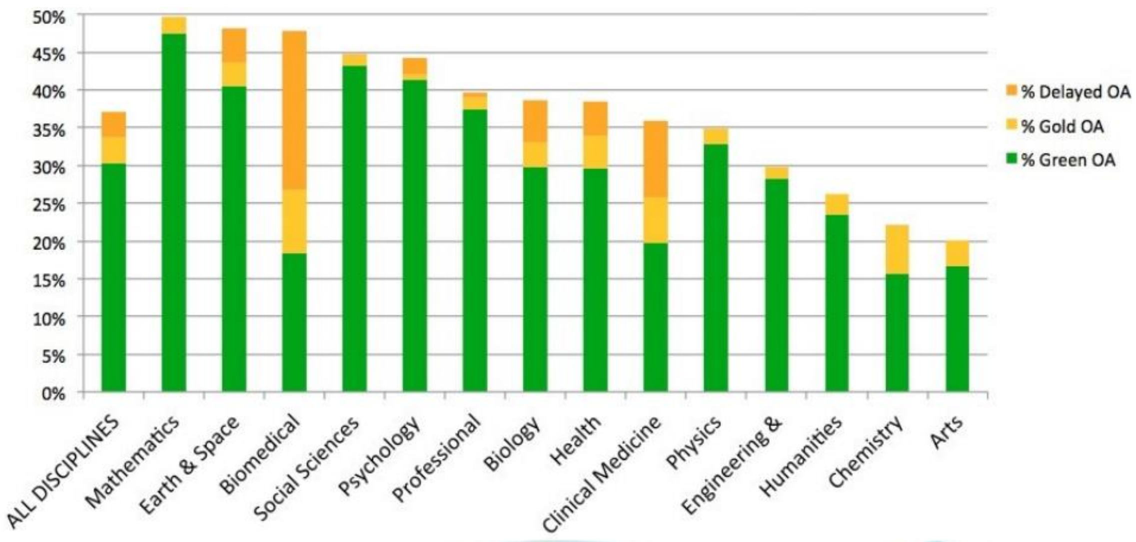


Figure 9: Percent OA by discipline (OpenAIRE https://www.openaire.eu - captured September 2015)

- **Open science and research data** is focused on publishing data from R&D which is defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings[57]. These contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia materials.[58] A research data set constitutes a systematic, partial representation of the

[57] OECD Principles and Guidelines for Access to Research Data from Public Funding http://www.oecd.org/sti/sci-tech/38500813.pdf

[58] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf.

subject being investigated, observations and results online, free of cost, of scientific activities available for anyone to analyse, to reuse and to distribute provided that the data source is attributed and shared alike.[59]

- The Human Genome Project[60] was a major initiative that exemplified the power of open data. It was built upon the so-called Bermuda Principles, stipulating that: "All human genomic sequence information (…) should be freely available and in the public domain in order to encourage research and development and to maximise its benefit to society'.

- The Dataverse[61] is an open source web application to share, preserve, cite, explore and analyse research data. It facilitates making data available to others, and allows you to replicate others' work. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive appropriate credit via a data citation with a persistent identifier (e.g., DOI, or Handle).

- RECODE[62] project identified two overarching issues in the mobilisation of open access to research data: a lack of a coherent open data ecosystem; and a lack of attention to the specificity of research practice, processes and data collections. As a result, this project recommended the development of aligned and comprehensive policies for open access to research data in Europe.

- **Open reproducible research** refers to free access to experimental elements for research reproduction. Including products and outcomes usually out of the scope of open research data, open reproducible research welcomes laboratory notebooks, preliminary analyses, and drafts of scientific papers, workflows, plans for future research, peer reviews, or personal communications with colleagues or physical objects. These experimental elements are offered online free of cost with terms that allow reuse and redistribution of the recorded material. Ground rules to assist with the recreation of research experiments and with the process of validating the research results are also provided.

  - The Research Ideas and Outcomes (RIO)[63] journal has been launched in September 2015 to publish all outputs of the research cycle, including: project proposals, data, methods, workflows, software, project reports and research articles together on a single collaborative platform, with transparent, open and public peer-review process.

The philosophy of Open Science goes far beyond just communication by opening the whole research process and results for peers and public, through digital media and collaboration efforts. Open reproducible allows stakeholder to acquire an holistic view of the scientific process along all the research cycle rather than

[59] Cf. Hodson, Simon and Molloay (2014): Current best-practice for research data management policies. http://www.zenodo.org/record/27872?ln=en
[60]http://www.genome.gov/10001772
[61]http://dataverse.org
[62]http://recodeproject.eu/
[63]http://riojournal.com

4.1.2 Other relevant topics related to open science

- **Linked Open Science**[64] is an approach to interconnect scientific assets as a combination of Linked Data [Datasets are encoded as Linked Data.], Semantic Web and Web standards, open source and Web-based online environments, Cloud Computing, and a machine-understandable technical and legal infrastructure.
- **Open peer review** is the process of peer validation conducted openly on the Internet revealing both authors and those who have reviewed the publications. Proponents of open peer review see its transparency as a way to encourage more civil and thoughtful reviewer comments — although others are concerned that it promotes a less critical attitude[65] and may show some deviations around close colleagues[66].
- **Open innovation science** refers to agreements of academic groups with companies which try to bring in the external ideas to complement in-house research. Rather than changing the nature of research, major motivations for companies to adopt open innovation are the decrease of costs while speeding up their product development cycle and the incorporation of external actors in the development process with a potential viral marketing.

  There is a conflict around patent issues and the rights of ideas and it is usual that in open innovation teams operate in secrete for instance in prizes for solutions to problems (e.g. Innocentive[67]).
- **Responsible research and innovation** also bridges the gap between the scientific community and society at large. RRI addresses the grand societal challenges arguing that they have better chance of being tackled if all societal actors are fully engaged in the co-construction of innovative solutions, products and services. Responsible Research and Innovation means that societal actors work together during the whole research and innovation process in order to better align both the process and its outcomes, with the values, needs and expectations of European society.
  - RRI Tools[68] is a project that aims empower all actors to promote not only excellent but also socially desirable science and technology. RRI Tools is developing a set of digital resources to advocate, train, disseminate and implement RRI under Horizon 2020.
- **Open metrics** are alternative ways to traditional impact metrics systems of evaluating the diverse impacts of the scholarly outputs.
  - Altmetrics reflect the broad, rapid impact that new, online scholarly tools allow the use of new filters to analyse academic literature and information collected from the Internet, such as social media sites, newspapers, and other sources. This matters because expressions of scholarship are becoming more diverse, ranging from raw

[64] Tomi Kauppinen, Alkyoni Baglatzi and Carsten Keßler. Linked Science: Interconnecting Scientific Assets. In Terence Critchlow and Kerstin Kleese-Van Dam (Eds.): Data Intensive Science. CRC Press, USA, 2012 (to appear).
[65] http://www.nature.com/news/nature-journals-offer-double-blind-review-1.16931
[66] Bias in Open Peer-Review: Evidence from the English Superior Courts, Jordi Blanes. *(2015)doi: 10.1093/jleo/ewv004*First published online: March 11, 2015
[67] http://www.innocentive.com/innovation-solutions/custom-challenge-programs
[68] http://www.rri-tools.eu

science with datasets, code, and experimental designs to nano-publication where the citeable units are arguments shared via blogging, microblogging, or even annotations on existing work rather than entire article[69].

- **Open science policy** aims to foster greater access to and use of scientific research. Through public policies and guidance from research funding agencies can facilitate the sharing of data resulting from publicly funded research. They can help research institutions better manage research data through the development of infrastructure and training. They can also provide guidance to researchers on compliance with the various policies governing data access and sharing (e.g. intellectual property rights, privacy and confidential issues).

- **Crowd funding of science** is not an entirely new phenomenon, but emerged as a new trend facilitated by the internet. Raising money for research from large crowds is particularly common in areas with clear social benefits, e.g. medicine. It remains to be seen where the limits are in terms of sustainability, research topics, and also regulation.[70]

4.1.3 Community perspectives on the open science concept

In the course of this study, the notion of open (digital) science was extensively discussed in interviews with researchers and research managers. Obviously, interviewees did not have a uniform understanding of the concept as it is still in its early stages. However, they all agree on a wide range of characteristics of the term. The following synthesis of the interviews can provide the reader with a clearer view of what is and what is not regarded as open science.

Two main recurring aspects are:

- a process of transforming (opening up, democratizing) science thanks to digital technologies.

- an intensive exchange environment of transparency (for transmit open knowledge to everybody) and participations Involving computers and brains at the same time, in real time. In this convergence, human focus on difficult tasks and automated tasks are left to the machines that need contents available and reachable. Some argue that ODS is semantics.

Status quo

BOOM OF PRACTICES

As several experts highlighted, open science and digital tools are currently expanding very rapidly and every day there are very exciting things happening. This in fact is one of the reasons why there is little agreement on terminology. Even *'open'* and *'digital'* are concepts with many different connotations subject of epistemological discussions. The open science ecosystem is formed by elements of a heterogeneous nature (such us technologies, practices, or players) around topics with multiple layers, different origins and intentions. The emerging virtual spaces can form a powerful digital infrastructure facilitating researcher blogs, online experiments, open labs, intranets, virtual research environments or social networks (like ResearchGate or even

[69] http://altmetrics.org/manifesto/

[70] R.E. Wheat, Y. Wang, J.E. Byrnes, J. Ranganathan, Raising money for scientific research through crowdfunding. In: Trends in Ecology & Evolution, Vol 28 (2), p. 71-72.

Facebook). It is pertinent to mention that several experts highlighted that the EU commission gives some orientation and a few experts explicitly mentioned the definition presented in the Digital Science concept paper.

ARE YOU REALLY OPEN?

Regarding what is not open science, discussions were more focused on practical issues and less theoretical regarding topics such as accessibility, ownership, engagement of social actors or the actual capacity of building up or disseminating knowledge. Scientists believe they do not yet produce many high-quality open science outputs. An important point is that published results may be understandable by both humans and machines and that they give insights not just in the outcomes but the whole research workflow.

Examples of bad practices are articles that nobody reads or when papers are just scanned. Also, hybrid proprietary models were mentioned as not practical. It was mentioned as a problem for citizen science when the social aspect is not fully considered and subjects only provide input in a one-directional fashion: for instance, Zooniverse is big and widely known, costly citizen science project. It is a big success, but it is not open. As a result, it is good for reaching people, but not good in empowering them. Consequently, it has a working business model for now, but may not sustainable one the sponsors leave.

WITH THE PEOPLE FOR THE PEOPLE

For many researchers, open science represents a change of relevance towards more balance between human and numerical sides. Working with a broad community can bring scientists closer to the people. In turn, open science offers knowledge to any person, similar to the open source movement. Everybody can be a content generator, mostly in international teams; but open science is a requirement for remote and isolated communities. There is also a big potential for policy making and responsible industries based on this. Open research projects are interesting for a better understanding of what it is going and informed policy-making. Research quality faces issues like data collection costs and management. Interpretation implies a shift of questions and of disciplines removing barriers between disciplines for cross fertilization. As an example, social media becomes embedded in behavioural data to be addressed by DSSH+big data (volume, variety, velocity, and veracity), humanistic/artistic/S&T research. Other trends such as miniaturization of technologies or the decrease of software development costs are helping a broader adoption of open science practices.

Opportunities

BETTER CHANCE TO ACHIEVE EXCELLENCE

The following positive effects of this hyper-connected scenario were recurrently stated:

- more efficient (faster, better updated, lower costs, more simultaneity, less duplication of efforts, more productivity, broader range of creative ideas)
- more transparent (more access, more rigour, more elaborated, better documented, better dissemination)
- better examination of research results (better quality control and better assessment, increase legitimacy)
- better conclusions (more data, broader participation, more questions, new approaches, new solutions, new results, new knowledge)

- larger impact (more beneficiaries, open innovation, better education and training, broader access to knowledge, more social)

Trends and uptake

MORE COLLABORATION, MIXED PERCEPTION

Everyday there are more and more agents, more and more technologies, more and more models of open science and these models are more open. In general, good reputation increases with open science work, but still the open practices are accepted only in a few fields such as astronomy, bioscience, crystallography... but not very accepted yet in other fields like chemistry, observational bioscience, laws, engineering, social sciences or health. It is fairly new that external actors outside academia like makers or artists are deeply involved in openness which is a major trend itself. However, in many cases the latter predominant create social or educational impacts rather scientific ones.

SCIENTIFIC ACCEPTANCE

The community states that there still is a lot of confusion about openness and scientists who do not make an effort do not understand it. Also, open science is still a long way from the practices of many researchers today, especially in areas where most groups (or individuals) work with small data sets. Most of them keep the data to themselves until publication. This practice will be hard to change. Some time ago, they used to have even their own computational tools. They had reservations to using shared resources which is slowly changing. There is a trend of scientists who are becoming more concerned with better data and open access to research data and publications, perhaps driven also by public institutions and funding agencies.

Citizen science as special topic in this area is not yet very broad, despite of its history. But most scientists have not used it yet and many are sceptical about data quality and other aspects. Similarly, altmetrics are considered important, but institutions still have reservation about broad applicability in not academia. Recognition systems focus more on peer review rather than other skills.

Among the skills needed, interviewees mentioned a vast set of issues like world-wide cooperation features, better visualization and data analysis capabilities, comprehensive narratives and technical infrastructure management.

Challenges

CULTURAL ENVIRONMENT AND MINDSETS

Current R&D systems D encourage competition more than collaborations. Researcher need money and orientate at how funding is provided (according to number of publications or patents). Few established researcher have the vision of win-win open science and early career understand it. They can become but they are frustrated where science funders do not see the need or its advantages. Open science may represent a loss of operational capacity and, because of that, a loss of spontaneity. Tasks such as data preparation for sharing, public engagement activities or the involvement in scientific policy are time and cost intensive and bring little scientific reputation. As a result, many actors keep their data for themselves. This tradition of self-replication even has the danger of leading the scientific community away from new directions and into closed communities.

BALANCING DEMOCRATIZATION

Democratizing science is also seen as a control mechanism. There will be more and more demands for accountability of researchers, but this will also increase pressure. Transparency can also decrease quality of research (e.g. research on unpopular or contentious matters) and critical discourses can be stopped at an early stage through very early broad participation. Participation can also lead to ruining fundamental dissent from the academic side. The challenge is make more actors to fully participate (institutions, groups, individuals) beyond current specifics groups. New profiles and new environments are required to support this work, mixing people and technologies from different backgrounds.

OPEN ISSUES: PRIVACY, IPR, EXPLOITATION, ODS ECONOMY, EXPLOSION OF DATA

Most interviewees mentioned confidential data (personal data, medical data, privacy) as an important challenge and large unresolved issue. Also mentioned were ownership of inventions, current models of publishing, the role of companies in partially funded research, and licenses. These issues depend very much on the research area and should actually be very critically examined and regulated. There is some discussion, but there are contradictions between public and market interests and little practical support for researchers today.

There is a need of better and clearer explanations on data issues from the research side. There is a need of monitoring the different aspects of open science, for example IPR along all the life-cycle of data.

There are very few references about economic impact studies of open science related practices, but it is agreed that the economic impact could be huge. In particular, this may be true for research that is never actually exploited. It was mentioned, for example that 85% of medical research is wasted as it is never properly shared (open practices could also help avoiding unwanted replications and recurring errors).

It was mentioned that the current explosion of publications – open or traditional – requires new tools including digital ones to fully embrace and exploit the potential.

EU weaknesses and strengths

EUROPEAN OPEN SCIENCE MINDS

Researchers in our interviews believed that Europe's citizens are civic/critic community members. There is a stronger focus on public goods in Europe than elsewhere and also on public funding. This is not yet fully exploited for public good. European citizens are more participatory, more activist, more creative and more humanist with a strong traditional culture of sharing compared to many other regions in the world. The public education level is usually high. Top institutions like CERN, ITER. EU central (EC) and national funding mechanisms play a leading role in the society where the science culture is historically grounded. The general public should not just be the provider of scientific data, but also a strong user of knowledge. The EU should step up its efforts to mobilise its resources to be able to compete. EGI and other similar forums like GEANT are very active and there are other pillars that can contribute to the formation of the open science ecosystem.

AREAS IN NEED OF CHANGE

Institutional attitudes sometimes decrease participation and culture of sharing. There are not many companies in Europe to manage content, we basically depend on companies from the USA (e.g. Google). Moreover, the EU focuses on the less sexy part (regulations). There are also intra-EU differences and overall a lack of harmonization (e.g. UK closer to US) regarding issues such as copyright. To flourish in open science, we need a copyright. In the universities, established researchers often do not give sufficient room for younger colleagues. Although these issues are addressed by policy actions, the current post-crisis situation still decreases public funding as well as employment. There is also some criticism among experts regarding H2020 which is well considered (e.g. funding programmes like SwafS) but sometimes it is still considered as a manual which limits creativity. EU support is very fragmented and the structure of open calls is very inefficient (very costly in time and money with extremely low rates of success). At national level, funding opportunities related to open science are considered sparse. Crowdfunding represents an alternative with some potential, but again only few countries and organisations have put in place clear rules.

4.2 ODS in the stages of science

Science and research are human activities that aim to understand the world, predict and control it, and to make a living. Much like the use of information and communication technologies (ICT) have been changing the lives of practically everybody, ICT is also changing practices of scientists and researchers[71] – both in academia and industry.

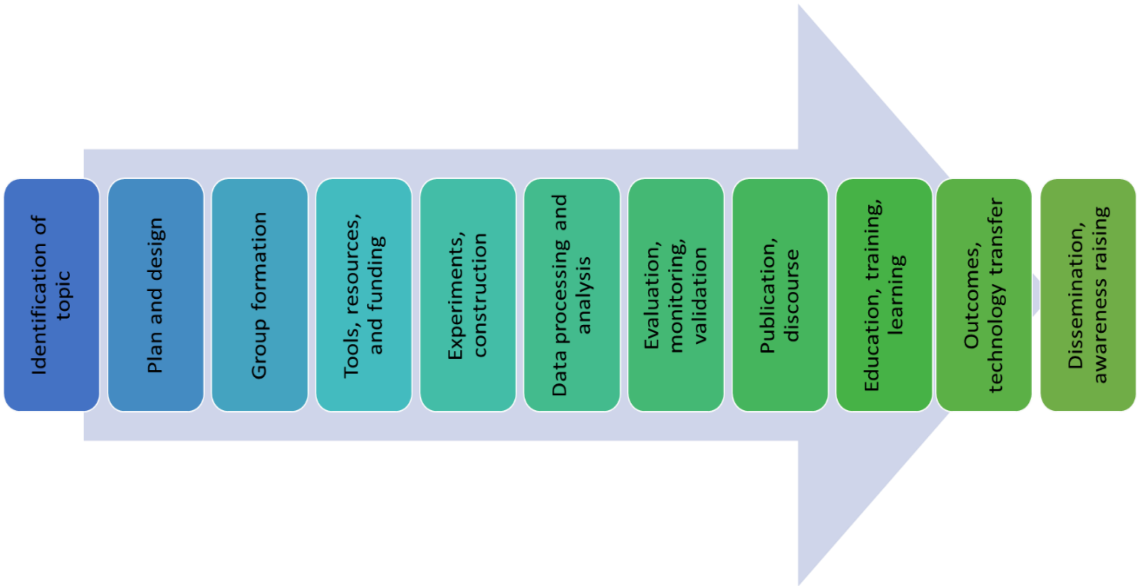


Figure 10: New and open scientific practices address every step in the scientific research process

The massive use of new ICT, the advent of massive data repositories, novel sensor systems, laboratory robots etc. have also transformed the daily routines of many researchers and scientists. Examples include:

- Disease management
- Social robotics
- MOOC
- Citizen empowerment
- Predictive systems based on computer modelling

[71] Science 2.0, science in transition 'Science 2.0'

– Connectivity for sustainable mobility
– Smart logistics
– Water resource management

ICT has impacted on all economic sectors and also changed most people's private lives. Whole economic sectors have undergone dramatic changes with the advent of ICT and some argue that ICT is changing the economy as a whole. It is therefore only natural that ICT is also impacting on science and research[72,73]. But it is far from clear today, which aspects of science, research and innovation will change in which direction. Potentially, ODS addresses all steps in the research process. The following table covers some specific cases of ICT usage in Digital Science.

Process step	ICT usage	Examples
Topic selection, funding	Open peer review, crowd funding of research, online problem data bases and open innovation systems	Generic crowdfunding platforms supporting science: Kickstarter[74], Indiegogo, Goteo, RocketHub. Specific crowdfunding platforms for science: - https://experiment.com - https://ilovescience.es (ES) - https://walacea.com (UK) - https://fundscience.org.au/ (AU)
Data collection	Big data, new sensor systems, automated data collection from 'internet-of-things', laboratory robots, interaction with citizens	Safecast is a global project to empower people with data, primarily by mapping radiation levels and building a sensor network, enabling people to contribute and freely use the data collected. http://blog.safecast.org/
Data analysis, generation of hypotheses and theories	Artificial Intelligence-based and statistical methods of knowledge discovery, data mining, interactive and visual data analysis; new computing infrastructure (including shared, distributed computing), interaction with citizens	Cognition systems - http://www.ibm.com/smarterplanet/us/en/ibmwatson/ is an artificially intelligent computer system capable of answering questions posed in natural language Wikisurveys - http://www.allourideas.org/ dedicated to creating new ways of collecting social data
Cooperation, discussion, evaluation and critical reflection	New electronic forms of discussions, collection of micro-knowledge, interaction with citizens and artists, new metrics, reputation and recommender systems	open-source electronics prototyping platform based on flexible, easy-to-use hardware and software. It's intended for artists, designers, hobbyists and anyone interested in creating interactive objects or environments - arduino.cc is an open-source hardware openly available, allowing the Arduino boards to be manufactured by anyone easily to use and re use. It is leading to a community of sharing based on

[72] EGI Open Science Commons
[73]https://www.kickstarter.com/discover/tags/science

[74]https://www.kickstarter.com/discover/tags/science

Process step	ICT usage	Examples
		openness beyond open source software and hardware.
Publication	Open access publication, open data (data re-use), open source software (software re-use), open methodology, collaborative writing, new media	- The Directory of Open Access Journals (DOAJ) is website that lists open access journals and is maintained by Infrastructure Services for Open Access. the database contains records for 10,000 journals[75] - http://arxiv.org is a repository of e-prints, of scientific papers with more than a million papers by the end of 2014. It is not peer reviewed but moderated. By 2014 the submission rate had grown to more than 8,000 per month - ResearchGate is a social networking site for scientists and researchers to share papers, ask and answer questions, and find collaborators with more than 7 million of users
Other aspects	Open methodology, open educational resources	Public Labs[76] is a community of practice using inexpensive DIY techniques to change how people see the world in environmental, social, and political terms.

Table 5: ICT use in Digital Science

[75] Adams, Caralee (5 March 2015). "Directory of Open Access Journals introduces new standards to help community address quality concerns".

[76] http://publiclab.org/

Example	Identify research topic	Plan and design	Formation of collaboration group, funding and supplies	Development of tools, pooling of resources	Experimentation observation and data acquisition	Data processing and analysis	Research results evaluation, monitoring, validation and sharing	Education, training and learning	Scientific outcomes and technology transfer, Innovation	Dissemination, awareness raising and outreach	
CERN				x		x			x	X	Data science, standards, e-infrastructures, open source, international collaboration, Large data, computational simulation, innovation
Fab labs	x	x	x	x				x	x	x	Digital prototyping, problem definition
Galayzoo						x		x		x	Crowdtasks. educational resources
Kickstarter			X							x	crwodfunding
Safecast			X	X	X	X	x	x	x	x	Crowdsensing, open access knowledge, open source, open methodologies, public engagement
Allourideas	X			X	X	X	X				Wiki surveys
Arduino	X	X	X	X	X	X	X	X	X	x	Open source, new media
ArXiv	x						x	x	x	x	OA, Open peer review
Public Labs	x	x	x	x	x	x	x	x		x	Open methodologies, crowdsensing,
Research Gate	X						x	x	x	x	Social network, altmetrics,
PDB	X										Online database

Table 6: ODS examples classified per step in the scientific process

4.3 ICT characteristics underlie open (digital) science features

Important law-like characteristics of ICT that also impact on Open Digital Science are: (i) the characteristics of group-forming networks; (ii) the zero marginal costs effect of software and similar information, and (iii) the power of formal modelling and simulation. There is reason to believe that these laws are also at the heart of ODS.

i) An intrinsic feature of many information and communication networks is the fact that the network value is linked to the number of connections, e.g. in telephone or FAX networks. For n users in a network, the network value is proportional to its $n^2 - n$ connections. This is known as Metcalfe's law [Shapiro & Varian 99]. It was pointed out [Reed 99] that there is exponential growth of value with the number of users in networks that facilitate the formation of groups, e.g. LinkedIn, Facebook and many others. Here, the connections are not only created between nodes or users of the network, but also between groups and therefore network value grows proportionally to 2^n.

It is important to understand that the dynamics of such networks is very different from purely non-group forming networks. Already small numbers of users (or other kinds of 'nodes') give rise to a large number of potential interaction between these users and between groups of users. Users of networks such as LinkedIn are experiencing this in interaction with not just single other users, but with different groups of users sharing an interest in a subject for which this group was created. These electronic group networks thus facilitate the exchange between arbitrary subsets of users that have expressed interest in the same subject. In going beyond this, it is even possible that the same set of users create different groups for different topics.

ii) Economists have pointed out that knowledge, like information, is characterized by zero marginal costs. However, in practice access to knowledge without the internet and without tools for efficiently searching large databases was far from being free. Costs for finding and accessing knowledge stored in books were high due to the time necessary. The advent of free and open access publications has changed this. While there is still a lot of argument and development regarding the precise conditions under which scientific results should be published, a plethora of knowledge is now in the open domain and it is free for everybody to use.

Open access alone did not change the game towards Science 2.0. In practice, most scientific libraries were already free decades ago and full of interesting results. Rather it is today's availability of free high-quality software for large-scale interest-driven and interactive search that supports both professionals and citizens in gaining access to scientific and research knowledge. In addition, more than just papers are provided and the scientific and research results include software, demonstrations, teaching material, construction plans, diagrams, video tutorials etc.

iii) Science – and in particular Science 2.0 – is a massive user of computer simulations. This is based on the use of formal models of the world developed in science (but also in the humanities, in engineering, and even in the formal sciences). Computers on the other hand provide general-purpose (universal) simulation engines capable of not only calculating numbers, but also manipulating non-numeric formal systems. This lends them perfectly for simulation of nearly all kinds of systems. (An example of the limits of computer simulation is discussed in [Feynman 82].)

The widespread use of simulation in Science 2.0 is not just due to recent increases in computing powers, support software and networked computing resources. Visualizations of data play an important role as well. Visual representations have become interactive for human users and relatively easy to produce with standardized software. This visualization is not just a nice side-effect of computing power, but may touch the essence of *'understanding'*, cf. [Breithaupt 06].

5 ODS in Europe and worldwide

One objective of the study is to improve our understanding of the current state-of-play of ODS in Europe and elsewhere. To this end, the study will also suggest metrics to measure ODS progress and components of an ODS observatory. First steps towards such an observatory have already been made by creating an internet platform that brings together a broad range of information on ODS and related topics. This web site, created by the study team and available online at http://opwendigitalscience.eu also serves to publish information about the study including interim results. This facilitates interaction with the scientific community in general and the community sharing an interest in ODS in particular.

5.1 ODS observatory

This sections describes the Open Digital Science web page available at http://www.opendigitalscience.eu. The main objective of this page is to maintain a live repository of ongoing projects and allow third party projects to add themselves. We're also analysing some Open-Digital-Science related twitter hashtags and we have created a collaborative map on Open Digital Science.

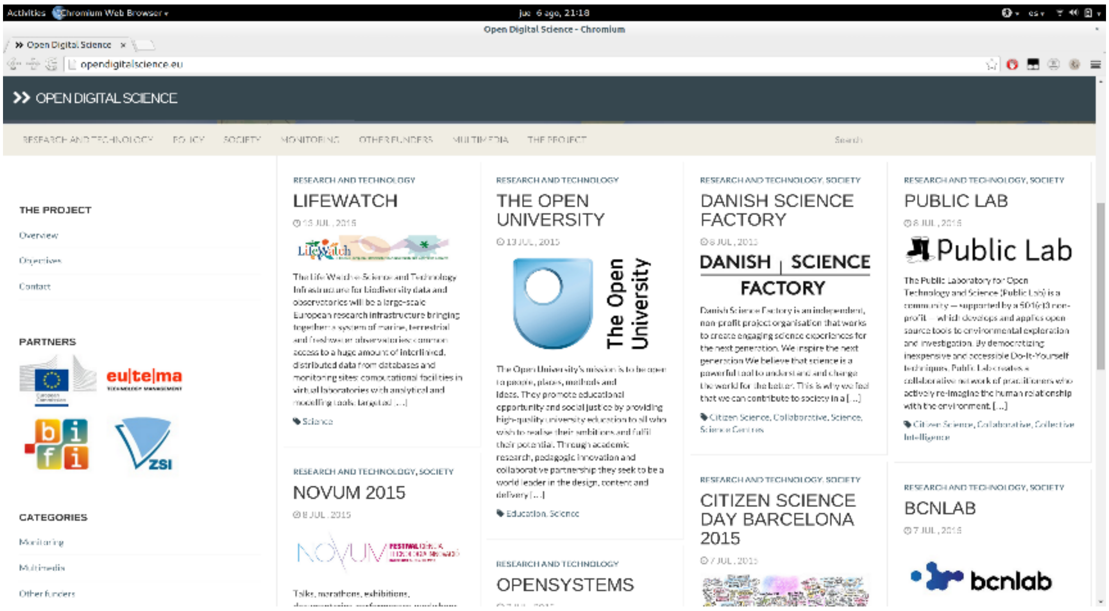


Figure 11: Open Digital Science web site main page

Most of the content described in this report (except analytics) is also accessible through the main web page. It was created using Wordpress [77], an open source and widely used content management system. It was installed and maintained in a virtual server provided by the University of Zaragoza, created using the OpenVZ[78] container-based virtualization. A responsive template provided by Marchetti Desing[79] was bought and modified according to our necessities. Some Wordpress plugings were added in order to provide some functions as automatic google-facebook-twitter register, calendar, StackOverflow like forum, etc.

[77] http://www.wordpress.org
[78] https://openvz.org
[79] http://www.marchettidesign.net/

Wordpress allows to classify content according to two orthogonal parameters, categories and tags. New content can be parametrized matching not only the main categories: Research and technology, Policy, Society, Monitoring, Other Funders, Multimedia, The Project but also to the needed tags: Science, Education, Collective Intelligence, etc.

A large amount of projects, best practices, etc. were added and categorized into our Wordpress CMS, but this is yet an ongoing task as we're adding new pages to our repository as they appears -or we found them.

As mentioned previously, some other tools such a calendar, a forum etc. has been added to the web page. We want to provide in this report a brief mention of the two most important tools, the Collaborative Map and the Twitter Observatory. Both can be found under the Monitoring category.

Collaborative map

We created a collaborative map on Open Digital Science using Ethermap[80], a real time collaborative and version controlled map editor. It was programmed and published in GitHub by Dennis Wilhelm. We forked this project and modified it according to our necessities. This is a real-time collaborative and version controlled map editor. Therefore, members of the scientific community can modify this map, add projects, institutions etc. in real time. It also allows you to see what other users are doing. We are storing versions of the database, preventing from damages by malicious users.

Figure 12: Collaborative map on Open Digital Science (http://opendigitalscience.eu:3000/map/ods)

Twitter Observatory

This Observatory is provided by the Kampal[81] spinoff. This tool can store and analyze in real time a preselected group of worldwide twitter hashtags. One can find an evolution of the selected

[80]https://github.com/dwilhelm89/Ethermap
[81]http://kampal.com/

topics, a general view, heat map, statistics, polarity and communities. We found this as a great tool in order to check out what is happening in twitter regarding Open Digital Science and other related topics. You can find this tool via the main web page inside the *Monitoring* category.

Figure 13: Real-time monitoring ODS hashtags, heap map (http://social.kampal.com/visualization/ods)

In this example, 14 initial hashtags were analysed: bigdata, opensource, crowdsourcing, openaccess, ods, citizenscience, openhardware, science20, opengovernment, escience, digitalscience, ictarts, gssopendata, openknowledge. Future updated versions of this statistical data will be refined in order to be used to compute indicators. Templates like those available on http://ec.europa.eu/digital-agenda/en/download-data and the DataCube [82] vocabulary will be used.

Analytics

We've been using the Google Analytics tool to collect information about the visitors to the web page. Since the beginning of the analytics -24th March 2015 – up to this moment – 6th August 2015 - 2390 users have visited the web page. The most visited section is *Research and Technology.* 60% of the visitors were between 25 and 35 years old and 36% of the people were women. Some other interesting statistics are provided in the following pictures.

[82]http://www.w3.org/TR/vocab-data-cube/

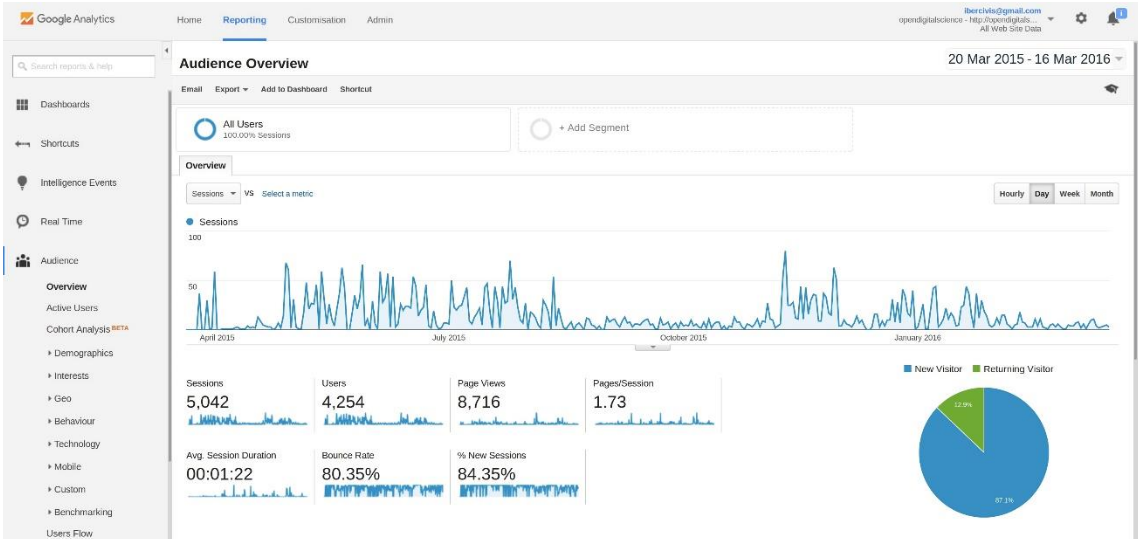


Figure 14: Google Analytics, audience overview

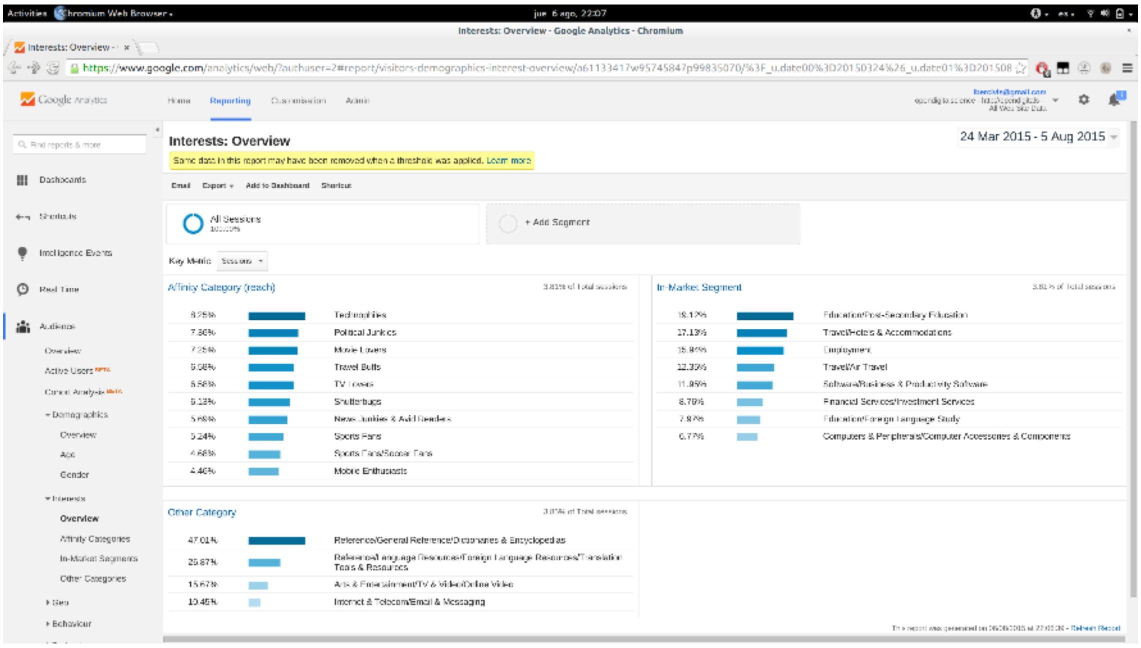


Figure 15: Google Analytics, interests

5.2 Selected cases

The following table presents some of the best cases collected on the OpenDgitialScience.eu website over the course of the study:

NAME	URL	Brief Description	Category	Subcategory	ODS practice	Why it is good case
OPEN DATA HANDBOOK	http://opendatahandbook.org	This handbook discusses the legal, social and technical aspects of open data. It can be used by anyone but is especially designed for those seeking to open up data. It discusses the why, what and how of open data – why to go open, what open is, and the how to 'open' data.	Resource	Guidelines	Open Data	It provides a clear vision of the concepts with concrete examples and tools that can be applied when working with open data.
OPENCORPORATES	https://opencorporates.com	The largest open database of companies in the world.	Resource	Repository	Open Data	It brings openness and big data potential to the private business environment with millions of entries.
LEONARDO	http://www.leonardo.info	The international society for the arts, sciences and technology	Network	Publication	Arts and science	Leonardo brings together different disciplines for cross-fertilization.
FWF TOP CITIZEN SCIENCE	https://www.fwf.ac.at/en/research-funding/fwf-programmes/top-citizen-science-funding-initiative/	Call for the "Top Citizen Science" (TCS) funding initiative, which has a total endowment of € 500,000 (€ 250.000 FWF and € 250.000 OeAD). Under this call, funding is made available for the expansion of "citizen science" components.	Funder	Program	Public engagement	It is one of the earliest and biggest contest specific for citizen science projects in the world. It promotes existing projects to adopt public engagement methodologies.
OPEN SCIENCE PRIZE	https://www.opensciencepri ze.org	The Prize provides funding to encourage and support the prototyping and development of services, tools or platforms that enable open content – including publications, datasets, codes and other research outputs – to be discovered, accessed and re-used in ways that will advance discovery and spark innovation. It also aims to forge new international collaborations that bring together open science innovators to develop services and tools of benefit to the global research community.	Funder	Program	Open Innovation	It mobilizes a global community, linking to a great number of resources available https://docs.google.com/spreadsheets/d/14_oCSKXcDWlcc4UBjLEh2KFidYN7ZBfCMNbz2iXu-2M/edit#gid=1595396026
UTTLEWORTH FOUNDATION	https://www.shuttleworthfoundation.org	The Shuttleworth Foundation is a small social investor that provides funding to dynamic leaders who are at the forefront of social change. We look for social innovators who are helping to change the world for the better and could benefit from a social investment model with a difference. We identify amazing people, give them a fellowship grant, and multiply the money they put into their own projects by a factor of ten or more.	Funder	Program	Open Knowledge	This program is funding innovative ideas with great potential for change. Applying is easy with the aim of brining ideas-makers rather than proposals-makers.
OPENDOAR	http://www.opendoar.org	OpenDOAR is an authoritative directory of academic open access repositories.	Resource	Repository	Open Access	Each OpenDOAR repository has been visited by project staff to check the information that is recorded here. This in-depth approach does not rely on automated analysis and gives

NAME	URL	Brief Description	Category	Subcategory	ODS practice	Why it is good case
						a quality-controlled list of repositories.
FUTUREEVERYTHING	http://futureeverything.org/	FutureEverything is an award-winning innovation lab for digital culture and annual festival, established in Manchester in 1995.	Laboratory	Festival	Arts and science	For more than 20 years FE has been exploring the meeting point of technology, society and culture. It is expanding to to other cities through a community network and regular events. It makes connections between thinkers, developers, coders, artists, designers, urbanists and policy makers – inspiring them to experiment and to collaborate in new ways.
ARS ELECTRONICA	http://www.aec.at	Since 1979, Ars Electronica has sought out interlinkages and congruities, causes and effects. The ideas circulating here are innovative, radical, eccentric in the best sense of that term. They influence our everyday life—our lifestyle, our way of life, every single day.	Laboratory	Festival	Arts and science	The Festival is a reference for networking and inspiration opportunities. The Prix as competition is pushing forward excellence and inspiration. The Center and the Futurelab as in-house education and R&D facilities. The Solutions is a company itself. Ars Electronica's It is a unique environment, and it is continuously reinventing itself.
FEDERAL CROWDSOURCING AND CITIZEN SCIENCE TOOLKIT	https://crowdsourcing-toolkit.sites.usa.gov		Resource	Repository	Public engagement	It recognizes the potential of citizen science for science and government and federal scale. It provides five basic process steps for planning, designing and carrying out a crowdsourcing or citizen science project. At each step, you'll find a list of tips you can use to keep your project on track. In addition to the tips, you'll find case studies with success stories and some of the challenges that project developers faced. The case studies can serve as models, inspiring you to plan your own project.
SOURCEFORGE	https://sourceforge.net	SourceForge is an Open Source community resource dedicated to helping open source projects be as successful as possible. It thrives on community collaboration to help creating a premiere resource for open source software development and distribution.	Resource	Repository	Technology	With the tools they provide, developers on SourceForge create powerful software in over 430,000 projects; they host over 3.7 million registered users. IT professionals come to SourceForge to develop, download, review, and publish

NAME	URL	Brief Description	Category	Subcategory	ODS practice	Why it is good case
						open source software. SourceForge is the largest, most trusted destination for Open Source Software discovery and development on the web.
POLYMATH PROJECT	http://polymathprojects.org	The Polymath Project is a collaboration among mathematicians to solve important and difficult mathematical problems by coordinating many mathematicians to communicate with each other on finding the best route to the solution. The project began in January 2009 on Tim Gowers' blog when he posted a problem and asked his readers to post partial ideas and partial progress toward a solution.	Network	Challenges	Mathematics	With ten challenges solved, this experiment resulted in a new answer to a difficult problem: online collaboration to solve complex math problem.
DANISH SCIENCE FACTORY	http://danishsciencefactory.dk/	Danish Science Factory is an independent, non-profit project organisation that works to create engaging science experiences for the next generation. It is funded by a mix of public and private money, getting support from most of Denmark's biggest foundations, organisations and companies.	Laboratory	Festival	Education	DSF collaborates extensively with all of the biggest science centres and relevant organisations in the informal education sector as well as a number of formal educational institutions, schools and universities.

Table 7: Best cases collected on the OpenDgitialScience.eu website over the course of the study

5.3 Trend analysis

Trends in computing

A recent study on Next Generation Computing [Prem 14] expects the following continuing trends in computing:

- *More with less:* Despite of many challenging limits of physics, computers and computer networks will continue to grow in computing power and speed, while shrinking in size. In many applications, however, it is likely that we will see many more resource-efficient computing devices, i.e. regarding energy or other scarce resources.
- *Software-driven world:* Hardware is becoming a commodity. The real power and intelligence of systems is going to be in software.
- *Cloud and hybrid:* The massive trend towards storing data in a cloud of internet-based servers is likely to continue. Even with technologies for data storage improving, there are clear benefits of relying on cloud-based services not just for storing data, but also for processing it.
- *Mobile computing, internet-of-things:* The massive trend towards mobile devices will continue. More and more artefacts – machines etc. – will become connected to the internet. New technologies will further reduce costs of connecting objects to the internet. This will facilitate the connection of even short-lived devices or perishable goods to the internet.
- *Open and do-it-yourself approaches:* In particular in the business world, the trend towards open systems is likely to persist. At the same time, it will become easier for both professional and private users to create applications from modular components at relatively low costs or to even buy services just for a single usage.
- *Converged and integrated systems:* The trend towards more and ever more complex systems controlled by electronics and software intelligence will persist.

All of these trends are likely to have an impact on open science. A trend towards an even broader roll-out of – in principle – connectable and new devices directly impacts on the options for researchers and scientists to create new experiments. The trend towards distributed computing and cloud services is also likely to support scientists in managing data, processing, open and interactive processes and speed in the creation, management, and evaluation of scientific research projects. New tools will facilitate the virtualisation of processing over large, distributed processing networks that are dynamically created and perhaps continually adapted due to changing requirements. The tendency to store data where it is produced or to rely on the clouds will persist. However, we are also seeing limits of data storage capabilities and data is already becoming processed where it is produced (e.g. satellite data).

Perhaps more interesting, it is likely that large software and cloud service providers will also participate in the small- and large-scale analysis of data. Their tools are likely to become very powerful, yet comparatively easy to use. Already today, players like Google provide solutions for the statistical analysis of data, for visualization and search in large data bases. These applications do not always fulfil all requirements for critical researchers, e.g. neither are the working principles fully clear nor are the data bases easily available. However, the tools are freely available, powerful, reliable, and they are often equipped with visually powerful graphics. This makes them difficult to resist, even for the critical scientist.

Societal trends

ICT trends have developed in parallel to societal changes and partially have also caused or at least supported these changes. To name just a few, more people than ever carry mobile phones, more people than ever are connected to the internet; the internet has become a communication network for the masses difficult to control and powerful in political processes; people find like-minded people using a range of social networks; and finally, an increasing number of robots supports not only industry, but also private households.

Analysts of societal trends in the developed world argue that the emergence of the 'digital native' is an important game changer in many economic processes and of society [Palfrey & Gasser 08]. It is the digital native that drives innovation in the mobile and software industry. The digital natives demand other types of services even from the most non-IT companies.

Digital natives have grown up to become scientists. It is not unreasonable to assume that they have no less expectations from the systems they use at work compared to those used for private purposes. These expectations may concern mobility, user experience and reliability, but also provision at no cost and high-speed interaction.

Megatrends

Science as a human activity is also embedded in global economic and societal trends: *Urbanisation, mobility, new energy systems, and the new normal economy* are just the most often cited ones of these trends. Although these trends are not ICT trends, the pervasiveness of ICT ensures that many of these either drive or are driven by developments in ICT. As an example, the increase in renewable energies and the shift in citizens from consumers to producers of energy is dramatically changing energy networks. [Rifkin 08] calls this development the 'third industrial revolution' based on an 'internet of energy'. He also points out the massive societal trend toward sharing rather than owning [Rifkin 14]. Again, it is important to understand that both these trends are ICT-mediated. Modern bi-directional energy networks require the kind of control and communication structures provided by modern ICT. And the sharing of products is often facilitated by chip-based and mobile systems, new tools and technologies for micro-payments, tracking, monitoring, remote maintenance etc.

The speed of innovation is likely increase in the now 'new normal economy'. As times are uncertain, investors are likely to put pressure on rates of return and thus on developers, researchers, and scientists to develop technologies required for new products and services. While the international centres of science and research are still heavily clustered in only a few countries, it is likely that new emerging economies will increasingly contribute results and researchers. These developments are likely to further drive internationalization of science and research, managed with new research project and team management tools and services supporting collaborative writing and publishing.

6 Open science scenarios: impact and uptake

To gain a better understanding of open science and the digital tools to realize it we described a set of open science scenarios. These scenarios should not be read as predictions of the future. They are views of what the future may bring and what might be necessary to make them happen – in particular with an emphasis on desired characteristics of such a future.

Scenario-based exploratory techniques have proven useful for strategic forecasting of technical trends [Geschka & Hanenwald 13, Mieke 07, Lizaso et al. 04] and more recently for research and technology policy [Prem et al. 14, Prem 14]. The different aspects[83] of open science are difficult to embrace in a single definition or description. Future scenarios are an intermediary conceptual step and help to communicate the ideas, discuss trends and their implications, understand issues, and collect policy needs.

The scenarios were developed in the form of plausible narratives. They assisted in communicating a clear vision, explore future trends, discuss options for measuring open science uptake and impact and generally stimulate the discussion with a broad community. They were developed in close collaboration with experts, input from the Advisory Board, and presented and discussed at two scientific conferences and online.

The six scenarios address different aspects of open science:

- ScienceFlex is about Citizen Science and its support with new online tools
- InnoSpeed discusses how open science can drive innovation, in particular for SMEs
- BlurredBounds describes a not-too distant future where the boundaries between academia and industry become nearly invisible as organisations dynamically create projects and recruit staff for virtual global project teams
- Digital Studies elaborates on how digital and open science changes the educational work of universities
- Policy Dialogue offers insights into the past choices of future policy makers
- Ancient Now describes discusses open science technology challenges from the perspective of the future

The scenarios were drafted to address a broad range of different stake holders from citizens to SMEs, universities, students, policy makers, and ICT infrastructure providers. The scenarios start with a short summarizing story line overview. Each scenario then describes the involved technologies, anticipated or current technological challenges. A few lines were added to provide the background and context and a short story then describes the concrete scenario. They are not complete pictures of the future and only highlight certain elements to stimulate the discussion.

[83] For example, citizen science, open publishing, open data, big data, funding ODS, innovation technologies, etc.

6.1 ScienceFlex

Story line

Julia and Claudia, veterinary researchers of a public-private research organisation, travel to meet with Citizen Scientists with whom they have been collaborating over a virtual platform. This will be their first real meeting and they are very excited

Technology/challenges

Research enrichment using new media and new actors. New experimental data sources (human as sensors, intra-body measurement). Crowdsourcing fine-grained level of monitoring. Validation through scientists.

Societal challenges such as climate, environment, energy or transport cannot be achieved without innovation. Scientific agenda co-creation. Decrease costs of research using resources from volunteers.

Online-platform for interactive citizen science – form funding to data collection to result dissemination. Learning is not limited to schools. Researching is not limited to laboratories.

Scientific culture. Making scientific knowledge more accessible. Democratization and appropriation of procedures and results, informed consent. Self-learning.

Physical spaces as mediators.

Active role due to increased proficiency. Citizens require right skills to play an active role. Basic research marginalized, more emphasis on sexy and easy to understand topics, in particular health.

- Simple tools
- Access to data
- Avoid misuse: trusted voting, reputation management
- Avoid anti-science?

Citizens can 3-d print at home a measurement device to perform experiments worldwide.

Background/macro perspective

How can Citizen Science be best combined with online support tools. How is it possible to attract and access large groups of citizens – for whatever purpose? How can both scientists and researchers on the one hand and citizens on the other be best supported in an interesting, non-trivial, dynamic fashion?

Story

When Claudia arrived to the Madeira airport she updated the ImHereApp which she uses every day to track her movement while monitoring how much energy and calories are consumed globally. This app also helps her to schedule best routes using collective voting systems in real time. Claudia is a veterinarian post-doc who is travelling with Julia, her principal researcher. They work for a public-private research institution with the focus on new pharmaceutical compounds. They are travelling for a participatory event promoted within ScienceFlex, a virtual community of citizen science. Julia is going to present her experience using the augmented reality SharedLAb tool for virtual meetings and remote synchronized activities. Claudia, who is even more excited, is more interested in the excursions, workshops and lectures, even cultural and social activities.

Their first activity is an excursion to the Wild Island where they visit a portion of the island with thousands of sensors and electronic devices installed in the wild. They join Francisco, the researcher in charge of this installation. This installation is quite popular: last month they reached 1.000.000 volunteers who analysed one of their pictures taken in the shore. “Last week, we reached our volunteer number 1.5 Million - a woman from Lithuania- who helped us to identify one mini orchid, can you believe it?”

Next day, they have a pop-up event about environmental diversity organized by CoWildMad, the regional institution devoted to promote citizen engagement in science under its open programme for public health monitoring. The event takes place in a socio-cultural centre in one little town in the northern side of the island, far from tourism. It is co-organized by several local communities (farmers, retired people, local crafters and artists…). Only few of them are used to work with low-cost-but-accurate sensors, but they know very well their surrounding environment. One of the most active moderators is Matthew, a passionate who is traveling all around the world supporting this kind of events and who came to Madeira supported by the Citizen Science Relationship Office from Lisbon.

The list of topics to address was open one month ago using an open source wiki-survey so that everybody could share their problems and ideas to develop reaching more than 50 concerns, some of them were subject of intense debate. From this list of topics, Claudia was especially attracted about the necessity to monitor one species of insect which during last year was creating some problems for the irrigation system, as this insect collapsed the stagnant water in the backyard of several neighbourhoods.

Scientifically, this challenge is interesting for her so she joins a table with ten local actors and other seven researchers to work around that little insect. Some people came very well prepared with several tools already 3D printed and assembled at home. These instruments will count the number of insects using high sensitivity microphones connected to the cloud to discriminate useful data from the noise of the field recordings. They also build up a mini pool in the centre, fully equipped with sensors, to monitor micro-scale behaviour. This mini pool is used also to test first compounds to neutralize the offspring. In order to monitor the medium scale behaviour, they also set up a bunch kites, balloons and drones used to do fine grain mapping. They used the OpenData portal from the government to surround the local irrigation and piping system. Most of the code deployed is a fork of the existing code available in Code4all, one famous online repository devoted to citizen science. For the data simulations, they used Nature Mobility Models available in the Complexity@All portal. In this case they run the simulations using volunteer computing platforms at home.

Once the event is finished, the sensors are deployed and reporting data in real time. Data is useful for Claudia since they are also validated by the QualityCheck tool. Claudia applies her artificial intelligence tool to discriminate those data tagged with her methodology of interest. Surprisingly, one creative team from Brazil did a funny visualization of the data, creating a kind of cartoon in the local language. Flying back to home, they are watching these videos on the plane. Claudia says to Julia, “I am going to ask my ScienceFlex colleagues to help me right now to translate this tool so that my children also enjoy this at home with their friends”.

6.2 InnoSpeed

Story line

ODS generates new tools and new opportunities for SMEs to access to global knowledge. Tools include data repositories, new software, tools for evaluation and infrastructure. Future ODS tools will be highly standardized with interoperable interfaces. SMEs will be able to tap into the global talent pool for short-term tasks, projects, or long-term engagement.

Technology/challenges

- Built-in multilingualism
- Semantic interpretation of scientific work using professional (rather than academic) vocabulary
- Interoperability of data bases, software, tools.

Background/macro perspective

In this scenario, the transfer from knowledge as a club good to a public good becomes possible by added context, training resources, consulting, access to students and software tools. This goes beyond simple access to only data (published in a paper), but also includes access to negative results/counter examples, preliminary data, software tools, a video interviews and background information. From the organisational side this includes subcontracting of data interpretation and analysis.

Story

Alex Tanto is the CEO of a small but highly specialised carpentry firm. All his life, Alex has been working with wood. As a small boy, he never imagined to be working in this field; but when his father died, he took over the troubled firm that produced furniture for local companies, restaurants, and families. The company now thrives: Alex soon realised that he needed to specialise if his company was to compete with low-cost overseas furniture importers. Today, his team provides flexible transport packaging solutions for the industry. His main customers include automotive, logistics, and even aeronautics companies.

Alex never attended university but, with the help of data brokers, managed to successfully tap into world class knowledge to make his packaging the best of its kind. He worked with mathematicians to optimise the wood cutting by reducing material while retaining strength. He recently completed a joint project with a team of Chinese researchers and an Australian SME whose goals was to develop a new wood preservation agent. In a decade-long study, the Chinese researchers had analysed over 5000 different timber preservatives and stored the recorded data in a database where it sat mostly unused for nearly 15 years. One day, the Australian SME, a specialist in developing online ship models, was tasked with modelling wood properties. The Australians found the timber preservation data as well as the responsible researchers through a multilingual search engine. They could quickly convince the latter to cooperate on developing new models that they then brought to the market.

Independently, Alex saw a need for new wood preservatives. His first activity was to consult an online search engine for innovations. The search engine, a spin off from an EU project, proposed the Australian company as a specialist in the area. With the help of a data interpreter, Alex was quickly brought up to speed concerning the potential of the extensive data for his work. Within three weeks, Alex was able to find two talented students that would work with the Australians on further exploiting the available data by systematically analysing over 1500 new combinations of wood preservatives for his packaging solutions.

The resulting model is now being implemented as a web service that accesses the Chinese data base, uses the procedures and material property simulations from Australia, and combines them with the designs of Alex' company.

The result was so successful that Alex shared his experience through a video blog, provided the data in one of the many established repositories, and even co-authored – in an old-fashioned way – a paper that was published in a recognised professional open access journal.

6.3 BlurredBounds

Story line

New types of organizations (NGOs, consultants, science entrepreneurs) dynamically create RTDI projects sourcing from global resources and competencies. They combine data and tools with accessing to specialized infrastructure. ODS facilitates this access and empowers individuals and specialized (dynamic) teams to engage in new RTD projects. The role of setting up projects and project management will resemble that of a film producer or entrepreneur today who combines required experts, talents and physical assets on a project-by-project basis. Motivation includes opportunities for market innovation, but also research and scientific challenges and opportunities for social innovation.

Technology/challenges

- Agile shared facilities
- Trusted recommender / broker systems (to find people, state-of-the-art approaches, …)
- Collaborative tools supporting the process, easy interaction with peers, specialists
- Stable and proof rating systems
- Scalability of computational power, data size, algorithms, computing paradigms
- Flexibility / stability for individuals (financial: payment of dues, insurance, agile lending…)
- Telepresence / robots
- Transferring old structures and entities to the modern demands of leading-edge-research
- Growth of freelance R&D and citizen science vs old hierarchical models (Universities, RTO structures) – funding schemes etc.

Background/macro perspective

Researchers no longer need to decide to pursue either an industry or an academic career. ODS facilitates non-linear and flexible career pathways. New researcher evaluation systems go beyond measuring simply the number of citations of a journal article but take into account new forms of publications, altmetrics, and quality factors. Policy makers and agencies support these kinds of projects with more agile funding mechanisms and with adjusted time horizons. Flexibility and scalability like the number of project partners, the different stakeholder groups that are involved, the required computation time, and the data size are major qualities of future research projects.

Story

2025-08-23 - "Late… late… always late". Peter M. is waiting for Linda S. to show up in their web-based collaboration network. <ding> – Linda is online. She is only 3 minutes late but Peter is a bit nervous because of the planned kick-off of their new research ensemble which they worked on over the past 3 weeks. Linda and Peter are the leaders of an open research match making company. They managed to find investors for a promising research concept on “enhanced 3D printing methods for printing metal structures” – they filtered the results of an open academia-industry innovation contest and formulated a project that is interesting for their company. 80 % of project budget comes from a private public partnership, the other 20 % have to come from private investors who Linda and Peter believe to have found in the U.S. and in France.

In the global fight for talent, Linda and Peter succeeded with their business model: to identify the best constellation of research actors (researchers across all scientific areas, SMEs, data centres, technology parks, etc.) for demand-driven innovation projects. At the same time they have high expertise on combining evaluation and impact data on researchers, which helps them to filter individuals with excellent performance records in specific stages of the scientific process.

Having established the core team to meet the demands of the challenging new project, the next step was to secure the required infrastructure. Since extensive simulation phases are planned and peak times are not known precisely, the choice was to rent elastic cloud computing power for the project duration. This is one factor to counter the unpredictability of their endeavour: to solve the envisaged metal structure challenges. Another factor is that they joined forces with a virtual specialist group of material experts who located in Asia and the U.S. Consequently, they gained access to an additional team of five required experts within 3 days of notice.

They decided to rent a remote semi-robotic lab for the second half of their project. After the conclusion of most simulations, this will be the time for experiments. Since those will produce a lot of heat, cooling is essential. Thus they decided to rent a shared facility in Norway. In contrast to their first desired location in Texas, this means that they have to upgrade two machines and have to deliver one additional heating oven to that location. But their savings in energy regarding cooling are outperforming the location in Texas. The facility in Norway will provide experienced staff to maintain the machines and the processes. Project members who are spread all over the world will be able to remotely control the majority of the processes 24/7 as the lab is highly automated. Two mid-sized data centres on different physical locations are also rented and already configured to act as backup-destinations for the simulation and project outcome and will mirror the restricted collaboration platform dedicated to the 3D printing research project.

After 30 minutes their project core team – the most valuable part of their ensemble - appeared online and they went through the most important parts to start the project. After their successful kick-off Linda and Peter changed to their virtual company chat-room each toasting to the success of the project.

Linda: “Do you recall the time when researchers had to decide between a university and industrial career and they typically were so stuck in their separate world? Today we have so many additional opportunities. Don’t you think?”

6.4 Digital Studies

Story line

Universities will remain, however they will split between few highly prestigious international branch campuses, big established universities and specialized niche suppliers. Some middle-tier higher education institutions and inefficient universities go bankrupt. Digital technologies deeply transform the way education is delivered and accessed. Universities are co-creators of dynamic digital learning ecosystems for lifelong learning on a global scale. Massive Open Online Courses (MOOCs) and hybrid courses are well established and describe a new a customer service/marketing orientation toward students. Half of students of MOOCs come from developing countries. New partnerships between universities, students, governments, industry, alternative education suppliers, NGOs, R&D organizations and the community shape the new organisation of administration and services. There is a cultural change towards the recognition of alternative educational pathways. The evaluation of universities relies on a new system of outcome-focused metrics which include indicators on average debt loan per student, job placement rates, alumni satisfaction, etc.

Technology/challenges

- Next step in Massive Open Online Courses (MOOC) will include not just courses, but also increased amount of data and tools (software, simulations, …)
- Virtual lab access
- For easy integration, new technologies will be needed that seamlessly integrate with each other and that are easy to use for the non-IT researcher
- There will still remain room for localized teaching (e.g. language, course requirements, regional aspects etc.), but teachers will have to opportunistically exploit, access and include globally available material and tools
- Competition may rise between universities to provide globally used tools. In SSH emphasis will still remain on access to publications, experts, and data; while in engineering and science it will also include tools, data, software
- ODS will make it easier to access and to scale global expertise. Opportunities for industry include access to specialized staff and tools to work on data and getting solutions. Challenge: access to confidential information.

Background/macro perspective

Universities usually put strong emphasis on both research and teaching. Therefore, as research moves into the realm of open digital science, changes can be expected in the area of teaching and student work. We may see less knowledge / training monopoly of universities and more cooperation. There will be an increase also in international course offerings and students' participation in international projects.

Story

After getting pregnant Maria realizes she wants to give another impulse to her live by becoming big data engineer. Maria has worked several years as a librarian and she is decided to get closer to technologies as a way to get closer to her future children. In the last few years she has realized that there is a huge gap between the dolls and games she experienced in her childhood and what kids are using these days, including all this amazing gadgetry connected to educational and community-based resources. But she is not that good with her hands, rather she loves to get deep into books cross-linking stories and scenarios. When she is pregnant with her first baby she takes some time to browse digital tools and online mappings of career pathways to explore various educational options that could lead her to her new career.

She finds the best match which allows her to balance university and child care and is conform to her financial possibilities:

- a top rated MOOC in computing which has been established in cooperation between two recognized European Universities
- an Institute of Technology
- a major telecoms cooperation
- and the Institute for Communication and Ethics

Maria performs very well in the online course which is also recognized by her project peers and mentors who follow her progress throughout the whole course via an online student performance monitoring tool. Despite of doing the activities pretty late in the night, she finds the workflow pretty comfortable.

Her mentor recommends her for a job and Maria gets the opportunity to work remotely some hours a week as an intern in a data centre for accounting and performance analysis. It was a bit tricky to set up the credential account manager in her laptop but once resolved the biometrics issue, remote operations worked very well and virtual presence allows her to interact pretty often with her colleagues.

When her child enters kindergarten Maria decides to attend university for two years in order to fully embrace the humanist side of technologies. Universities have started to offer more networking opportunities, which makes it attractive for Maria to be physically at site and build social capital. Fortunately, digital courses have become more intertwined with existing curriculum and acceptance of MOOC credits is easy. And Maria can still choose to follow some video lectures online at home and then spend class time for the discussion of problems and practice with peers.

Maria can advance at her own pace and balance studies, child care and her need to earn a living. At the same time universities have become more flexible and family friendly and she can enjoy certain child care support services. As money is getting shorter, Maria starts a part-time job in her last year while finishing her studies online. She finds interesting additional specialization courses in edX which allow her to sharpen her career profile. Maria finds a good job shortly before finishing her studies due to her diverse portfolio of credentials including online certificates, work-relevant project enrolment and lifelong learning experience.

6.5 Policy dialogue

Story line

Two policy makers meet in 2025 at the World Summit on Science and Research Policy. They look back to their policy decisions in 2015-16.

Technology/challenges

Policy makers have an opportunity now to steer ODS in terms of incentives, regulations, infrastructure etc. The range of topics is very broad in principle. Several topics such as open access have in the meantime been addressed at national level – often even differently within a single member states. Some countries have also started with support initiatives for ODS, e.g. services or infrastructure. A major challenge concern the level of initiative (global versus local); another challenge for policy makers is how to best align with industry on the one hand and with academia on the other.

Background/macro perspective

The main issues of ODS policy have been privacy issues in the context of data collection, exchange, aggregation, mining etc.; intellectual rights; open access; research infrastructure and infrastructure accessibility rules. An important area of influence concerns rules of funding agencies, but also alternative ways of measuring scientific quality (altmetrics).

Story

ODS 2025 – World Summit on Science and Research Policy 2025 in the Open Science Tower build in 2017 as a symbol for our joint efforts to implement our new open science and research policies

– Hardly anyone calls it Open Science any more –says the Indian senior official in his early fifties.

– True –ponders the EU policy-maker for a moment– the name was given back when nearly everyone agreed that things needed to change. It was hard to get everyone to move in the same direction.

– Well, put together a programme and fund it well… and everyone will fall in place.

– Our budget was tighter and funding harder to mobilise. But that was not the hardest part; such a paradigm shift required that everyone did their part. It required concerted efforts. The way scientific researchers worked had already changed in that hardly anyone waited for their research results to be published in form of a paper. Not only did they exchange research results much earlier, many shared their lab notes, their code, and the like. Some put even the identification of the research question in the open –she smiles–. What I mean is that change was accelerated on the individual level; it was more difficult to bring about change also on the organisational level –she grins–. Our research organisations retain a fair level of self-government, which is why it was important to get them aboard. The new system was negotiated with them and with all the important stakeholders. Them and all the other organisations involved in the conduction, funding, and evaluation of research, including also publishers.

-– So you brought them to the table to start a lengthy discussion –the gentleman resumed the conversation–.

– We tried to make the topics, and especially the results of the process, as concrete as possible. One of the most important ones was standards, not just for the exchange of research data, code, and especially results but also for the acknowledgement and recognition of contributions, you know, the currency in the academic world. Another topic was fixing the broken peer review process.

– Acknowledgement of scientific contributions, why did you focus on them?

– Back then, impact was a proxy for us to learn whether public money was spent well. Counting publications and how often they were cited were the easiest things to count. But… new ways of doing science required new ways of putting value to a scientific inputs, outputs and impacts, be it providing an important insight to a discussion or a vital revision of code, or participating in and contributing to a research project. The goal here was to appreciate a wide variety of activities.

– And why you have such rigorous regulation in place concerning openness and privacy.

– You do not mean to suggest that openness just comes about without concerted actions, do you? That you can leave the whole issue to the dominant players? –After seeing her colleague's non-committal shrug she continues–. Leaving it to the market forces just would not do it. I know you went a completely different route."

– Indeed! I do not argue the openness in principle; we have made it all open for our businesses.

–You mean you had the CEO of Google write the regulations for you? –the EU policy-maker grins knowingly.

– That is an open secret. However, the utter transparency advocated by Google provides tremendous opportunities, especially for the private sector. It is astounding what they can do with all the personal information out there. Custom-made things that suit your personal needs that were unimaginable before," enthuses the Indian gentleman.

– But then, your openness seems rather selective, doesn't it? It does not apply to business enterprises, or does it?

– We needed, and still need, to protect intellectual property is what makes them innovate at all.

– European citizens would never have accepted this lack of privacy.

– You must be joking! –snorts the Indian gentleman.

– Not at all. There was a period where it seemed that they did not care, putting personal information out there on the internet for everyone to see. As it turned out, they were not disinterested in privacy they just cared more about comfort. So the solution was to ensure that privacy would not come at the expense of comfort. Built-in licences did the trick with data sets containing personalised data which could be (re-)used easily without violating any privacy regulation.

– Quite the hassle, I must say. Good luck with that! –grins the Indian colleague.

– Cheers! –winks the EU colleague and raises her glass

6.6 Ancient Now

Story line

ODS is transforming the way how research is done as regards speed, how people interact, how results can be proven, or how outcome can be re-used. In the future, we might look back on today as a period in which a lot of the ingredients were already available and several ODS trends and benefits already started to emerge. But also as a period where a lot of effort was required to get rid of barriers.

Technology/challenges

- Trusted methods to contact and network with people (no Spam, guaranteed delivery,…)
- Trusted rating systems (beware of rating system optimisation approaches as SEO, Google bombing Google washing etc. today)
- Shorten time-to-use (time required to make use of published data, methods, tools, …)
- Define and establish the right metrics (flexible, open enough <-> rigidity to ensure required continuity)
- Interoperability and re-usability (Interface, data and visualisation standards, …)
- Remove barriers (from publishing houses to technological barriers such as "wrong version of program – cannot open data")
- Integration of data and software, processes and tools

Background/macro perspective

A burden of legacy systems that seem to be no longer up to the task prevents us from the full potential of the digitalisation of research. Starting from an ancient email system that is mainly driven by uncertainty of delivery, spam and lack of trust, the encapsulation of data, methods and tools in their own spheres and standards, to the long publishing cycles of established journals. An ecosystem with metrics, trust mechanisms, critical reflection, data curation and certification for online tools and data repositories has still to emerge from emerging ODS technology enablers. Known enablers require further shaping, new enablers are still to be identified from removing today's barriers.

Story

Teacher: "Listen up children! Today's lesson is about a time not far away from our year 2025. It was on the edge of a global change in how people were doing research and science. It was the period towards 2015/16 that had almost all necessary ingredients available to form what we see normal today:

To have a trusted research and science environment based on reputation and openness with our today's understanding of standardised interfaces ensuring interoperability, metrics and quality measures. Back in time, people had to rely on results of a few dominant search engines and especially on proprietary algorithms / approaches that were not transparent or even public.

Therefore, each individual was more or less an expert in trying to find quick proofs for new research findings and searched for forum discussions etc. to judge if the content found was of a reliable source and of certain quality or not. To a large extent, the same search engine was used to find proofs. All of this happened although - amongst other risks - it was well known that search engine optimisation (SEO) was a living business branch.

If people wanted to contact people of their own or other research areas – for the purpose of exchanging ideas or in their strive to find proof for their found search results or research work – they had to try to find different ways through systems that were more or less efficient (lack of security, trust, guaranteed delivery, etc.)

Spam was dominant – the community failed to transfer the very old email system had failed to bring up a new trusted incarnation 2.0. People were drowning in messages of various platforms such as Linked In, Xing, Facebook, Research Gate etc. It was hard to distinguish real/fake or valuable/useless contacts. To prove if a person had a certain reputation or to get an impression of the work of a person the same search engine was used.

They were not able to use the outcome of another research team in a comfortable manner as we are able today. Often it was nearly impossible to reproduce a result or to build on someone else's results. It sometimes took years until someone else could prove results wrong.

As I told you before – almost all ingredients were nearly available at that time of 2015. New powerful computing systems were produced every year – computation power was not the main issue. But they had to aggregate a lot in a certain order, redistribute foci, create standards and old barriers had to be removed. One of those barriers were private publishing companies. They commercially published research outcome even where the research was publicly funded. Another barrier was the time and effort required to use and aggregate data, tools, methods and results encapsulated in their own formats, standards, logic and versioning.

But it was this lack of interoperability, the need for approaches to handle big data, create promising and flexible interface standards that was driving their funding programs towards 2020. This trend together with the growing number of digital natives (Generation Z) and the influence of the fast growing digitalisation of research formed what we can rely on today in 2025. We now have an ecosystem with metrics, trust mechanisms, critical reflection, data curation and certification for online tools and data repositories. A fast, reliable and very clever way to perform leading-edge state-of-the-art research and produce benefits for our valuable society!"

Teacher: “Now your exercise is, to go back in time and define the main actors such as digital natives, policy makers, professors, universities, companies, etc. Then come up with your interpretation of their role and influence in the transfer the old research system into our well know Open Digital Science based approach. And define the barriers of that time per main actor and what each actor had to gain and loose.“

7 ODS metrics

Interviews and literature review suggest that we have clearly entered a re-evaluation process of how research performance should be assessed. It is still highly unclear, though, what the destination of this journey will be. That said, what is clear already is that the transition will eventually lead to a new 'social contract' between science and society that will require new policy frameworks and newly developed core indicators for good science and research.

The consensus amongst experts with regards to performance indicators is that the current one – that merely considers the quantitative dimension of research performance and popularity/quality values (e.g. numbers of citations) – is outdated and needs to be complemented or amplified by indicators which take into account new opportunities offered by ODS. A new era of multi-dimensional metrics will, in addition to performance, have to equally consider other dimensions such as quality, relevance, transparency, or effectiveness in science. This is especially relevant for open and responsible science. However, the development and particularly the implementation of a new metrics system will take time and most likely replace the old system gradually rather than radically.

Ideas for alternative metrics circulate but both the discussion and new concepts are still in their infancy. The majority of experts argue that quality indicators need to become more important in the ODS era to reduce publication barriers, increase the transparency of and trust in the publication processes as well as the whole process of conducting science and research. The drastically increased dynamics of data production, provision, and re-use pose a challenge to maintaining data trustworthiness. Paired with the provision or dissemination of code, the new generation of indicators is required to assess transparency and reproducibility of both data and results. For policy makers, this means a shift away from the mere measurement of research output and so-called excellence.

The development of a new metrics system must be accompanied by a discussion on the theory of change that underlies the selection of core indicators. It is important that governments, HEI managers, and research organisations stop thinking that metrics are an end in their own right and start determining where the application of metrics is important and which metrics are relevant/helpful under which circumstances.

7.1 Objectives

7.1.1 Why do we need Open Science indicators? (Results of literature review)

Open science does not only open up new ways of creating and sharing knowledge, but also of disseminating results and individual components of the research process. How this diversification or diffusion of research is adopted – not just by the scientific research community but – by society is not only a matter of technological developments but also of changes in cultural practice. It is yet not clear how to monitor and measure the uptake and impact of OS practice, especially societal impacts.

Expectations of OS impacts are high. As summarised in a recent report of the OECD[84], the following positive factors are associated with OS:

[84] OECD:2015:18

- Improving efficiency in science
- Increasing transparency and quality in the research validation process
- Speeding the transfer of knowledge
- Increasing the knowledge spill-overs to the economy
- Addressing global challenges more effectively
- Promoting the engagement of citizens in science and research

It is argued that especially researchers in low income countries profit from OS and the possibilities to use and re-use data from other researchers. The same might be said for SMEs which rely on research data.

In the literature review and in our interviews with OS experts there is a general agreement that new indicators for the monitoring and assessment of scientific production and its impact need to be agreed on, in light of a major redesign of the scientific process provoked by OS. However, there is yet very little substance to build upon. Altmetrics show potential but have not taken off (yet). Challenges in the development of indicators are that the reliability, statistical validity, and generalisability of new forms of data are not yet fully understood. Open data are even more sensitive than OA publications and therefore less straight-forward in their promotion. Many questions concerning the legal framework, ownership issues (especially in the case of public-private partnerships), licensing systems, copyright law, etc. remain yet unsolved.

7.1.2 What is already being measured?

There have been several attempts to set new indicators for the successful uptake and impact of OS practices. One such example is the so called *open access citation advantage*[85], meaning the tendency that open access publications receive more citations relative to non-open access ones. Some studies analyse the correlation between citations counts, publication format (OA or non-OA), and the quality of articles to see if there is a quality advantage or quality bias according to publication format. In other studies the increased returns on investment in *research outputs with increased accessibility* have been calculated. The European Commission estimated that their open data initiatives can be expected to generate a yearly income of EUR 140 bn. (cf. EC 2010, *How Europe can gain from the rising tide of scientific data*). Other available statistics reveal the percentage of published scientific papers according to different levels of OA (green OA, gold OA).

A number of qualitative surveys have been released to capture the attitude of researchers towards OA publishing (e.g. “What scientists think about OA Publishing”, SOAP project, project-soap.eu). The survey revealed a significant gap between the researcher’s conviction that OA is beneficial for their research fields and the factual low ratio of articles published in OA journals.

The OECD Working Party of National Experts on Science and Technology Indicators (NESTI) announced to launch soon a survey on the behaviour of scientists and researchers which will include questions on open access and open data. Open data readiness is being monitored by the Open Data Barometer (opendatabarometer.org) by means of a cluster analysis of open government data readiness and impact variables. The Barometer measures also impacts of open data through mentions of data use and impact in media and academic channels.

Altmetrics

New evaluation systems are needed, because evaluation of research is currently based on teaching and bibliometric indicators that do not take into account a whole array of contributions

[85] OECD:2015:11

to and resulting from the research process (data, methods, codes, insights, ideas, trainings, participations in all kinds of activities, etc.). With this realisation, the idea of Altmetrics was born. The main dimensions that altmetrics cover in terms of impact was categorised by PLoS[86] as follows:

- Viewed – HTML views
- Downloaded/saved – as viewed plus saved, including saved on sites like CiteULike, Mendeley, and other social bookmarks
- Discussed –journal comments, science blogs[87], Twitter, etc.
- Recommended – F1000Prime
- Cited – citations captured by PubMed, Wikipedia CrossRef, Web of Science, Scopus, etc.

One can see the efforts to go beyond the traditional citation metric but it is yet unclear what the consequences of using altmetrics will be. “There is scarcely any research on the comparability of altmetrics and virtually no research on their potential manipulations and network effects.” (Fecher and Friesike:2014:43). Mingers and Leydesdorff (2015) go a step further and argue that there are a number of problems such as that altmetrics can be gamed by fake likes and tweets or that a high score may not mean that a work is good but controversial or fashionable.

The scope of altmetrics is fairly limited; the concepts involved in OS exceed that scope by far. An issue of special interest is, how open science accomplishes to stimulate “good science” in the terms of minimising current shortcomings such as “questionable proof-generating means, intolerance against uncommon theses and approaches, citation-based ‘truth generation’ and inflexible cultures of scientific approaches within disciplines” (Friesike & Bartling:2014:12). One of the objectives of our indicators is to find proxies for the assessment of “good science” in the philosophy of OS.

7.1.3 What do we want to measure?

One of the main objectives of this study is to propose a framework for an OS observatory which monitors the progress of OS in Europe on a continuous basis. Our OS indicators shall therefore be useful to monitor the uptake and impact of Open Science with an emphasis on the *digital* dimension (as the technological facilitator of OS.

Guiding questions for the development of the indicators were:

- How widely is ODS accepted/adopted in practice?
- How have different phases in the scientific process changed due to OS?
- What changes are being perceived on the system level?

Indicators shall measure if OS practices make science more accessible for a wider audience, whereby Fecher and Friesike see accessibility in the double sense: (a) accessibility of the research process and (b) comprehensibility of the research result[88]. This understanding suggests that the relationship between science and society must be reflected in the indicators in any case.

[86] Public Library of Science

[87] Or an aggregator like ScienceSeeker

[88] Fecher & Friesike:2014:19

7.1.4 Who are the key stakeholders in Open Science?

Different stakeholder groups have different motivations for and expectations from promoting OS, or from providing incentives for its uptake. We argue that the proposed indicators take into account the diversity of the main stakeholder groups. In their policy paper "Making Open Science a Reality", the OECD (2015) defined the key actors as follows:

- Researchers: their drivers to get active in OS are (a) motivations (cultural values, reputation, incentives, etc.), (b) necessities (standards, funding), and restrains (publish or perish, competition, additional burden: legal implications and contractual rules)
- Governmental bodies: interested in increasing the value of existing results; evaluation of new national OS strategies; they are bearing the costs for offering OA to and preservation of data; develop legal frameworks to create an open-science friendly environment; OS as part of the national innovation strategies; open government data;
- Research funding agencies: key actor in promotion of OS through definition of funding requirements and incentives; support the development of research infrastructure for OA and OS. Funding agencies may give incentives for the use of new metrics that take into account OS uptake in the research practice. The selection of evaluation criteria and indicators applied by research funding agencies is a strong instrument to shape researcher's behaviour.
- Universities & public research institutes: have certain autonomy in defining STI strategies and therefore OS policies and adopting OS mechanisms. May provide infrastructure and support; OS skill development & training; May uptake research on OS.
- Libraries, repositories, data centres: providing infrastructure for and provision of digital material: preservation, curation, publication and dissemination of digital scientific material; new importance of online repositories;
- Private publisher: offer services and infrastructure around OA publishing and related key services blogs, compute alternative metrics for research papers, support development of apps)
- Business sector, SMEs: beneficiaries; demand-side, public-private partnerships
- Supra-national entities: definition of international standards & principles; promotion; ensure interoperability of systems and standards and repositories; bring in position of developing countries;

7.2 Methodology

We want to emphasise that the here presented indicators only offer a first approximation and do certainly not cover all aspects of OS uptake and impact.

The crucial question is the combination of indicators which measure factors that change the research process.

The figure below depicts the activities and the kind of results they created. It shows the whole process from start to the writing of this report. Based on the 'future scenarios' created in the course of this project, a first set of indicator candidates were created. An literature review informed this initial set and provided further insights and some additions. A series of in-house workshops at the Centre for Social Innovation (ZSI) led to an improved, tailored set. Through a project-internal workshop, this set was made ready to be presented to a wider audience of experts in, or close to, OS in form of an online assessment/survey. This scrutiny led to a selection of the indicators that are considered most relevant to measure the uptake and impact of OS.

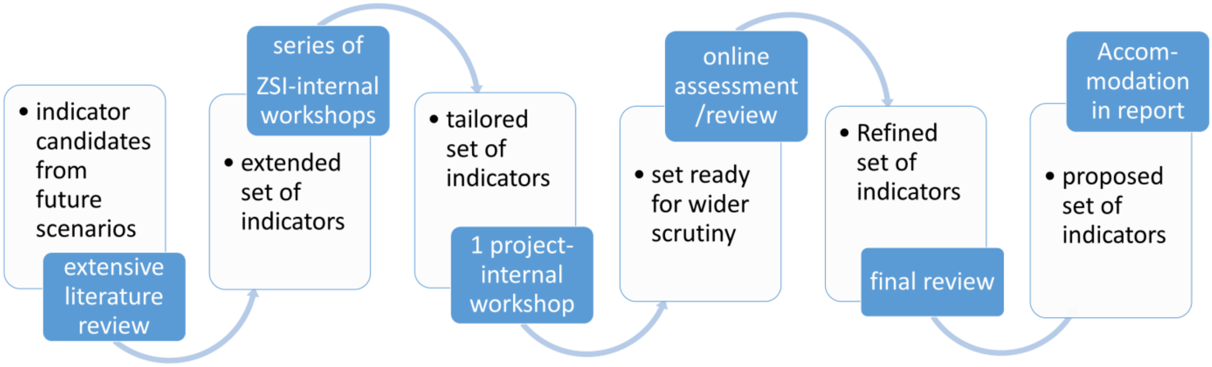


Figure 16: The process behind the proposed set of indicators

7.2.1 Online Assessment/Survey

As described above, the online assessment was a vital step in compiling the set of indicators proposed below. The survey was designed in such a way that a participant would only need to assess those dimensions that [s]he felt most experienced and interested in. The questions themselves were straightforward in that, after the introductory part, they simply asked each indicator to be assessed in terms of relevance to gain insights into OS uptake or/and impact. The following section presents the structure of the main part of the survey. The complete list of indicators we asked to be assessed can be found in the annex (cf. p. 93).

7.2.1.1 Survey structure and answer scale

The indicators were categorised into two major groups/dimensions (A and B, see below) and seven sub-dimensions. The scientific process is not regarded as static or linear but as highly dynamic and rather circular/iterative:

A) the scientific process:

- conceptualisation and data gathering/creation
- analysis
- diffusion of results
- review and evaluation

B) the system level:

- reputation system, recognition of contributions, trust
- open science skills and awareness
- science with society

- Answer scale: 0-10 (0..no relevance, 10.. highest relevance), see image below

Section C: **Scientific process: Conceptualisation & data gathering/creation**

C1. **Please rate which indicator you find most relevant to gain insights with regard to OS uptake or/and impact. (0..no relevance, 10..highest relevance)**

	0	1	2	3	4	5	6	7	8	9	10
# of research organisations that do open consultations to decide which topics to pursue in their research	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
# of proposals applying for funding of OS infrastructure creation and use	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
# of shared laboratories (online)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
early involvement of citizens	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 17: Exemplary rating of items (here, in the dimension *scientific process - conception and data gathering/creation*)

7.2.1.2 Promotion of the survey

A select number of roughly 120 individuals were addressed personally and invited to participate in the online assessment. The respondents were also asked to extend the invitation to colleagues they thought might contribute insights/experiences as well. For this reason, the total number of reached individuals is unknown (but expected to be low).

7.2.2 Challenges met

Developing adequate indicators is a challenge in itself and there are many questions, e.g. concerning the legal framework, ownership issues (especially in the case of public-private partnerships), licensing systems, copyright law, etc. that remain yet unsolved.

Our initial approach was to start with a literature review on the current status of open science in Europe and build on its results in terms of developing future scenarios and compiling an initial set of uptake and impact indicators. However, our literature review showed that little has been done so far with respect to indicators that is of substance, which is why we had to adapt our initial approach, i.e. we chose to

- ⇒ conduct creation future scenario first, then
- ⇒ scan future scenarios for potential indicator candidates,
- ⇒ do some desk research to check their feasibility and find complementary ideas, and
- ⇒ refine indicators with both internal and external expertise.

As a logical consequence, the online assessment came fairly late in the project; however, the results were still in time for the final report.

We included both quantitative and qualitative indicators. Most of the latter are perception indicators which can only be assessed by asking the target group, e.g. by sending a survey to researchers. We need perception indicators, because literature reveals a serious discrepancy between the theoretical appraisal of OS and the willingness/ability of the individual researcher to invest additional time and efforts to put OS into practice. Scheliga and Friesike (2014) address here a social dilemma: “[w]hat is in the collective best interest of the scientific community is not necessarily in the best interest of the individual scientist.” This issue cannot be resolved by us; it will have to be negotiated in the various science communities.

7.2.3 Lessons learned

- ODS not accepted as a new concept
    - Initially, the project consortium was tasked to test the concept of Open Digital Science; however, it became apparent that the general view was that Open Science is typically *digital* and that there was an agreement, that the *digital* facilitates Open Science, and, as a consequence, that a new label was not needed (even potentially confusing). NB: the online assessment was shortly adapted to reflect this notion
- Stakeholders need more prior involvement for them to spend their (valuable) time and share their experience, especially in the light of the many other on-going consultations (or recently concluded ones). Although those may not have been on metrics specifically but on the overall topic, they still require time and are a deterrent for any further/parallel involvement. Nevertheless, participants expressed an appreciation for the topic of OS indicators.

7.3 Analysis of results

This section concerns itself with only those indicators that were assessed sufficiently high. We picked as the minimum an average rating of 7.5 (on a scale from 0 to 10, 10 being the maximum). The full initial set of proposed indicators can be looked up in the annex.

Below, each sub-section is dedicated to one of the seven dimensions covered by the online assessment (cf. *Survey structure*, p. 65) and provides details for each indicator, such as the rating and proposed entities responsible for adopting/monitoring said indicator – the following colour codes and abbreviations are used:

R	researchers
RO	research (conducting) organisations
RFO	research-funding organisations
PM	policy-makers
PU	publishers

Table 8: Indicators that were assessed sufficiently high

7.3.1 Socio-demographic characteristics of respondents and responses per assessment dimensions

Overall, 34 full responses were submitted to our database. Out of 120 invited individuals, this means a return rate of about 28 %. 12 of the respondents stated to be female, 21 to be male; 1 person chose not to answer this question.

The following table shows the attribution of respondents to stakeholder groups (selecting multiple items was possible):

Stakeholder group	No. of participants
Researcher	19
Research funding	9
Research management	6
Policy maker	5
Private sector	2
Other	2
Citizen	1
Publisher	0

Table 9: Attribution of respondents to stakeholder groups (selecting multiple items was possible)

The two respondents in the category "other" specified as their stakeholder group *public administration* and *policy analysis*, respectively.

The following table shows the research domains which researchers stated to primarily work in:

Research topics	No. of researchers
Applied Sciences	9
Arts & Humanities	2
Economic & Social Sciences	9
Health Sciences	1
Natural Sciences	2
General Sciences	0
Other	2

Table 10: The research domains which researchers stated to primarily work in

The two elements in the "other" category are not truly other research topics but specification of the stated primary field (educational sciences and science of science and innovation).

The participants hailed from 16 different countries (see table below). However, the geographic distribution is tilted towards Austria, i.e. roughly a third of all respondents.

Country	No. of participants
Austria	12
Belgium	3
Brazil	1

Country	No. of participants
Bulgaria	1
Czech Republic	2
France	1
Germany	2
Lithuania	1
Netherlands	1
Poland	2
Portugal	1
Slovenia	1
South Korea	1
Spain	1
Sweden	1
United Kingdom	2
No answer	1

Table 11: The participants hailed from 16 different countries

As stated in section 7.2.1 (p. 65), the online assessment covered six dimensions. The table below shows the no. of responses in each of those dimensions. Most respondents wanted to comment on the review and evaluation block.

Research topics	No. of researchers
Scientific process – conceptualisation and data gathering/creation	14
Scientific process – analysis	13
Scientific process – diffusion of results	16
Scientific process – review and evaluation	21
System level – reputation system, recognition of contributions, trust	17
System level – open science skills and awareness	10
System level – science with society	16

Table 12: No. of responses in each of the dimensions

7.3.2 Cluster I: Conceptualisation & data gathering/creation

Important questions in this dimension are whether the quality of data and information is adequate, e.g. whether the data were properly cleaned, whether they are curated, are metadata provided, etc.

Brase et al. (2009) observe that “[i]n many scientific communities there is yet no standard data quality assessment protocol as it exists for scientific publications.” The OECD lists a number of criteria to determine what *good quality data* implies beyond the accessibility: data need to be intelligible, trustworthy and, of course, reusable. (cf. OECD:2015:58)

Open access policies are more mature than open research data policies, as data sets are in many cases difficult to define or not properly cleaned or linked to metadata. Other critical issues related to open data are privacy, security concerns, and intellectual rights issues (OECD:2015:87).

Recent policy trends have revealed that the majority of policy initiatives for the promotion of OS involve mandatory rules and requirements (most commonly, funding agencies mandate public access to funded research), and the development of infrastructure to enable OS. Fewer initiatives relate to non-monetary incentive mechanisms like the definition of new reward/promotion systems. (OECD:2015:87)

Several comments in the online survey made it clear that indicators for scientific work must no longer be restricted to measuring final products (such as articles), but should measure the development of the individual steps of the scientific workflow. Furthermore, results will differ according to disciplines, fields, or data types. The RIO[89] Open Science Journal, for instance, accepts the submission of research ideas, research proposals, review articles, data papers, software descriptions, data management plans, software management plans, research presentations, case studies, Wikipedia articles, and many other types.

Requirements from research funders	mean rating (0..10 max.)
% of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards	7.9
	RFO PM

Accessibility	mean rating (0..10 max.)
accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects	9.1
	R RO RFO

Machine-readable	mean rating (0..10 max.)
% of machine-readable data / metadata	7.9
	PU R RFO

Availability of metadata	mean rating (0..10 max.)
availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research)	7.5
	PU R RFO

Quality of metadata	mean rating (0..10 max.)
quality of metadata (versioning, volume, data format, description of fields, etc.)	8.2
	PU R RFO

Simulation results	mean rating (0..10 max.)
usability of simulation results (models, data, and code)	7.5
	R RFO PU

[89] Research Ideas and Outcomes, ISSN 2367-7163, URL http://riojournal.com

It should be noted that – with a rating of 7.1 – the related indicator *access to simulation results* was slightly below the threshold of 7.5, the two lowest ratings were given by researchers from *Economic & Social Sciences*.

Data services	mean rating (0..10 max.)		
(types of) open data services offered	8		
	PU	R	RO

Data compilation/publication costs incorporated	mean rating (0..10 max.)		
% of funded projects incorporating costs for data compilation / publication and maintenance (of the repository/data sets)	7.6		
	PM	RFO	RO

Long-term availability	mean rating (0..10 max.)		
is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no))	8.2		
	RFO	RO	PM

This indicator received one critical comment saying that the long-term availability is very important but impossible to guarantee in most cases. Therefore, it is more wishful thinking than a suitable indicator.

Sharing policies	mean rating (0..10 max.)
# of sharing policies in research organisations (sharing of data, organisms, etc.)	7.6
	RO

Apart from the indicators mentioned above, the "# of research organisations that do open consultations to decide which topics to pursue in their research" as well as the "early involvement of citizens" each received a rating of 5.6; interestingly, the variance of responses is fairly high. It may be worth picking them up in further discussions as some research fields may be more open to engaging citizens.

It was recommended to include in the metadata the methodologies of data gathering and also distinguish data types (e.g. input, output).

Suggestions of new indicators on the basis of survey comments:

- reuse-friendly licences of data and/or code (either separate or an explicit part of the accessibility of open code/data indicator)
- It was suggested that he absolute and relative numbers of published data management plans, grant proposals, lab notebooks, final reports, etc. would be much more informative, as would the delay between their creation and publication.

7.3.3 Cluster II: Analysis

None of the proposed indicators (cf. ANNEX) received a rating above the threshold of 7.5; the closest one is the "actual use of open methods". In the open answers part, the respondents argued that open methods contribute to improving the reliability of research results but that the impact of the open methods is still marginal because their use is still not spread widely in the research community. Nevertheless, the potential seems to be there, the respective indicator needs to be developed further.

As regards the data citations[90] and code/software citations, one respondent took issue with the wording, i.e. it was not made explicit that the data and code citations were citations of open data and open code – the rating would be substantially higher in the latter case. Another responded noted that citing code and data was only proper scientific conduct. Eliminating these two answers would push the rating of the # of data citations and # of code citations to about 7.5; we would argue that these two are not off the table yet.

Suggestions of new indicators on the basis of survey comments:

- access to underlying data and data structures
- access to the software
- # of open data citations
- # of open source citations
- # of open access citations
- # of open content citations (video, sound, blog, etc.)

7.3.4 Cluster III: Diffusion

We deliberately chose the term "diffusion" (of results) instead of the term most commonly used term in academia: "publication". We want to stress that *diffusion* can and – some would argue – should start well before the results are in. As Bourne (2010) concludes, "[…] the final published work does not map well to the workflow of the scientific endeavour used to create it. In the digital era there is no excuse for not doing better." In our online assessment, several comments underpinned the need to get away from the traditional paper publishing models and find indicators that gauge the growth of dissemination channels other than journals. Participants stated that journals are becoming irrelevant in many fields already; some fields strongly prefer conference papers, some active papers, some preprints, others books, or simply running code. We suggest monitoring the growth of each of those and the emergence of new channels. One respondent put forward that every step of the research cycle should be communicated, and feedback invited on it before that step even starts. The impact of OS can then be captured in those cases where that kind of open communication and responsive attitude to feedback has actually changed the trajectory of research, e.g. a side-line turned into the main thing, a bug/design issue was detected, or the project just responded (or even emerged in response) to what is happening in society. At the same time, it will be worth to observe whether a turn in research culture is taking place that allows scientists to feel safe to publish also negative or unsatisfactory results. It would further be important to measure the effective diffusion of scientific and technological knowledge transfer to specific target groups. This can probably only be tracked in the individual case, by comparing target groups (as defined in the research proposal), dissemination activities and actual outreach and feedback. Outreach to citizen will depend to a certain degree on the nature of the research

[90] platforms that may provide data on data citation: DataCite, ORCID , Figshare, The Dryard Digital Repository, ReseacherID.

project, respectively if the research question is of public concern. A RRI[91] indicator could be a proxy: "Documented ELSI/ELSA[92] project component and/or transdisciplinary component that addresses societal relevance and ethical acceptability (presence/frequency; qualitative descriptions; best practices)"[93]. Furthermore, it was suggested that the cooperation with publishers should be monitored and if the European Commission accomplishes to conduct negations with publishers, which would add value to the dissemination process. On the policy level, it can further be investigated whether well-designed mandates for openness are implemented by research funding and research performing organisations in the EU.

Open standards	mean rating (0..10 max.)
% of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)	8.2
	RO R

Free licencing	mean rating (0..10 max.)
% of publications with free licencing (public domain, attribution, all kinds of sharing)	8.6
	RFO RO PM

It was suggested in our online assessment that, instead of asking red tape questions, research funders should ask in their application form, how research objects – such as data, code, or materials – that the researchers shared in the past have been reused by others. The expectation is that this could help getting people interested in others using their "stuff" and that, instead of hindering it, as it often happens today, they would enable this.

Suggestions of new indicators on the basis of survey comments:

- measure the effective diffusion of scientific and technical knowledge transfer to specific target groups

7.3.5 Cluster IV: Review and Evaluation

Currently, peer review is the standard practice to assure quality of scientific output. Traditional peer review has well known shortcomings, though, such as little credit given to reviewers, lack of transparency and limited verification of scientific results[94]. Open peer review is often mentioned as an alternative, but not without the same amount of criticism. In the Open Science community, however, there is certain agreement that transparency measures need to be taken in the review and evaluation process. A multitude of suggestions is up in the air, some being considered as "incremental", meaning that they would not do much harm to the current review procedure, and others as "radical" or quite transformative. Adding transparency to the review process can happen at various stages of the scientific process and therefore be more or less transformative. One option would be to make grant proposals publicly accessible at various points of time (after the project has ended, along with the final project reports, at the beginning of a project, at the point

[91] Responsible Research and Innovation
[92] Ethical, Legal, Social Aspects/Implications
[93] EC:2015
[94] OECD:2015:50

of announcing funding decisions, upon submission to the funder and during the drafting phase)[95]. Another would be to make the peer review public. This can again happen in an incremental form, meaning that some knowledge within the peer review process is made openly accessible, or in a radical form, meaning that transparency of knowledge becomes a separate pillar of legitimacy itself.[96] Open peer review is currently a highly contested field and so is the choice of respective indicators. This can also be said for the question how societal relevance of research should be treated and assessed in evaluation. A rather easy measure could be to make the “impact statement” of a proposal publicly accessible. A labelling system for expected impact (oriented on e.g. the Sustainable Development Goals) could be an option to create clearer evaluation references.

Openness in calls for proposals	mean rating (0..10 max.)
openness in call for proposals (open proposals, open submissions, open review)	7.8
	PM RFO RO

Review criteria	mean rating (0..10 max.)
% of peer reviews that include reproducibility and transparency as review criteria	7.7
	RFO PU

Suggestions of new indicators on the basis of survey comments:

- % of transparent reviews (with distinction of different levels of transparency):
    - % of funding bodies that publish the final assessment of funding proposals
    - % of funding bodies that publish the successful proposals (along with their reviews and/or assessment summaries and/or final reports)
    - % of funding bodies that publish members of the review panels (ex-ante or ex-post)
    - % of funding bodies that publish the “impact statement” which is part of the grant application
    - % of reviews performed by non-scientists
- Evaluation criteria (of funding agencies and/or journals) require references of OA activities of applicant (Yes/No)
- % of open access publications referenced in publication lists submitted for evaluation
- Evaluation criteria include commitment to potential impact on society (e.g. social impact labelling based on internationally recognized criteria (Societal challenges in H2020 or SDGs, see Rio practice) (Yes/No)
- % of retractions of articles
    - also: the evolution of retractions

7.3.6 Cluster V: Reputation system, recognition of contributions, trust

The uptake of OS practice in the research process is unlikely to flourish if researchers fear it is not properly acknowledged and officially recognised. This is underpinned in the initially mentioned

[95] Mietchen, D, The Transformative Nature of Transparency in Research Funding

[96] D., Milanesi, E., Koenig, T.(2014): Grant Application Review: The Case of Transparency

surveys on researchers attitudes towards OS, which reveal low factual progress in putting OS into practice. Reward mechanisms for data sharing are currently especially weak and researchers might choose rather not to spend a serious amount of time in cleaning and curated their data for the re-use of others. Some organisations (datacite, ORCID, Figshare, Dryad Digital Repository, ResearcherID) have propositions for data citation tools which would credit authors for data and metadata sharing, but "in most countries the existing framework does not promote sharing efforts, especially with respect to results, data sets or other research material at the pre-publishing phase"[97]. Formal recognition of a variety of contributions along the scientific process (e.g. to the selection of research topics, formulation of hypotheses, project participations, review activities, etc.) have yet to be adopted.

Data communication as valued scientific contribution	mean rating (0..10 max.)		
data communication recognised as criterion for career progression (yes/no)	7.5		
	RO	R	PM

Suggestions of new indicators on the basis of survey comments:

- % of publications in Open Access Journals
- % of publications in Open Access Journals with no Impact Factor
- availability of means to easily publish negative results

7.3.7 Cluster VI: OS skills & awareness

This was the least "sexy" dimension of the indicators set in terms of number of respondents – only ten individuals chose to provide their assessment.

OS-related skill development across disciplines will be a crucial factor for the maturation of OS in Europe. Researcher's skills in OS (e.g. curating and maintaining large data sets) differ across disciplines due to different traditions or training opportunities in digital tools and data handling. There is a substantial need for further training of researchers and scientists in handling big, multi-layered and complex data sets. Proper curation and dissemination of data sets is challenging, costly and time consuming and will probably require adequate incentives in order to be adopted by researchers across disciplines. At the same time, citizens need to be trained to make use of and get involved in OS.

Skilled personnel	mean rating (0..10 max.)	
% of research personnel / research disciplines skilled in OS	8.3	
	RO	R

Active personnel	mean rating (0..10 max.)	
% of research personnel active in OS	8.4	
	R	RO

Curricula for the development of OS skills	mean rating (0..10 max.)
% of curricula that include OS skills (also prior to higher education)	8.1

[97] OECD:2015:89

	RO	

Researchers aware of standards	mean rating (0..10 max.)	
% of research personnel aware of standards (is there a standard (relevant to open science), how to adhere to it, etc.)	8.1	
	RO	

Researchers familiar standards	mean rating (0..10 max.)	
% of research personnel familiar with those standards	8.4	
	RO	

Researchers signing an open science pledge	mean rating (0..10 max.)	
# of researchers having signed an open science pledge	7.5	
	R	RO

Researcher organisations signing an open science pledge	mean rating (0..10 max.)	
# of research organisations having signed an open science pledge	7.9	
	RO	

Suggestions of new indicators on the basis of survey comments:

- # of trainings hold in writing/data sharing/coding/oer/dissemination/colaboration/evaluation/reproducibility

7.3.8 Cluster VII: Science with Society

This cluster is about finding indicators, which assess effects of OS on the promotion of the engagement of citizens in science and research. As Mietchen, Mounce, and Penev (2015) observed, most of the research process is hidden from public view through multiple layers of obfuscation as a heritage of conventions and habits from the paper era. This has begun to change, also because digital technologies enable engagement and popularisation. Popularisation activities are understood as targeting a wide audience and a non-specialised public. So far there has been a tendency that peers devaluate scientists who engage in popularisation, because it had a somewhat attributed negative correlation with scientific excellence. A statistical study[98] on the correlation between dissemination activities and academic records of more than 3500 scientists shows that scientists who are engaged in dissemination, contrary to general belief, are on average academically more active. The authors of the study therefore suggest to “taking into account ‘scientific culture popularisation actions’ for the evaluation of researchers”. Openness for the sake of science with and for society is of course more than popularisation activities and is an essential permanent factor throughout the scientific process.

[98] Jensen, P., Jean-Babstiste, R., Kreimer, P., Croissant, Y. (2008): Scientists who engage with society perform better academically, http://arxiv.org/abs/0810.4672

Citizens' engagement in (open) science	mean rating (0..10 max.)
increase in % of citizens engaging in open science	7.6
	R RO PM

Research communication beyond academia	mean rating (0..10 max.)
circulating and communicating research results outside the academia is standard (yes/no)	9.1
	RO PM

Accessibility of data that is of public interest	mean rating (0..10 max.)
provision of affordable sets of public interest data / metadata	8.1
	PM RFO

Closing the global gap in access to information	mean rating (0..10 max.)
advancement in closing the gap between the information rich and the information poor	7.7
	PM RFO

Suggestions of new indicators on the basis of survey comments:

- % Institutions rewarding researchers active in popularization of science
- # of research projects initiated/commissioned by civil society organisations or citizens
- # of ex-ante proposal evaluations including of civil society actors
- % of scientists active in OS and in popularization as % of total active scientists
- % of scientists active in OS and in industrial collaboration as % of total active scientists
- % of scientists active in OS and in teaching as % of total active scientists

7.4 Selection of indicators for an OS observatory

The Open Science observatory is the idea of an online platform that offers data on a number of observed indicators to measure the uptake and the impact of OS. As many of the indicators proposed in previous sections are not (yet) monitored systematically, it may be necessary to pick a mixed approach of observing the ones that are easier to monitor from a technological/technical point of view, design new ways with other major stakeholders, and conduct periodic interviews with concerned parties to gauge attitude changes and developments that are difficult to monitor.

The table below consists of indicators that, in our estimation, should be fairly easy to monitor from a technical point of view; it remains to be seen what their relevance is once a more comprehensive set of indicators has been developed.

It should be noted that it might be easier to observe actual numbers rather than relative ones (percentage) – as they are used in the indicators below – because the total/overall figures may not be known and/or readily available. Therefore, it may be a compromise to observe actual numbers as a precursory step. They could be related to total figures as soon as those became available.

Indicator	Cluster
% of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards	Data Gathering
accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects	Data Gathering
% of machine-readable data / metadata	Data Gathering
quality of metadata (versioning, volume, data format, description of fields, etc.)	Data Gathering
availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research)	Data Gathering
usability of simulation results (models, data, and code)	Data Gathering
(types of) open data services offered	Data Gathering
is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no))	Data Gathering
% of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)	Diffusion
% of publications with free licencing (public domain, attribution, all kinds of sharing)	Diffusion
% of peer reviews that include reproducibility and transparency as review criteria	Review
data communication recognised as criterion for career progression (yes/no)	Reputation System
% of research personnel / research disciplines skilled in OS	Skills
% of research personnel active in OS	Skills
% of curricula that include OS skills (also prior to higher education)	Skills
% of research personnel aware of standards (is there a standard (relevant to open science), how to adhere to it, etc.)	Skills
% of research personnel familiar with those standards	Skills
# of researchers having signed an open science pledge	Skills
# of research organisations having signed an open science pledge	Skills
openness in call for proposals (open proposals, open submissions, open review)	Science & Society
increase in % of citizens engaging in open science	Science & Society

Indicator	Cluster
circulating and communicating research results outside the academia is standard (yes/no)	Science & Society
provision of affordable sets of public interest data / metadata	Science & Society

Table 13: Overview of core indicators

8 Conclusions

8.1 Digital and open science: moving targets in a dynamic environment

This exploration of recent and developing trends in science has once more clarified the huge impact of digital technologies on current practices of researchers and scientists. Our analysis suggests that it still early days in this development. A new generation of researchers – known as the digital natives – is likely to approach key processes in the scientific work environment fundamentally different from previous generations. The feedback from the community also suggests that the changes are happening quickly, not uniformly across scientific disciplines and also in an environment that is strongly developing driven by digital trends, but also reacting to these trends. This provides a huge challenge for policy makers and research managers. As digital and open science are evolving in a dynamic environment they are basically moving targets that - while still not fully understood - require configuration and formation.

We have seen that at the conceptual level the development of open science is paralleled by a multitude of terms used to describe recent phenomena – e.g. *Science 2.0* or *e-science*. Most of these terms are neither clearly defined nor clearly delineated from each other. Some terms play a key role in policy papers and in the formation of communities and there are various promises associated with them. For example, *Science 2.0* is often connected to new forms of communication that are expected to facilitate inter-science discourse, but also to support the communication with stakeholders outside the scientific arena. Similarly, *Citizen Science* carries an element of re-contextualising science much broader in society and *Open Access* is a term that is now widely used to describe emerging changes in the triadic relation of science, society, and publishing institutions.

These conceptual diversity reflects the multiplicity of policy challenges and design options in research, science, and innovation management connected to current trends. It may even be fair to say that we are facing the biggest challenge for RTDI policy makers since the advent of the modern organization of science. This is one the one hand due to the enormous hopes and promises connected to open science, such as democratisation of science, broad and free access to scientific results, and improved innovation capabilities. On the other it is due to what may soon be radical changes to traditional scientific institutions. Open science has already started to significantly change the publishing market, but it may also impact on education and more generally role and self-image of universities. There may a new role for public labs, for research facilitators (science entrepreneurs), public associations and citizens to play – all facilitated by open science and new technologies.

These technological underpinnings of open and digital science that we aimed to clarify in this study indicate a strong proximity of open science and digital technologies that root deeper than just using new ICT for publishing. It are indeed the very technology characteristics of computing and networking that facilitate what we mean with *science* such as networked exchange of views, critical reflection and discourse, exchange between groups of experts, or reproducibility – to name just a few.

We have seen how difficult it is to precisely grasp and measure the evolution of open science and related practices. Although there is substantial discussions on new forms of metrics as regards impact assessment, for example, no commonly agreed set of indicators has emerged yet. It is not even clear, which facets of the new development deserve most attention. Our analysis points to a few broadly accepted tracers such as those related to data collection, diffusion or skills. Several of these indicators should also be relatively easy to implement – most notably those based on digital data – while others will inevitably require more elaborate and costly evaluation and qualitative assessment.

8.2 Metrics recommendations

Designing indicators to measure the uptake and impact of Open Science is a challenge, not least of all because the concept itself is still evolving. In addition, Open Science is necessarily broad because it is composed of many dimensions (e.g. along the scientific research process) and embedded in a larger system that involves e.g. new skills, a new reputation scheme, or the wider public. Apart from measuring the right "thing", an indicator needs to be sufficiently precise and measurable to allow its monitoring over time. However, most indicators proposed in this report are new and not gathered/surveyed/evaluated automatically (yet). Consequently, a first vital step is put the necessary mechanisms in place. To achieve this, the stakeholder groups that are primarily involved in/responsible for an indicator[99] are provided in the previous chapter.

In general, an essential precondition for indicators to work as intended is that all concerned stakeholder groups are involved in their design and evolvement. They all need to agree on what an indicator should measure (and what it should) and how it should be used (and what it must not be used for). Furthermore, indicators need to be flexible enough to accommodate differences, e.g. in research fields, and allow the emergence of new developments. The differences in research fields can be considerable, as is the pace at which OS is being adopted. Those differences will need to be elaborated and reflected in the relevant indicators.

Together, the concerned stakeholders could develop use cases and determine existing impediments for sharing and re-using contributions and results, and decide how many of those impediments can be resolved under certain conditions. It may also be insightful to observe how such impediments evolve over time.

When developing indicators, it will be necessary to determine which one makes sense on which level, e.g. the individual, organisational, national, or supranational level. Importance should be lent to developing indicators that measure the quality of research, not just the quantity of outputs. This was a recurring request from the experts included in our research. Since the two overlap, the Open Science indicators should be in line with Responsible Research and Innovation indicators and concepts.

It is generally recognised that citation counts are inadequate to measure the quality of a scientific work, let alone its impact. At the same time, data/code/content/etc. citations are expected to become functional in overcoming the reliance on text citation counts. Having said that, the planning of phasing out citation counts in favour of more adequate measures should start sooner rather than later.

Our overall recommendation is for all stakeholder groups to engage and, based on the work previously done, continue the discussion on how Open Science – and especially the surrounding concepts – can be measured in terms of uptake and impact on scientific research per se, the RTDI system, and society as a whole. Finally, new indicators need to be tested – not just discussed – before being adopted on a larger scale. This can be done in small experiments with individual, selected indicators.

[99] designing, measuring, interpreting, and/ or adapting an indicator

8.3 General recommendations

It is evident that the current challenges and changes related to open science should be of concern to all actors in the policy arena: this includes policy makers, funding agencies, research organizations, researchers and industry.[100]

Policy makers and open science funding agencies

The dynamic nature of both digital technologies and open science make it necessary to continue in-depth discussions on how to shape the future of open research. This includes coordination of policies and in some areas even the discussion on general objectives for policy-making. Given its open nature, these discussions need to be designed as participatory processes and it is recommendable that online tools play an important role here.

It will also be necessary to strengthen regulatory frameworks to develop a coherent and more harmonized environment. This regards, for example, a more coherent approach to open publication, but also open data. It is important to recognize the currently different approaches that exist not just across agencies and funding schemes, but also in the different scientific disciplines.

There is an opportunity now to actively shape robust and sustainable science with society engagement and industry implications as new tools and platforms are emerging and while the interest in this area is strong.

- FRAMEWORKS: Promotion of frameworks, in particular in Europe that include common standards to ensure a broader uptake. Areas include legal issues, resources management policy (e.g. data, shared infrastructures, analysis tools) and the promotion of these standards.
- MAINSTREAMING INCENTIVES: Launching programmes specific to open science and digital tools in order to increase the uptake and understanding in this strategically important area.
- INCREASING KNOWLEDGE: Public programmes should contribute to a deeper analysis of open science practices and outcomes and to promote EU-wide policies (e.g. data, privacy, IPR, access, publications, ethics etc. in the context of European Digital Agenda and Digital Single Market). It will remain important to identify, monitor and disseminate key elements as well as best practices.
- FINDING OPORTUNITIES: Broadening uptake of new funding schemes (crowdfunding, capital seeds, challenges, prizes….) to ease the participation of grassroots initiatives, digital communities or independent citizen scientists in a practical and direct way. Examples include open calls for individuals for fellowships mixing public-private investments.
- COORDINATING EFFORTS: Sharing best practices and promoting open science-related joint efforts (such as monitoring or roadmap development) to further harmonize policies.

[100] The following recommendations were derived during the process of the study and emerged either from interaction with stakeholders or from the analysis of the study team. They were presented both to the Advisory Board and on the internet site of this study for publication and discussion with the scientific community.

- OPENING the share of knowledge at a global scale in agile and inclusive formats such as challenges. Publishers must ensure that all publicly funded research is freely shared

Research institutions

Open science can be an opportunity for research institutions to strengthen excellence, creativity and create increased impacts. Harvesting these opportunities will require strategies and concrete policies that clearly support these strategies.

- PROMOTING CULTURE: To support a culture of openness, scalability and share promoting new collaboration channels with long-tail researchers and society. Key issues include quality, reliability, and interoperability in open science.
- COLLABORATION: Innovative methodologies and transdisciplinary groups are needed to address the wide range of impacts.
- ENSURING ACCESS: Research organizations, industry and infrastructure providers can contribute by extensively sharing resources (source code, public data sets, tools, methods, computing and storage facilities) and broadening access to science infrastructures, e.g. open science commons. The maintenance of accessibility services and usage of those resources should be adopted as criteria in existing and future funding schemes
- PUSHING FORWARD: Refine ODS management plans, ODS metrics and ODS quality assurance & assessment plans. Expand current research and innovation evaluation criteria to account for social impact and engagement.
- PLAN THE LONG-TERM: To collaborate avoiding duplications and developing sustainable collaboration and business models
- PERFORMING & SHARING Methods and infrastructures should be clearly defined, published and verified subject to reproducibility , accountability, re-usability, adaptability to new applications.

Researchers groups, communities and practitioners

At the level of research groups and communities, it is particularly important to strengthen human-technical capabilities and assessment. This includes the training of researchers on current open science aspects and technologies – both for established scientists and young researchers.

- TRAINING: Design of effective training programmes addressing specific issues of ODS such as interdisciplinarity, analytical and social skills. Provide educational resources as well as guidelines and dissemination plans for all the stakeholders.
- CAREERS: Consider new mechanisms for career recognition and awards that take into account the new trends towards open science.

Industry, including supply-side and demand-side

Open science offers new opportunities for the industry as regards access to research results, new forms of industry-academia cooperation and staff training. The challenges regarding protection of intellectual property may require revised policies and clear strategies in close co-operation with research organizations.

The trend towards digital tools will continue. Given the expected growth of open science, there will be huge market opportunities. This will provide a range of opportunities for businesses including university start-ups and small businesses.

- COLLABORATING: Innovative methodologies and transdisciplinary groups are needed to address the wide range of impacts.
- STRATEGY: Industry actors experienced in research cooperation should take the lead in developing workable agreements for open access to infrastructure, publications, and data that exploit the advantages of open science while achieving a balance with industrial interests to protect its resources.

9 Annexes

9.1 Data management plan

Project Name: Open Digital Science Study; Grant Title SMART 2014/0007 Open Digital Science This study is performed under the European Commission tender SMART

DESCRIPTION:

2014/0007 Open Digital Science was prepared by eutema GmbH, the Institute for Biocomputation and Physics of Complex Systems of the University of Zaragoza, and the Centre for Social Innovation (ZSI). It is based on our understanding of your objectives and specified deliverables defined in the tender specifications and includes our view of the key issues and challenges for Open Digital Science.

Funder European Commission (Horizon 2020)

Principal Investigator / Researcher Erich Prem

DMP QUESTIONS

For each one of this issues, please find related information in the next section “DATA SOURCES”.

Scientific research data should be easily:

1. Discoverable

Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?

→ Yes

2. Accessible

Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses?

→ Yes

3. Assessable and intelligible

Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review?

Data generated by the ODS portal is provided so that judgments can be made about their reliability and the competence of those who created them.

→ Yes

4. Usable beyond the original purpose for which it was collected

Are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data?

→ Yes

5. Interoperable to specific quality standards

Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc?

→ Yes

Among other technologies, ODS uses: node.js, Leaflet + Leaflet.draw, AngularJS, socket.io, CouchDB, Grunt, Bower, csv, xml, DW Q&A, Anspress, GoogleAnalytics, Wordpress

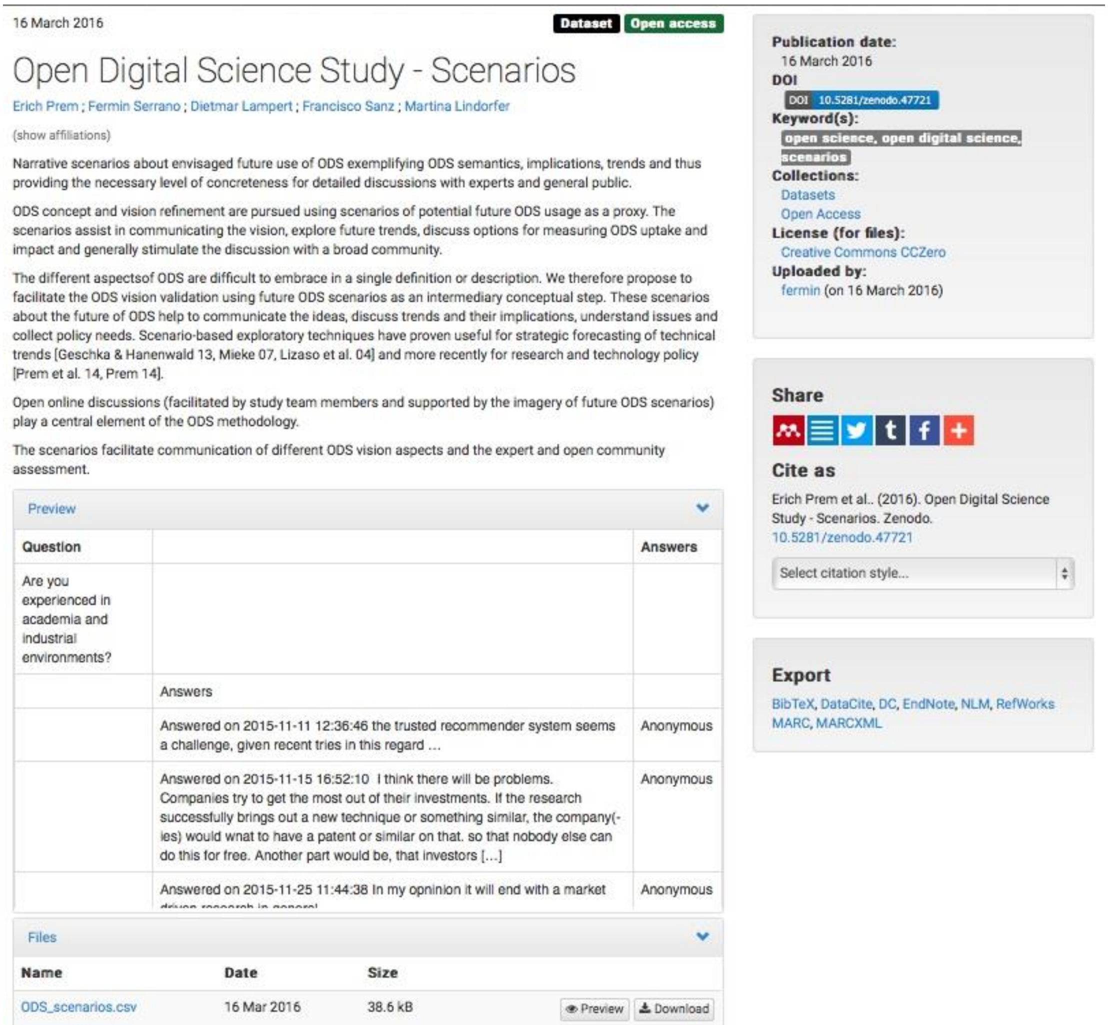
DMP DATA SOURCES

1. Scenarios data (story, background, comments)

https://zenodo.org/record/47721

DOI: 10.5281/zenodo.47721

License (for files): Creative Commons CCZero



16 March 2016 Dataset Open access

Open Digital Science Study - Scenarios

Erich Prem ; Fermin Serrano ; Dietmar Lampert ; Francisco Sanz ; Martina Lindorfer

(show affiliations)

Narrative scenarios about envisaged future use of ODS exemplifying ODS semantics, implications, trends and thus providing the necessary level of concreteness for detailed discussions with experts and general public.

ODS concept and vision refinement are pursued using scenarios of potential future ODS usage as a proxy. The scenarios assist in communicating the vision, explore future trends, discuss options for measuring ODS uptake and impact and generally stimulate the discussion with a broad community.

The different aspectsof ODS are difficult to embrace in a single definition or description. We therefore propose to facilitate the ODS vision validation using future ODS scenarios as an intermediary conceptual step. These scenarios about the future of ODS help to communicate the ideas, discuss trends and their implications, understand issues and collect policy needs. Scenario-based exploratory techniques have proven useful for strategic forecasting of technical trends [Geschka & Hanenwald 13, Mieke 07, Lizaso et al. 04] and more recently for research and technology policy [Prem et al. 14, Prem 14].

Open online discussions (facilitated by study team members and supported by the imagery of future ODS scenarios) play a central element of the ODS methodology.

The scenarios facilitate communication of different ODS vision aspects and the expert and open community assessment.

Preview

Question		Answers
Are you experienced in academia and industrial environments?		
	Answers	
	Answered on 2015-11-11 12:36:46 the trusted recommender system seems a challenge, given recent tries in this regard ...	Anonymous
	Answered on 2015-11-15 16:52:10 I think there will be problems. Companies try to get the most out of their investments. If the research successfully brings out a new technique or something similar, the company(-ies) would wnat to have a patent or similar on that. so that nobody else can do this for free. Another part would be, that investors [...]	Anonymous
	Answered on 2015-11-25 11:44:38 In my opnion it will end with a market driven research in general	Anonymous

Files

Name	Date	Size	
ODS_scenarios.csv	16 Mar 2016	38.6 kB	Preview Download

Publication date: 16 March 2016
DOI: DOI 10.5281/zenodo.47721
Keyword(s): open science, open digital science, scenarios
Collections: Datasets, Open Access
License (for files): Creative Commons CCZero
Uploaded by: fermin (on 16 March 2016)

Share

Cite as

Erich Prem et al.. (2016). Open Digital Science Study - Scenarios. Zenodo. 10.5281/zenodo.47721

Select citation style...

Export

BibTeX, DataCite, DC, EndNote, NLM, RefWorks MARC, MARCXML

Figure 18: Scenarios data (story, background, comments)

2. ODS practices data Posts - Cases

https://zenodo.org/record/47770

DOI: 10.5281/zenodo.47770

License (for files): Creative Commons CCZero

You may use this file to transfer that content from OpenDigitalScience.eu site to another. This file is not intended to serve as a complete backup of OpenDigitalScience.eu site.

This is a WordPress extended RSS file generated by WordPress as an export of the OpenDigitalScience.eu site.It contains information about OpenDigitalScience.eu site's posts, pages, comments, categories, and other content. You may use this file to transfer that content from OpenDigitalScience.eu site to another.

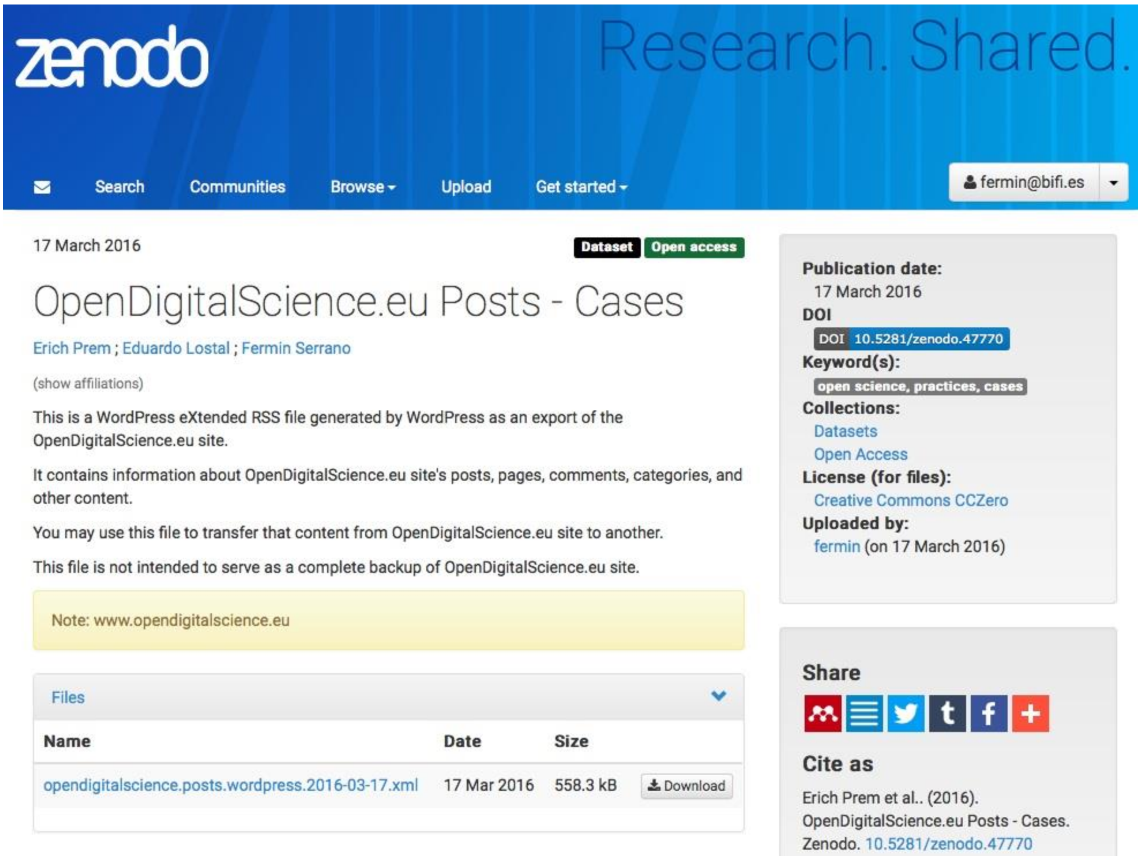


Figure 19: WordPress extended RSS file generated by WordPress as an export of the OpenDigitalScience.eu site

About ODS practices, data is provided with instructions to import this information into a WordPress site follow these steps:

```
<!-- 1. Log in to that site as an administrator. -->
<!-- 2. Go to Tools: Import in the WordPress admin panel. -->
<!-- 3. Install the "WordPress" importer from the list. -->
<!-- 4. Activate & Run Importer. -->
<!-- 5. Upload this file using the form provided on that page. -->
```

```
<!-- 6. You will first be asked to map the authors in this export file
        to users -->
<!-- on the site. For each author, you may choose to map to an -->
<!-- existing user on the site or to create a new user. -->
<!-- 7. WordPress will then import each of the posts, pages, comments,
        categories, etc. -->
<!-- contained in this file into your site. -->
```

3. ODS map source code

https://github.com/frasanz/Ethermap.git

License Apache v2.0

- Ethermap is a real-time collaborative map editor allowing:
- synchronization of geoobjects between all clients
- visual highlights of changes creating user-awareness
- watching other users or show their current workarea
- basic feature version control (browse older revisions and revert changes)
- communicating about specific features within the chat

Technologies

- node.js
- Leaflet + Leaflet.draw
- AngularJS
- socket.io
- CouchDB
- Grunt
- Bower

Install dependencies (Ubuntu)

It is assumed that you have installed node.js (developed using 0.10.26)

```
sudo apt-get install couchdb
npm install -g grunt-cli
npm install -g bower
npm install -g forever
```

Run for Development

```
npm install
bower install
grunt serve
```

Run for Production

```
npm install
bower install
grunt build
NODE_ENV=production forever -o out.log -e err.log start dist/server.js
```

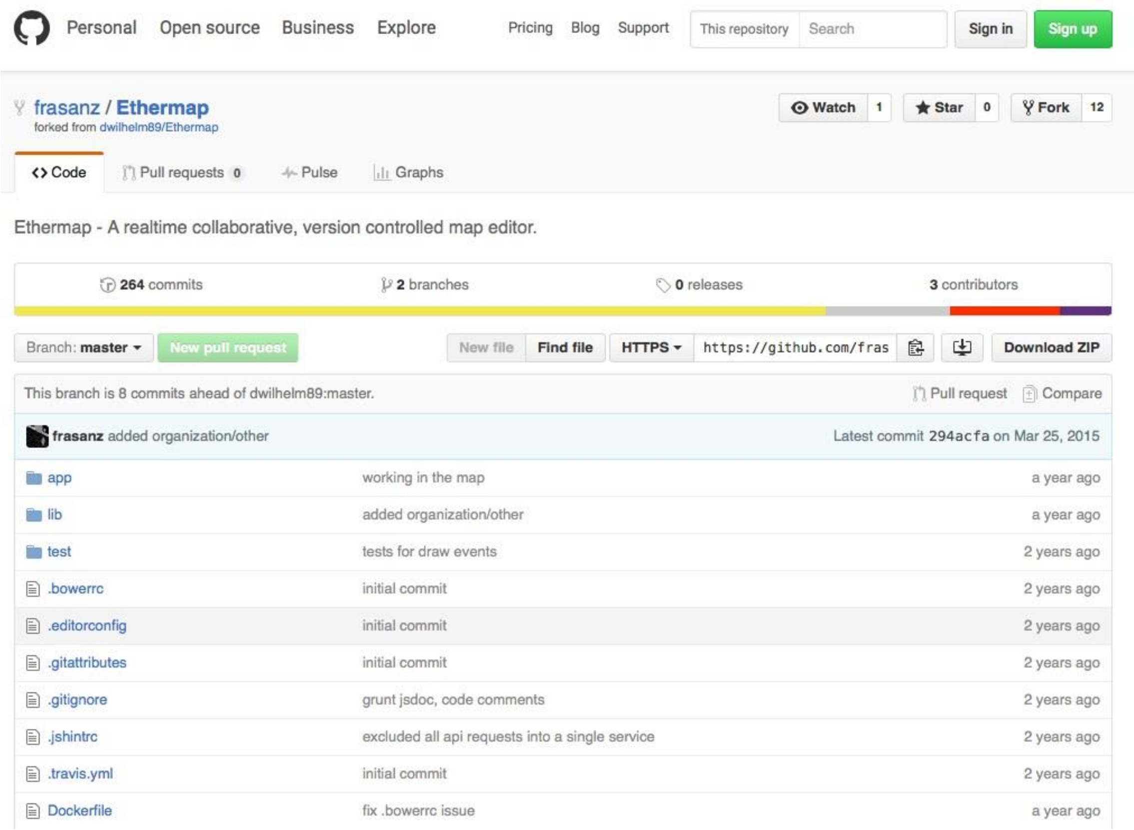


Figure 20: Source code repository

4. ODS initial set of indicators

https://zenodo.org/record/48991#.VwOpEIu-nBJ

DOI: 10.5281/zenodo.48991

License (for files): Creative Commons CCZero

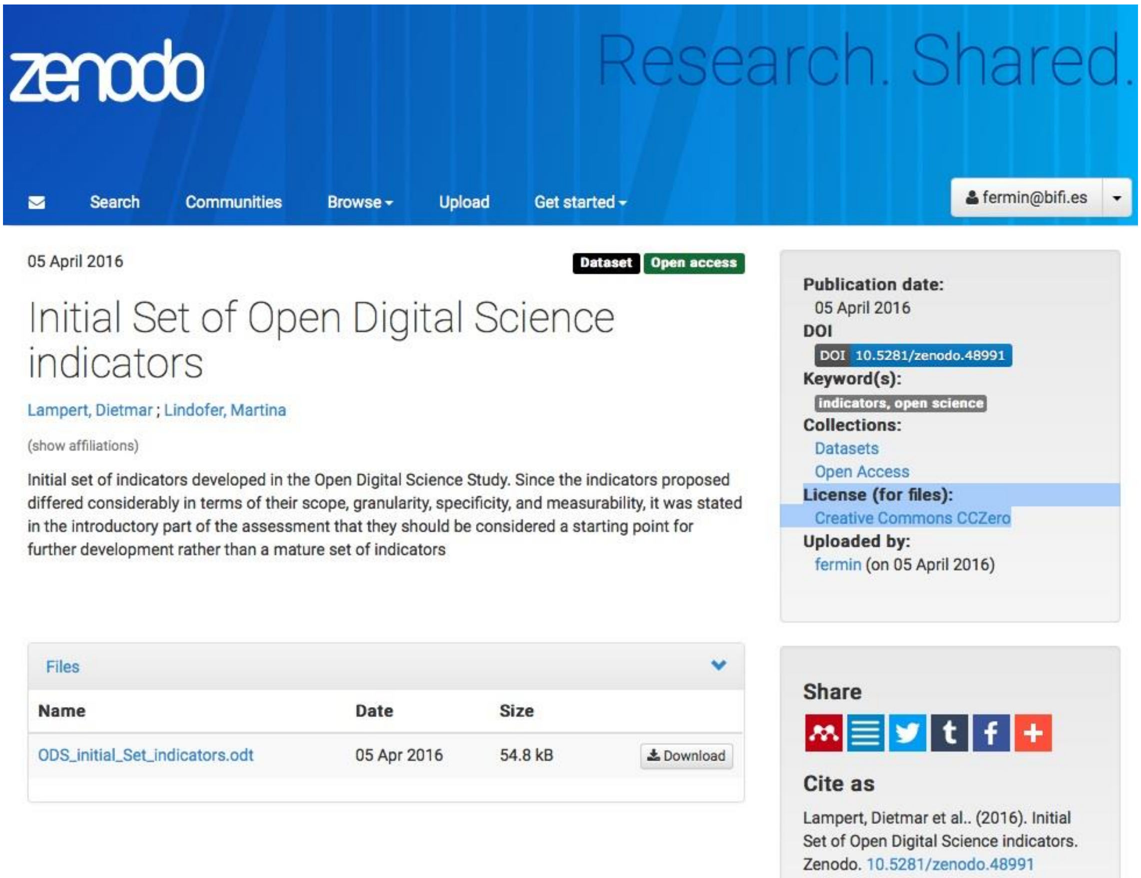


Figure 21: Initial set of Open Digital Science indicators

9.2 Towards a permanent observatory

During the course of the study, the European Commission asked RAND Europe, Deloitte, Observatoire des Sciences et des Technologies (OST), Altmetric and Digital Science to develop a monitor for tracking trends in Open Science in Europe. The Open Scinece Monitor[101] will include focus groups and user testing with an emphasis on policy makers.

The monitor as it is developed is likely to include bibliometrics, altmetrics, data mining and interview. It is currently ongoing and can benefit from taking a broader perspective that also includes the broader view on open science presented in this study. It is therefore suggested to design the monitor to address the various stages of the scientific work flow, in particular:

- Data collection and analysis
- Diffusion of results
- Review
- Reputation system
- Open science skills
- Open science and society

[101] https://www.rand.org/randeurope/research/projects/open-science-monitor.html

In a first step, the focus should be on mostly quantitative indicators that can be evaluated more easily than qualitative indicators, in particular the following:

Indicator	Cluster
% of machine-readable data / metadata	Data Gathering
(types of) open data services offered	Data Gathering
% of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)	Diffusion
% of publications with free licencing (public domain, attribution, all kinds of sharing)	Diffusion
% of peer reviews that include reproducibility and transparency as review criteria	Review
% of curricula that include OS skills (also prior to higher education)	Skills
# of researchers having signed an open science pledge	Skills
# of research organisations having signed an open science pledge	Skills
openness in call for proposals (open proposals, open submissions, open review)	Science & Society
increase in % of citizens engaging in open science	Science & Society
provision of affordable sets of public interest data / metadata	Science & Society

Table 14: Selection of quantitative indicators

In a second step, it will be necessary to include qualitative analysis to cover the whole range of indicators proposed in Section 7.

9.3 References

Alperin J.P., Babini, D., Fischman, G.E.: Open Access indicators. Assessing growth and use of OA resources from developing regions - the cases of Latin America and Africa (2013), developed by UNESCO, CLACSO, PKP and FLACSO Brasil

Bartling, S., Friesike, S. (2014): Opening Science. The evolving guide on how the internet is changing research, collaboration and scholary publishing, Springer Open

Bornmann, L. (2014): Measuring the broader impact of research: The potential of altmetrics, publsished by researchgate

Buschmann, Katrin, Stefan Kasberger, Peter Kraker, Katja Mayer, Falk Reckling, Katharina Rieck, Michela Vignoli (2015): *Open Science in Österreich: Ansätze und Status* (Open Science in Austria: Approaches and status), in Information. Wissenschaft & Praxis 2015; 66(2–3): 1–9. DOI 10.1515/iwp-2015-0025

Delfanti, A., Pitrelli, N. (2015): *Open science: revolution or continuity?*, In: Open Science, Open Issues, Series UC Davis Previously Published Works

EC (2015): Indicators for promoting and monitoring Responsible Research and Innovation. Report from the Expert Group on Policy Indicators for Responsible Research and Innovation, EUR 26866 EN

EC – socientize: Green Paper on Citizen Science. Citizen Science for Europe Towards a better society of empowered citizens and enhanced research

Finnish Ministry of Education and Culture (2014): Open science and research leads to surprising discoveries and creative insights. Open science and research roadmap 2014-2017. Reports of the Ministry of Education and Culture, Finland 2014:21

Galloway, Linda M. and Pease, Janet L., "Altmetrics for the Information Professional: A Primer" (2013). Libraries' and Librarians‘ Publications. Paper 105.

Mietchen D, Mounce R, Penev L (2015) Publishing the research process. Research Ideas and Outcomes 1: e7547. doi: 10.3897/rio.1.e7547

Mietchen, D. (2014): The Transformative Nature of Transparency in Research Funding, Published: December 30, DOI: 10.1371/journal.pbio.1002027

Mingers, John and Loet Leydesdorff (2015): A Review of Theory and Practice in Scientometrics, European Journal of Operational Research

Nicol A., Caruso, J, Archambault, E. (2013): Open Data Access Policies and Strategies in the European Research Area and Beyond, produced for the EC DG Research and Innovation by Science -Metrix Inc.

OECD (2015): OECD Digital Economy Outlook 2015, OECD Publishing, Paris. DOI: dx.doi.org/10.1787/9789264232440-en

OECD (2015): “Making Open Science a Reality”, *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en

Open science and research leads to surprising discoveries and creative insights. Open science and research roadmap 2014-2017. Reports of the Ministry of Education and Culture, Finland 2014:21

Open Access perspectives in the humanities and social sciences. A collection compiled and edited by the LSE Impact of Social Sciences blog, #HSSOA, presented by SAGE and LSE.

Recommendations for the Transition to Open Access in Austria Date (2015), Working Group “National Strategy” of the Open Access Network Austria (OANA).

Scheliga, K., Friesike, S. (2014): Putting Open Science into practice: A social dilemma?, http://firstmonday.org/ojs/index.php/fm/article/view/5381/4110

Sledge, L., Fishman, T.D.: Reimagining higher education. How colleges, universities, businesses, and governments can prepare for a new age of lifelong learning, A GovLab report, Deloitte University Press

9.4 Initial set of indicators used in the online assessment

NB: since the indicators proposed below differed considerably in terms of their scope, granularity, specificity, and measurability, it was stated in the introductory part of the assessment that they should be considered a starting point for further development rather than a mature set of indicators

A) Research process: Conceptualisation & data gathering/creation

Indicators	Mean value
# of research organisations that do open consultations to decide which topics to pursue in their research	
# of proposals applying for funding of OS infrastructure creation and use	
# of shared laboratories (online)	
early involvement of citizens	
% of funding programmes supporting the promotion of data-intensive research	
% of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards	
accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects	
% of machine-readable data / metadata	
quality of metadata (versioning, volume, data format, description of fields, etc.)	
# of sharing policies in research organisations (sharing of data, organisms, etc.)	
% of harmonised sharing policies	
availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research)	
% of funded projects incorporating costs for data compilation/publication and maintenance (of the repository/data sets)	
observing the merging or natural (re-)formation of science disciplines	
access to simulation results	
usability of simulation results (models, data, and code)	
(types of) open data services offered	
is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no))	

B) Research process: Analysis

Indicators	Mean value
# of data citations	
# of code citations (software code)	
use of digital visualisation services (static or interactive online visualisations)	
actual use of open methods	

C) Research process: Diffusion of results

Indicators	Mean value
% of published works using researcher IDs	
# of research organisations where OS is strategically anchored (e.g. in guidelines, strategic documents, target agreements)	

# of agreed policies, principles, or contracts of openness (national, EU-level)	
# of directives from the European Commission for openness	
ratification of those directives by EU member states / adoption by research organisations	
% of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)	
instruments for evaluation of the status of OS	
# of efforts to make open data that are most relevant for the public interest	
% of publications with free licencing (public domain, attribution, all kinds of sharing)	
# of researchers NOT publishing in journals	
# of PhD theses using OS OR on OS-relevant topics	
# (and quality) of publications on OS (as a subject)	

D) Research process: Review and evaluation

Indicators	Mean value
% of decrease of bad science / fraud	
% of researchers perceiving a research career as attractive	
% of review of results from society perspective (social relevance)	

E) System level: Reputation system, recognition of contributions, trust

Indicators	Mean value
data communication recognised as criterion for career progression	
credibility of science in the opinion of the public	
formal recognition of a variety of contributions along the scientific process (e.g. to the selection of research topics, formulation of hypotheses, project participations, review activities); vs. publish and perish	
# of pilot initiatives for new reward systems	
decrease of # of researchers who have negative attitude towards failure (negative results of research efforts)	
% of research funders acknowledging the value of failures (in stimulating actual innovative ideas/approaches)	

F) System level: OS skills & awareness

Indicators	Mean value
% of research personnel / research disciplines skilled in OS	

% of research personnel active in OS	
% of curricula that include OS skills (also prior to higher education)	
# of curricula for data science or other new roles	
# of graduates in data science per year	
% of research personnel aware of standards (is there a standard (relevant to open science), how to we adhere to it, etc.)	
% of research personnel familiar with those standards	
# of researchers having signed an open science pledge	
# of research organisations having signed an open science pledge	
# of mandates and assigned roles (catalysts/evangelists)	

G) System level: Science with society

Indicators	Mean value
# of research projects using crowd funding	
openness in call for proposals (open proposals, open submissions, open review)	
# of research organisations involving citizens in ethical matters	
# of initiatives/training programmes for citizens to engage in science/research	
increase in % of citizens with science literacy	
increase in % of citizens engaging in open science	
circulating and communicating research results outside the academia is standard (yes/no)	
% of researcher who acknowledge Citizen Science as valid form of research	
provision of affordable sets of public interest data / metadata	
advancement in closing the gap between the information rich and the information poor	
Decrease of "emotional gap" between science and society	
% of non-academia (citizens, civil society organisations) represented in advisory boards for research projects/programme	