OPEN SCIENCE MONITOR

STUDY ON OPEN SCIENCE: MONITORING TRENDS AND DRIVERS
Reference: PP-05622-2017

D.2.4 Final report

13th of December 2019

Consortium partners:	The Lisbon Council ESADE Business School Centre for Science and Technology Studies (CWTS) at Leiden University
Subcontractor:	Elsevier

Table of Contents

List of Figures

List of Tables

1 Introduction

Open science has emerged as a powerful trend in research policy. To be clear, openness has always been a core value of science, but it meant publishing the results or research in a journal article. Today, there is consensus that, by ensuring the widest possible access and reuse to publications, data, code and other intermediate outputs, scientific productivity grows, scientific misconduct becomes rarer, discoveries are accelerated. Yet it is also clear that progress towards open science is slow, because it has to fit in a system that provides appropriate incentives to all parties.

The European Commission has recognized this challenge and moved forward with strong initiatives from the initial 2012 recommendation on scientific information (C (2012) 4890), such as the Open Science Policy Platform and the European Open Science Cloud.[1] Open access and open data are now the default option for grantees of H2020.

The Open Science Monitor (OSM) aims to provide data and insight needed to support the implementation of these policies. It gathers the best available evidence on the evolution of open science, its drivers and impacts, drawing on multiple indicators as well as on a rich set of case studies. [2]

This monitoring exercise is challenging. Open science is a fast evolving, multidimensional phenomenon. According to the OECD (2015), "open science encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives". This very definition confirms the relative fuzziness of the concept and the need for a clear definition of the "trends" that compose open science.

Precisely because of the fast evolution and novelty of these trends, in many cases it is not possible to find consolidated, widely recognized indicators. For more established trends, such as open access to publications, robust indicators are available through bibliometric analysis. For most others, such as open code and open hardware, there are no standardised metrics or data gathering techniques and there is the need to identify the best available indicator that allows one to capture the evolution and show the importance of the trend.

Today, especially at European level where competences on research are limited, data and indicator play a powerful role in executing policies. Conversely, the absence of robust data can hinder the implementation of the policy.

This is precisely the objective of the Open Science Monitor.

1.1 Objectives

The OSM covers four tasks:

1. To provide metrics on the open science trends and their development.
2. To assess the drivers (and barriers) to open science adoption.

[1] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform; https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

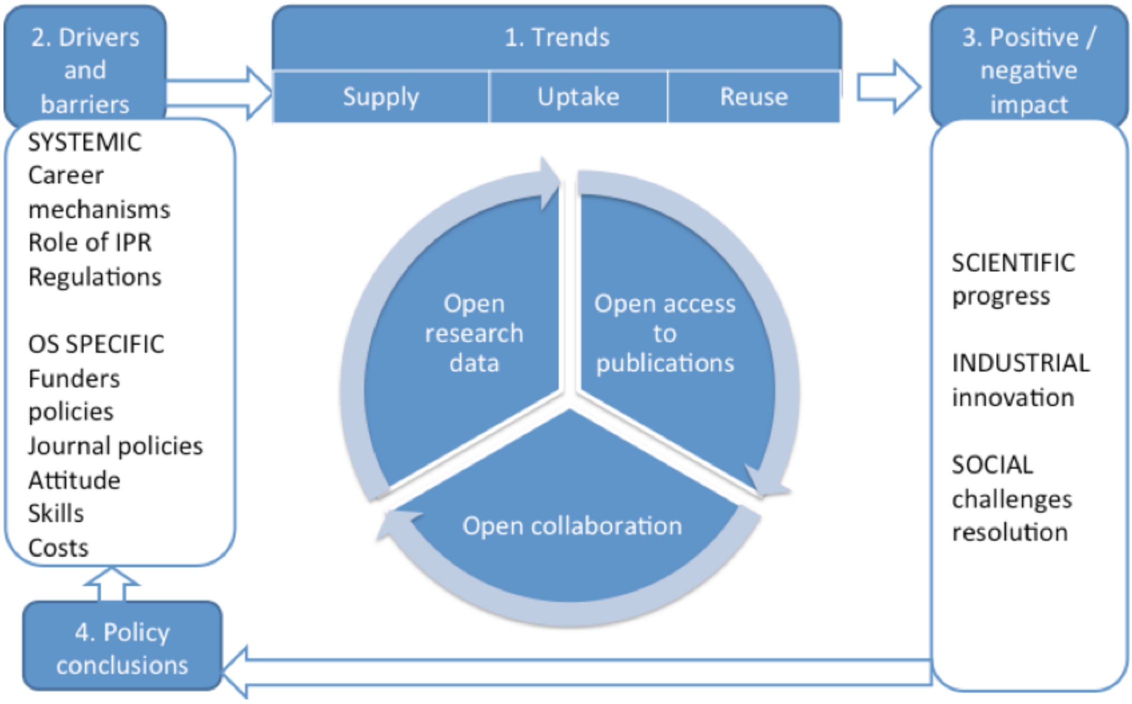
[2] The OSM has been published in 2017 as a pilot and re-launched by the European Commission in 2018 through a contract with a consortium composed by the Lisbon Council, ESADE Business School and CWTS of Leiden University (plus Elsevier as subcontractor). See https://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor

3. To identify the impacts (both positive and negative) of open science.
4. To support evidence-based policy actions.

The indicators presented here focus mainly on the first two tasks: mapping the trends, and understanding the drivers (and barriers) for open science implementation.

The chart below provides an overview of the underlying conceptual model.

Figure 1: A conceptual model: an intervention logic approach



The central aspect of the model refers to the analysis of the open science trends and is articulated alongside three dimensions: *supply*, *uptake* and *reuse* of scientific outputs.

In the OSM framework, *supply* refers to the emergence of services such as data repositories. The number of data repositories (one of the existing indicators) is a *supply* indicator of the development of open science. On the demand side, indicators include, for example, the amount of data stored in the repositories, the percentage of scientists sharing data. Finally, because of the nature of open science, the analysis will go beyond usage, since the reuse dimension is particularly important. In this case, relevant indicators include the number of scientists reusing data published by other scientists, or the number of papers using these data.

On the left side of the chart, the model identifies the key factors influencing the trends, both positively and negatively (i.e. *drivers* and *barriers*). Both drivers and barriers are particularly relevant for policy-makers as this is the area where an action can make greatest difference, and are therefore strongly related to policy recommendations. These include "policy drivers", such as funders' mandates. It is important to assess not only policy drivers dedicated to open science, but also more general policy drivers that could have an impact on the uptake of open science. For instance, the increasing reliance on performance-based funding or the emphasis on market exploitation of research are general policy drivers that could actually slow down the uptake of open science.

The right side of the chart in the model illustrates the *impacts* of open science to research or the scientific process itself; to industry or the capacity to translate research into marketable products and services; to society or the capacity to address societal challenges.

1.2 Scope

By definition, open science concerns the **entire cycle** of the scientific process, not only open access to publications. Hence the macro-trends covered by the study include: open access to publications, open research data and open collaboration. While the first two are self-explanatory, open scientific collaboration is an umbrella concept to include forms of collaboration in the course of the scientific process that do not fit under open data and open publications.

Table 1: Articulation of the trends to be monitored

Categories	Trends
Open access to publications	• Open access policies (funders and journals), • Green and gold open access adoption (bibliometrics).[3]
Open research data	• Open data policies (funders and journals) • Open data repositories • Open data adoption and researchers' attitudes.
Open collaboration	• Open code, • Altmetrics, • Open hardware, • Citizen science.

New trends within the open science framework will be identified through interaction with the stakeholder's community by monitoring discussion groups, associations (such as Research Data Alliance - RDA), mailing lists, and conferences such as those organised by Force11 (www.force11.org).

The study covers **all research disciplines**, and aims to identify the differences in open science adoption and dynamics between diverse disciplines. Current evidence shows diversity in open science practices in different research fields, particularly in data-intensive research domains (e.g. life sciences) compared to others (e.g. humanities)

The **geographic coverage** of the study is 28 Member States (MS) and G8 countries, including the main international partners, with different degrees of granularity for the different variables. As far as possible, data has to be presented at **country level**.

[3] According to the EC, "'Gold open access' means that open access is provided immediately via the publisher when an article is published, i.e. where it is published in open access journals or in 'hybrid' journals combining subscription access and open access to individual articles. In gold open access, the payment of publication costs ('article processing charges') is shifted from readers' subscriptions to (generally one-off) payments by the author.[…] 'Green. open access' means that the published article or the final peer-reviewed manuscript is archived by the researcher (or a representative) in an online repository." (Source: H2020 Model Grant Agreement)

1.3 Methodological approach

The Open Science Monitor gathers data from a plurality of sources. While the details are fully explained in the methodological note, it is worth summarizing the main aspects here.

The study produces key indicators and metrics related to the above-mentioned open science trends. These indicators are generated using multiple data sources which can be categorized in three main types. First, an analysis of existing data and metrics, for instance as provided by the Sherpa or Re3data database. Indicators here are produced by analysing the underlying data provided by these services. Secondly, the study generates its own metrics, namely on open access to publications, by analysing publications based on the Scopus and Unpaywall data. Thirdly, the study included a survey of about a thousand researchers worldwide, performed by Elsevier.

The study also provides an analysis of policies, drivers, barriers and impacts. In this case, the evidence is provided mostly by 28 case studies and the related meta-analysis.

Table 2: overview of the methodology

	Existing metrics	Own metrics	Survey	Cases
Trends	X	X	X	
Publication	*X*	*X*		
Data	*X*		*X*	
Collaboration	*X*			
Drivers			X	X
Barriers			X	X
Policies	X			X
Impact			X	X

2 Overview of trends and indicators

This section presents an overview of updated indicators of Open Science Monitor. The initial overview of trends is based on the intervention logic of the structure of the methodology proposed in the inception report:

- Open access to publications,
- Open research data,
- Open collaboration.

All the data included in this section are taken from the online dashboard of the Open Science Monitor, available at: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en . Full details on the methodology are available on the website and in annex to this report.

2.1 Open access to publications

The indicators of *Open access to publications* are measuring to what extent it is possible to freely access research publications. The indicators cover bibliometric data on publications, as well as data on funders' and journals' policies.

Concerning the **availability of open access publications**, the total number of open access publications (gold, green, hybrid and bronze) had been growing between 2009 and 2018 (the amount of open access publications has almost doubled in this period from 361 thousand to over 684 thousand), though, a significant slowdown in growth since 2016 was registered (8% growth as compared to growth ranging between 12% to 16% in the previous years). Data for 2017 and 2018 show a decline, however in our consideration this is simply due to the embargo period for green open access so that they will become available over the next months – as shown by the fact that gold open access, where embargo does not apply, is continuously growing.

Figure 2: Percentage of open access publications

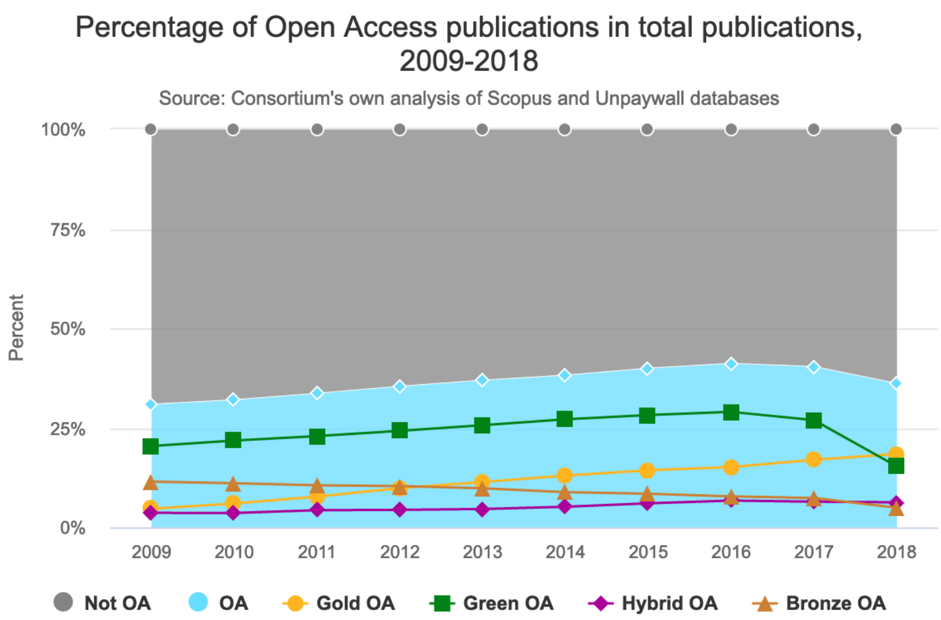


Table 3: Open access publications (gold, green, hybrid and bronze) by year

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Not OA	807672	824427	876706	917504	971428	1010275	1017187	1044540	1086564	1208464
Not OA % of change		**2.1%**	**6.3%**	**4.7%**	**5.9%**	**4.0%**	**0.7%**	**2.7%**	**4.0%**	**11.2%**
Open Access	**361932**	**389100**	**445341**	**503875**	**569939**	**625503**	**677841**	**731812**	**736656**	**684636**
OA % change		**7.5%**	**14.5%**	**13.1%**	**13.1%**	**9.7%**	**8.4%**	**8.0%**	**0.7%**	**-7.1%**
Gold OA	54910	72608	101740	140147	175309	212999	242785	269360	309593	349589
% change		**32.2%**	**40.1%**	**37.8%**	**25.1%**	**21.5%**	**14.0%**	**10.9%**	**14.9%**	**12.9%**
Green OA	239013	264567	303047	344491	395195	444121	478279	514907	490630	289461
% change		**10.7%**	**14.5%**	**13.7%**	**14.7%**	**12.4%**	**7.7%**	**7.7%**	**-4.7%**	**-41.0%**
Hybrid OA	42571	43850	57033	62363	70992	84624	102384	119348	117541	119019
% change		**3.0%**	**30.1%**	**9.3%**	**13.8%**	**19.2%**	**21.0%**	**16.6%**	**-1.5%**	**1.3%**
Bronze OA	134029	134547	139945	147194	151145	144268	143379	138292	134500	93216
% change		**0.4%**	**4.0%**	**5.2%**	**2.7%**	**-4.5%**	**-0.6%**	**-3.5%**	**-2.7%**	**-30.7%**

The closer look into the data reveals that the growth was mainly driven by the rapid growth of the number and share of open access gold publications. In 2018, the open access gold publications outnumbered for the first-time open access green publications (green – 289 thousand to gold – 349 thousand).

At the country level, top five countries with the biggest share of open access publications (total of gold, green, hybrid and bronze) are: United Kingdom (52.3%), Switzerland (51.8%), Croatia (50.8%), Luxembourg (50%) and Netherlands (49.9%). On the opposite spectrum bottom five countries with the lowest share of open access publications are: Russian Federation (23.9%), China (27.8%), India (29.9%), Greece (34.7%), Canada (37.1%).

However, if we look at the absolute numbers of open access publications, the situation changes significantly with US clearly leading the way to open access (27.6% of all open access publications), followed by China (10%), UK (9.9%), Germany (6.6%) and Japan (5.4%). Together open access publications of these five countries represent nearly 60% of all open access publications.

Figure 3: Percentage of open access publications by country.

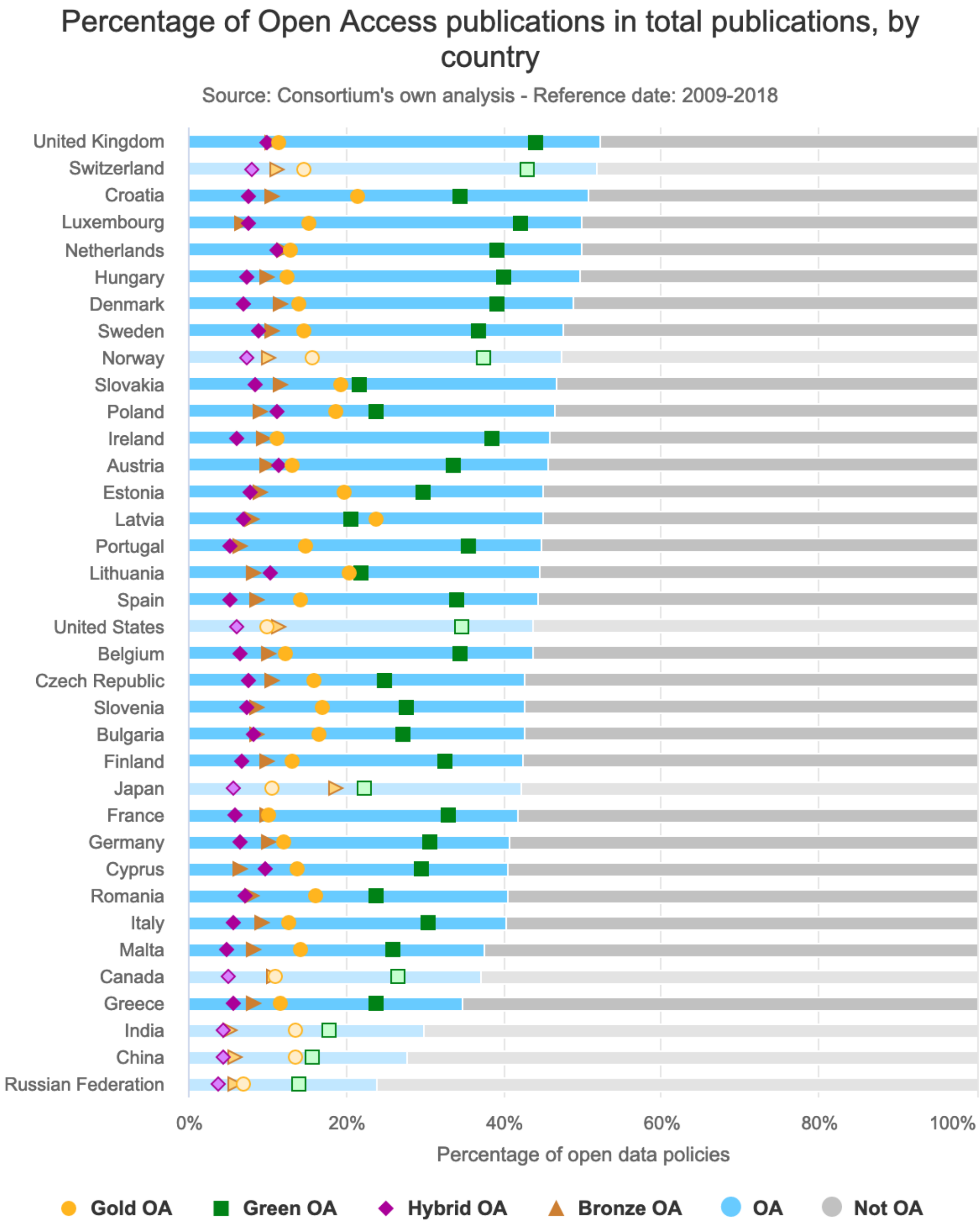


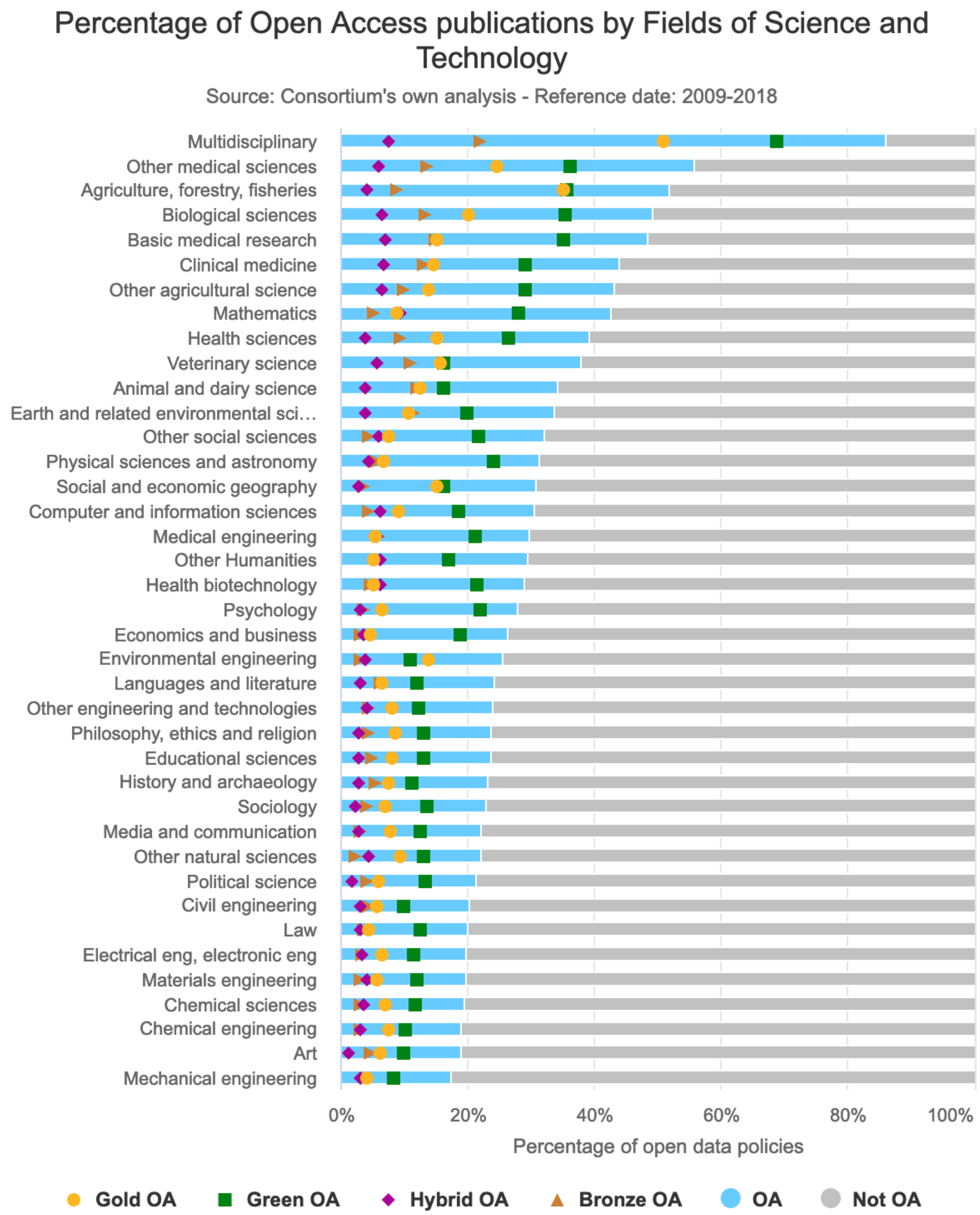
Table 4: Open access publications (gold, green, hybrid and bronze) by country

Country	Not OA	% total publications	Total OA	% total publications	Gold OA	% total publications	Green OA	% total publications	Hybrid OA	% total publications	Bronze OA	% total publications
United States	2346486	56.3%	1821541	43.7%	405057	9.7%	1442292	34.6%	246303	5.9%	467031	11.2%
China	1708392	72.2%	657838	27.8%	316526	13.4%	366048	15.5%	98571	4.2%	134676	5.7%
United Kingdom	598347	47.7%	656613	52.3%	141665	11.3%	551290	43.9%	122050	9.7%	131031	10.4%
Germany	638049	59.3%	438618	40.7%	127508	11.8%	329374	30.6%	68385	6.4%	108179	10.0%
Japan	491666	57.8%	358961	42.2%	88791	10.4%	188554	22.2%	46637	5.5%	157998	18.6%
France	443312	58.2%	318791	41.8%	76929	10.1%	250206	32.8%	44432	5.8%	73995	9.7%
Canada	432232	62.9%	254432	37.1%	75349	11.0%	181893	26.5%	33374	4.9%	72921	10.6%
India	469394	70.1%	200609	29.9%	89727	13.4%	119116	17.8%	28209	4.2%	34971	5.2%
Italy	386379	59.6%	261600	40.4%	81818	12.6%	196304	30.3%	35523	5.5%	59025	9.1%
Spain	322038	55.7%	256071	44.3%	81736	14.1%	195526	33.8%	30110	5.2%	48826	8.4%
Russian Federation	304400	76.1%	95631	23.9%	27301	6.8%	55391	13.8%	14888	3.7%	23102	5.8%
Netherlands	197547	50.1%	197091	49.9%	50399	12.8%	153865	39.0%	43976	11.1%	46967	11.9%
Switzerland	141623	48.2%	152342	51.8%	42377	14.4%	125883	42.8%	23402	8.0%	32597	11.1%
Sweden	139000	52.5%	125797	47.5%	38628	14.6%	96852	36.6%	22947	8.7%	27843	10.5%
Poland	127287	53.4%	110866	46.6%	44012	18.5%	56340	23.7%	26540	11.1%	21301	8.9%
Belgium	120476	56.4%	93231	43.6%	26070	12.2%	73279	34.3%	13593	6.4%	21457	10.0%
Denmark	87605	51.2%	83522	48.8%	23658	13.8%	66681	39.0%	11670	6.8%	19560	11.4%
Austria	81960	54.3%	68938	45.7%	19528	12.9%	50633	33.6%	16895	11.2%	14714	9.8%
Portugal	75802	55.2%	61490	44.8%	20317	14.8%	48572	35.4%	7126	5.2%	8922	6.5%
Norway	70452	52.7%	63296	47.3%	20851	15.6%	49821	37.2%	9766	7.3%	13292	9.9%
Finland	73935	57.6%	54413	42.4%	16564	12.9%	41603	32.4%	8408	6.6%	12576	9.8%
Greece	76546	65.3%	40684	34.7%	13529	11.5%	27864	23.8%	6592	5.6%	9497	8.1%
Czech Republic	63721	57.3%	47531	42.7%	17663	15.9%	27529	24.7%	8222	7.4%	11538	10.4%
Ireland	47793	54.1%	40622	45.9%	9732	11.0%	33934	38.4%	5198	5.9%	8292	9.4%
Hungary	35158	50.2%	34848	49.8%	8595	12.3%	27879	39.8%	5139	7.3%	6882	9.8%
Romania	29898	59.6%	20291	40.4%	7975	15.9%	11850	23.6%	3540	7.1%	3998	8.0%
Slovenia	20434	57.3%	15199	42.7%	6003	16.8%	9828	27.6%	2611	7.3%	3012	8.5%
Slovakia	18042	53.3%	15811	46.7%	6474	19.1%	7322	21.6%	2812	8.3%	3872	11.4%
Croatia	15835	49.2%	16350	50.8%	6886	21.4%	11039	34.3%	2390	7.4%	3377	10.5%
Bulgaria	12104	57.4%	8979	42.6%	3453	16.4%	5688	27.0%	1703	8.1%	1820	8.6%
Lithuania	10982	55.3%	8860	44.7%	4001	20.2%	4319	21.8%	2019	10.2%	1627	8.2%
Estonia	10049	55.0%	8235	45.0%	3575	19.6%	5407	29.6%	1393	7.6%	1630	8.9%
Cyprus	7455	59.4%	5086	40.6%	1708	13.6%	3681	29.4%	1190	9.5%	803	6.4%
Luxembourg	4905	50.0%	4896	50.0%	1477	15.1%	4117	42.0%	722	7.4%	638	6.5%
Latvia	4330	55.0%	3544	45.0%	1868	23.7%	1607	20.4%	542	6.9%	620	7.9%
Malta	2062	62.4%	1243	37.6%	465	14.1%	854	25.8%	156	4.7%	271	8.2%

In terms of the science and technology field, *Multidisciplinary* field has the biggest share of open data publications – 86.2%, followed by *Other medical sciences* 55.7%, *Agriculture, forestry, fisheries* (51.9%), and life sciences: *Biological sciences* (49.2%), *Basic medical research* (48.5%) and *Clinical medicine* (44%).

The most closed fields with the lowest number of open access publications are: *Mechanical engineering* (17.3%), *Art* (19%), *Chemical engineering* (19%), *Chemical sciences* (19.4%) and *Materials engineering* (19.8%).

Figure 4: Percentage of open access publications by Field of Science and Technology



It is worth noticing, however, that in terms of absolute numbers the situation is slightly different. *Clinical medicine, Biological sciences, Basic medical research, Physical sciences and astronomy* and *Earth and related environmental sciences* have the highest aggregated number of open access publications, whereas *Other agricultural sciences*, *Art*, *Media and communication*, *Law* and *Political science* are the fields with the lowest number of open access publications.

2.2 Open research data

The Open research data indicators measure to what extent data underpinning scientific research results has no restrictions on its access. The main source for these indicators is the survey run by Elsevier in 2018, but some indicators are developed through existing metrics.

2.2.1 Researchers' behaviour

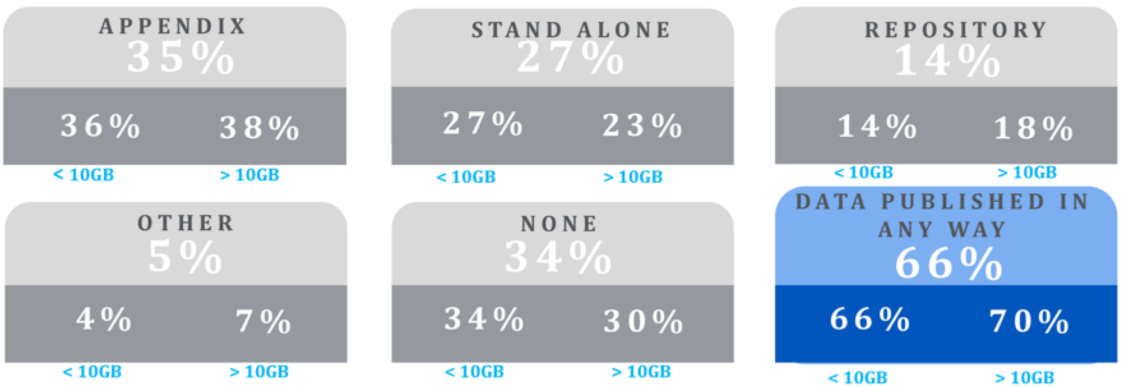
Data are a fundamental input in all research activities. Observational data are the most common data category, used by 48% of researchers, followed by experimental data, used by 35% of researchers. In terms of size, the vast majority of researchers (70%) generates less than 10 GB of data in their last research project.

Researchers share data in different ways. The most commonly used method is as an appendix to an article (35%), while only 14% of researchers share it through data repositories. Overall, a third of researchers declare not to share their data.

This is mostly good news: the vast majority of researchers declare to share data. However, a closer look makes these data less good than they look:

1. This is data sharing as self-reported by researchers. Not only it is prone to the usual limitations of the method compared to observation, but also it does not reflect the article as a unit of analysis. Other sources show that the vast majority of articles do not include research data.
2. Repositories are considered the most suitable way to make data genuinely accessible and reusable for other researchers, but they are yet used only by a small minority of researchers (roughly one in seven).
3. These results are practically identical to the 2016 survey, both in the aggregate and in the individual option, showing that limited progress has been achieved.

Figure 5: Have you published the research data that you used or created as part of your last research project in any of the following ways?

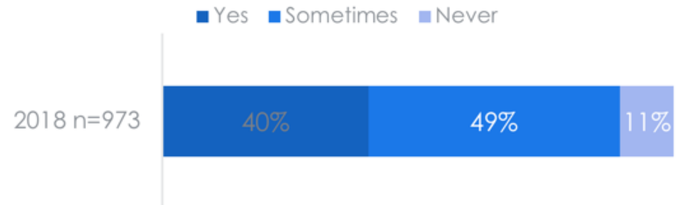


On the other hand, researchers need data sharing: 42% declare that they rely on data from other researchers, and 74% declare they would benefit from access to others' research data.

2.2.2 How data are shared

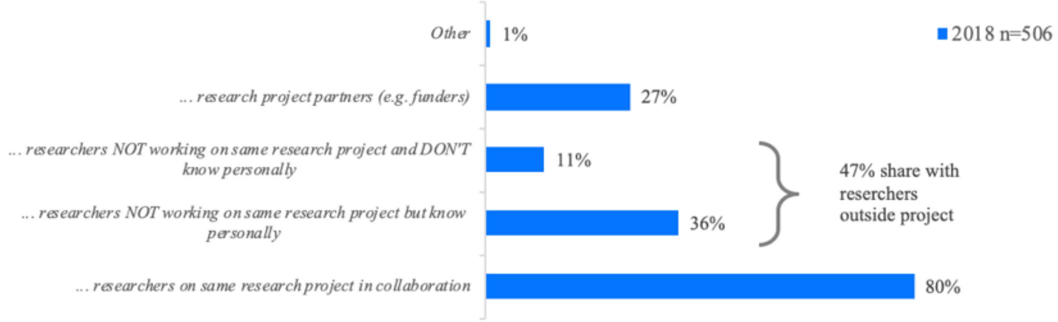
Regarding approaches to research data management, the majority of researchers do take steps to manage their research data for potential future reuse (89%), however, this is not always carried out in a consistent fashion with 49% stating that these steps were taken “sometimes”. Again, there is no progress visible from 2016.

Figure 6: Do you take steps to manage your research data and/or archive it for potential re-use by yourself and/or others?



When focussing on data sharing practices, research data sharing is carried out mostly between collaborators on the same projects (80%). Only 38% of those who share, do it with researchers outside of their own project. This suggests that researcher do not adopt fully open data approaches, but rather discriminatory approach, sharing with selected partners on a case by case basis.

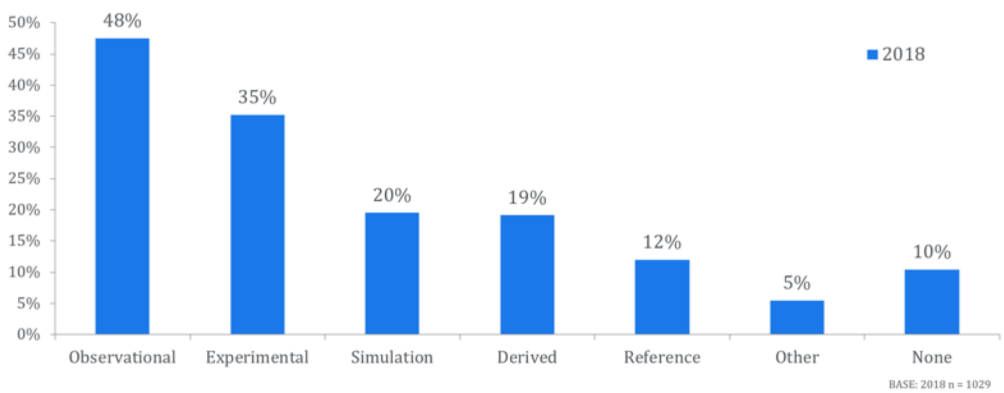
Figure 7: Have you done any of the following with any or all of the research data that you used or created as part of your last research project?



Asked to researchers who have shared data they used/created with others

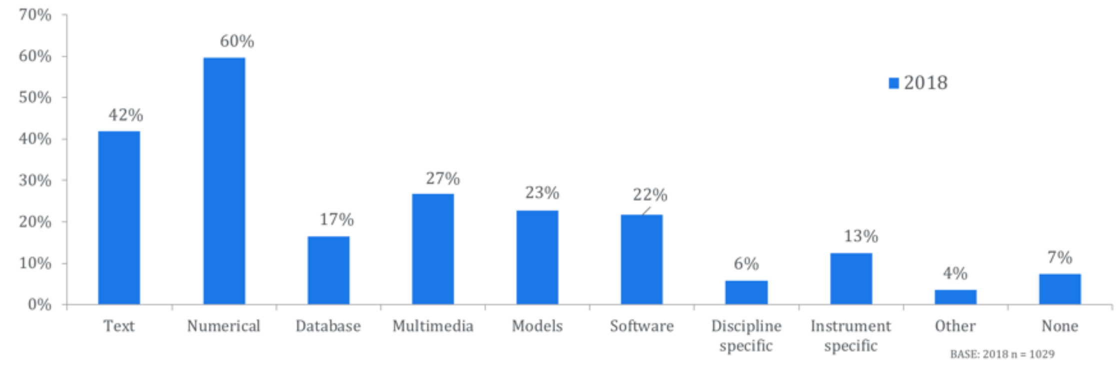
The number of researchers sharing data in their last project has remained stable, with no growth shown over the past two years. Observational and experimental data continue to be the main type of data used by researchers (48% and 35% respectively).

Figure 8: Categories of research data used in your last research project.



When analysing the research data formats used in research projects, numerical and text formats are the most popular (60% and 42% respectively) amongst researchers.

Figure 9: Research data formats used or created as part of your last research project



2.2.3 Data ownership

Regarding research funding, a country/subject specific funder remains the primary source of funding. When observing data ownership, the data was divided into two sections: 1) before publication and 2) after publication. The majority of researchers say they own the data pre-publication (62%).

Figure 10: What was the primary source of funding for your last research project?

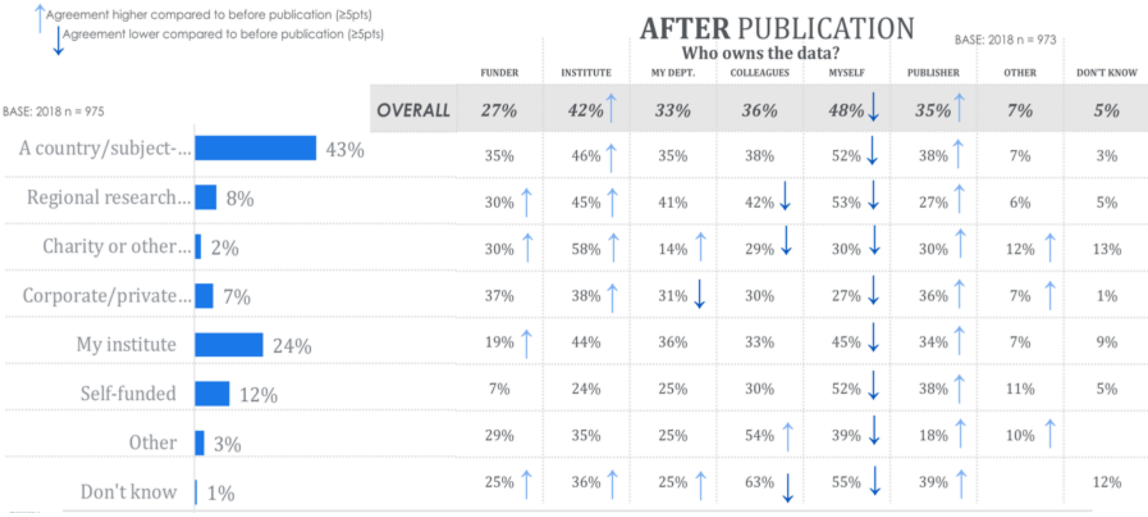
BEFORE PUBLICATION

BASE: 2018 n = 975 — BASE: 2018 n = 973

		Who owns the data?							
		FUNDER	INSTITUTE	MY DEPT.	COLLEAGUES	MYSELF	PUBLISHER	OTHER	DON'T KNOW
	OVERALL	23%	36%	32%	39%	62%	4%	4%	5%
A country/subject-specific funder	43%	32%	40%	31%	42%	68%	5%	3%	4%
Regional research funder	8%	21%	33%	45%	48%	67%	2%	4%	3%
Charity or other 'third sector' funder	2%	22%	51%	9%	41%	39%	6%	7%	13%
Corporate/private/commercial organization	7%	34%	31%	42%	33%	42%	8%	2%	1%
My institute	24%	14%	40%	37%	34%	57%	3%	4%	6%
Self-funded	12%	5%	21%	21%	31%	67%	5%	7%	6%
Other	3%	31%	32%	22%	39%	47%		3%	3%
Don't know	1%		26%	7%	70%	63%			12%

After publication, the perception that the researcher owns the data decreases and there is a strong growth in the perception that the publisher owns the data with an increase from 4% (prior to publication) to 35% (after publication).

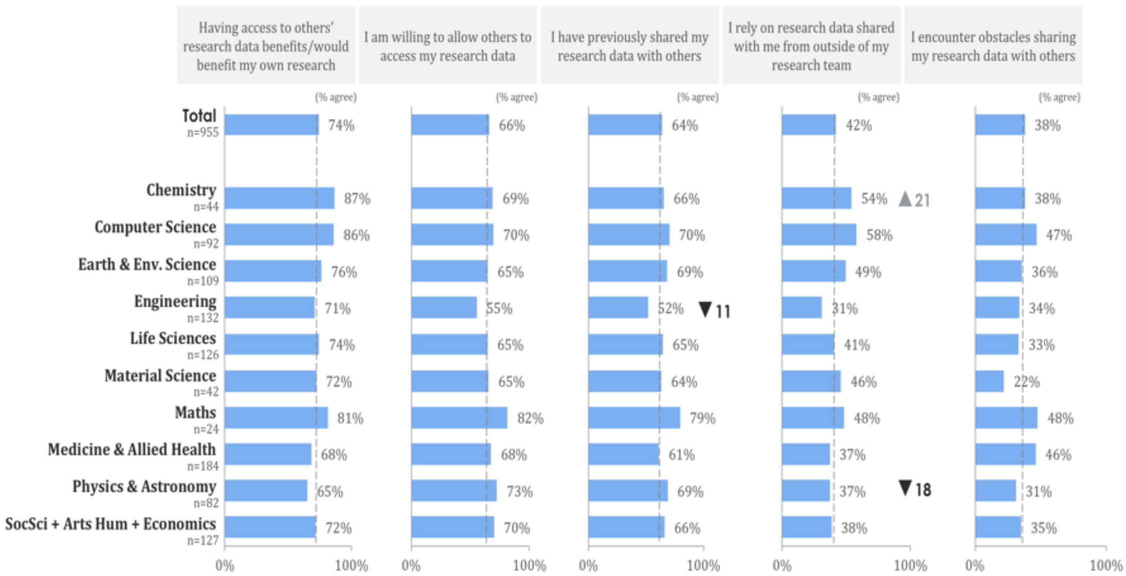
Figure 11: What was the primary source of funding for your last research project?



2.2.4 Attitudes across disciplines

Regarding researcher attitudes to data management and sharing, Chemists (87%) and Computer Scientists (86%) are most likely to benefit from data sharing and have the strongest reliance on research data from outside of their research team (54% and 58% respectively). Mathematicians stand out as the group who are most willing to allow others to access their research data (82%) and who have shared their data with others in the past (79%).

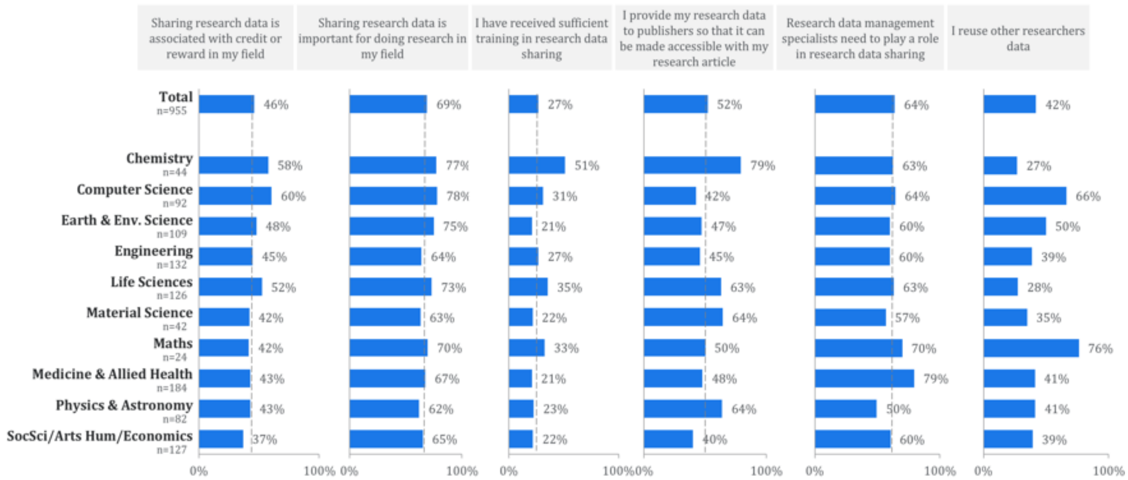
Figure 12: Attitudes to data management by subject (I/II)



Computer Scientists are most likely to be rewarded for sharing (60%) and a strong link between data sharing and reward can also be seen amongst Chemists (58%) and Life Scientists (52%). Meanwhile only 30% of researchers in the field of Social Sciences, Arts and Humanities and Economics stated that sharing research data is associated with credit and reward in their field. In terms of training in research data, Chemists believe they have received the most experience in this area (51%). However, the perception of a general lack of training can be seen across the board in the other subjects. Chemists also stand out as the group that makes their research data available to publishers to make it accessible with their research articles (79%).

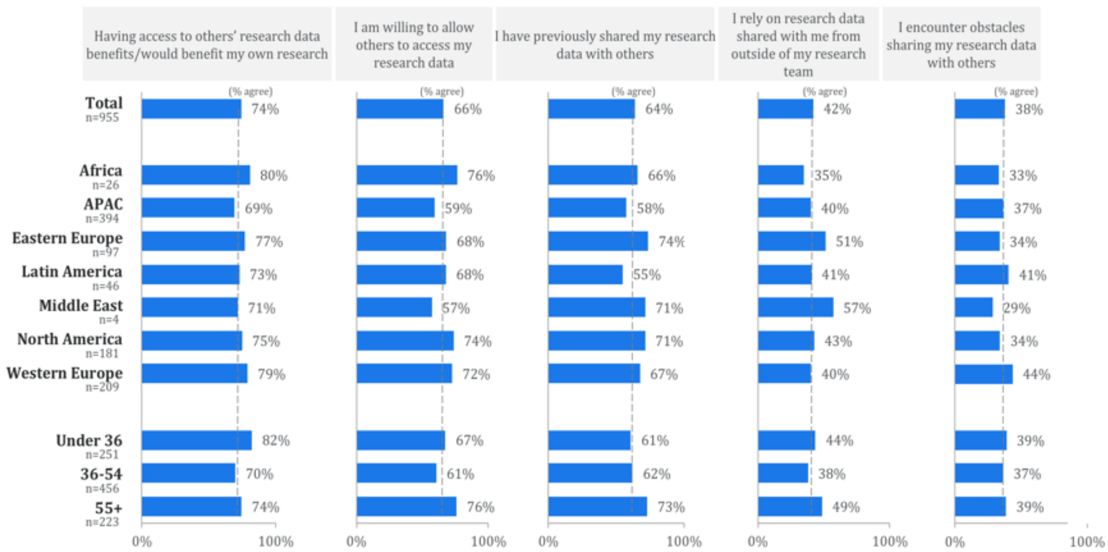
Those researchers in Medicine & Allied Health most believe there is a role for research data management specialists in research data sharing. Finally, both Mathematicians and Computer Scientists are the groups that most reuse other researcher's data (76% and 66% respectively).

Figure 13: Attitudes to data management by subject (II/II)



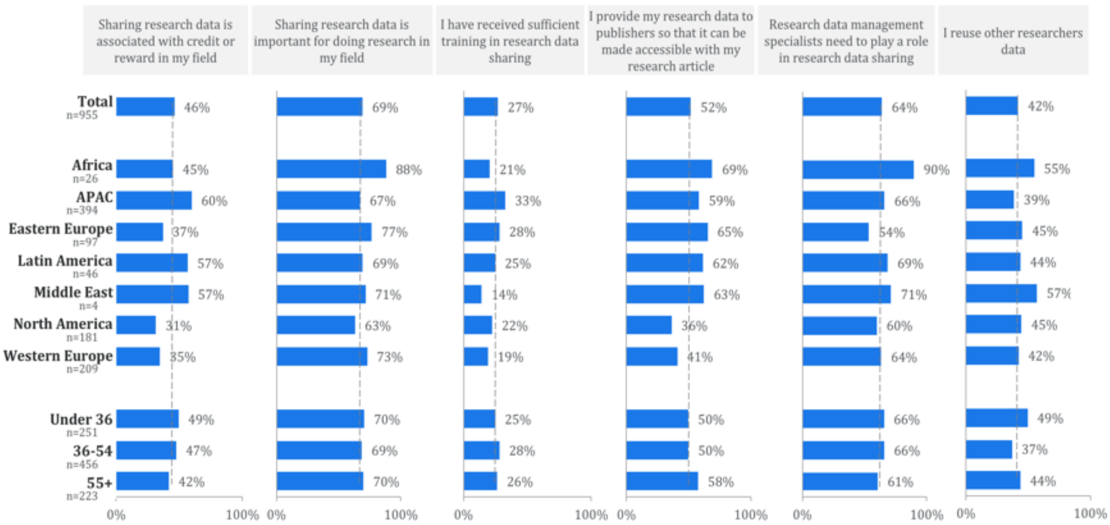
When analysing the attitudes to data management by region and age, it can be seen that those in Asia Pacific (APAC) are least likely to share their data (69%) and those in Western Europe are finding it more difficult to access other researcher's data (44%). Those researchers under the age of 36 are the group that most agree that access to other researcher's data would benefit their own research (82%). However, the strong willingness to allow access to research data by the over 55s should also be highlighted (76%).

Figure 14: Attitudes to data management by region/age (I/II)



In Asia Pacific there is a stronger link between sharing research data and the perceived credit and reward that it brings to researchers (60%). Africa (90%), the Middle East (71%) and Latin America (69%) stand out as those regions where there is a need for research data management specialists in research data sharing.

Figure 15: Attitudes to data management by region/age (II/II)



2.2.5 Data repositories

Regarding the **availability of data repositories**, there were no less than 3449 repositories listed on Re3data.org in October 2019, up from 2986 in 2018, out of which repositories of life science constituted 36%, natural science 33%, humanities and social science 21% and engineering sciences 10%. The vast majority of data repositories provide open access to the data – over 94% of all repositories for which this information is available.

Figure 16: Data repositories by subject

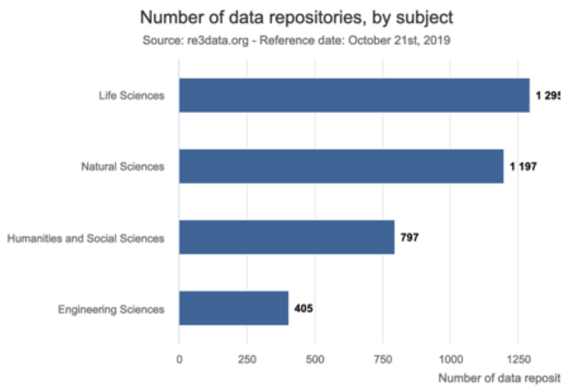
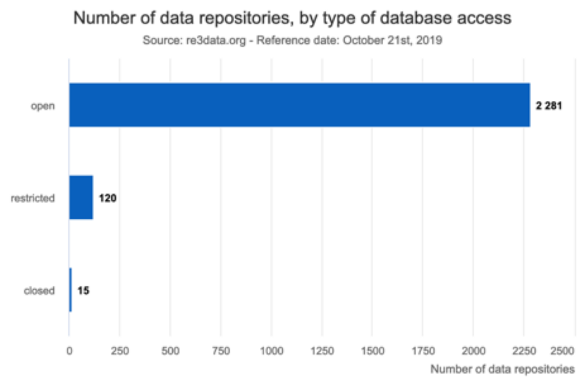


Figure 17: Data repositories by type of access



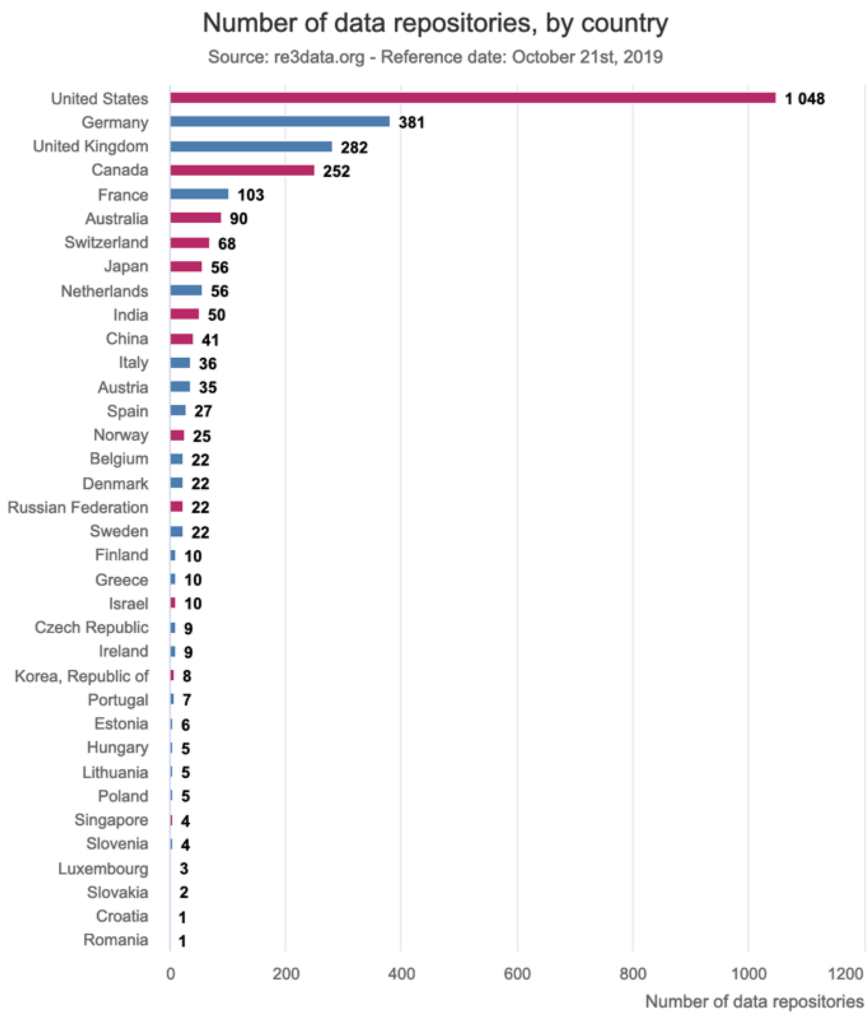
The vast majority of data repositories is located in US – 39,2% (1048 repositories) – almost three times the number of data repositories of Germany – a second country on the list with 14% of data repositories and the EU country with the highest number of data repositories.

Looking at top 10 countries with the highest number of data repositories only four EU countries are prominent: Germany (381 repositories), United Kingdom (282 repositories), France (103 repositories) and Netherlands (56 repositories).

From non-EU countries Canada (252 repositories), Australia (90), Switzerland (68), Japan (56) and India (50) are placed in top 10 countries.

Romania and Croatia are the countries with the lowest number of data repositories.

Figure 18: Data repositories by country, blue bar indicate European countries

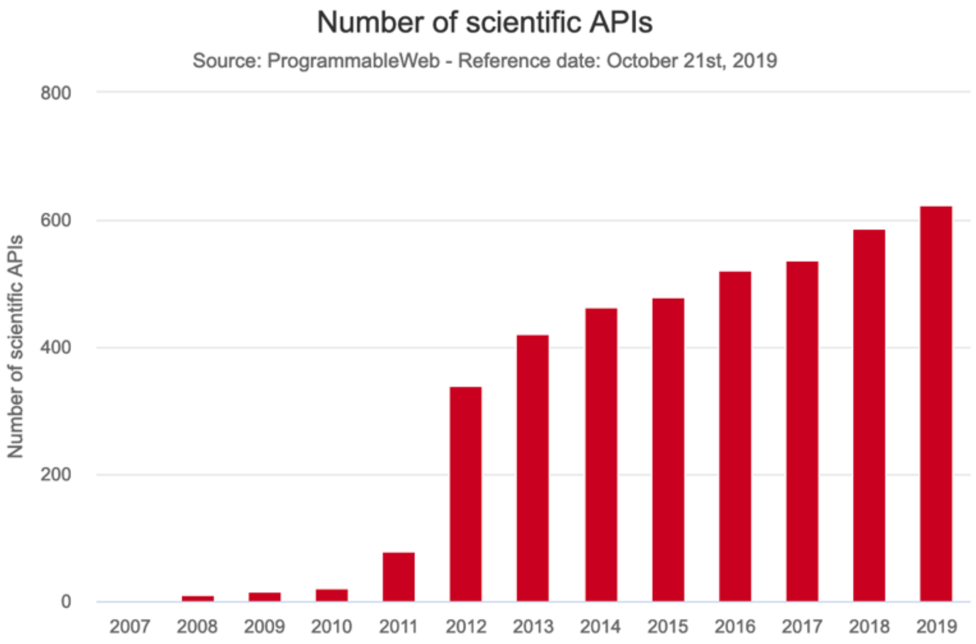


2.3 Open collaboration

Open scientific collaboration refers to the forms of collaboration in the course of the scientific process that go beyond open data and open publications. Measuring open scientific collaboration includes measuring of different type of outputs such as open code, open hardware, the use of collaborative platforms between scientists and the "citizen-science" phenomenon.

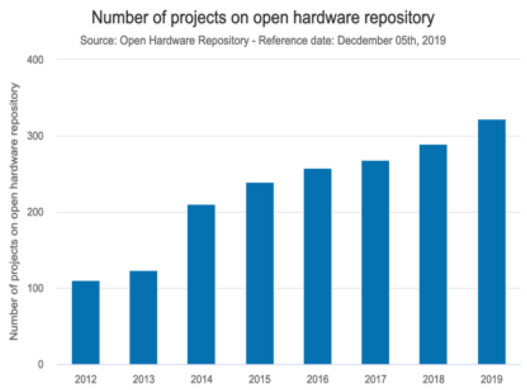
Regarding the **availability of scientific Application Programming Interfaces (APIs**), the number of scientific APIs is growing over the course of last ten years. In 2012 there was a sudden increase in a number of APIs (it´s numbered quadrupled) and since then a number of APIs grew steadily to 625 scientific APIs in 2019. While APIs are not new, they are showing continuous growth and becoming commonplace also in the scientific context.

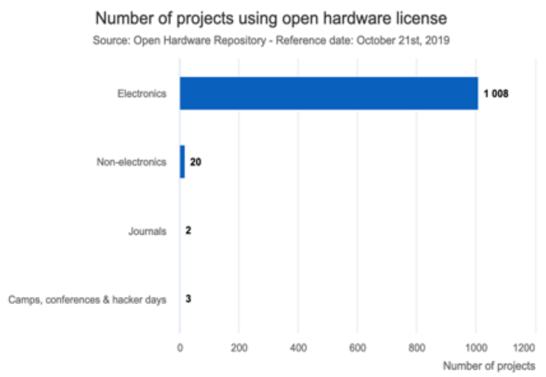
Figure 19: Number of scientific APIs



Just as the number APIs, number of **projects on Open Hardware Repository** is growing steadily, amounting to 323 projects in 2019.

Figure 20 & Figure 21: Open hardware projects





Another growing phenomenon in the realm of open collaboration is **open citizen-science**. Open Science Monitor is using data from to science crowdsourcing platforms SciStarter and Zooniverse. By October 2019 the SciStarter recorded 4506 projects, up from 4095 in 2018, out of which the majority referred to life sciences and environment. Similarly, out of 217 projects recorded by Zooniverse till October 2019, more than half of them concerned life sciences and environment.

Figure 22: Projects in Scistarter by discipline

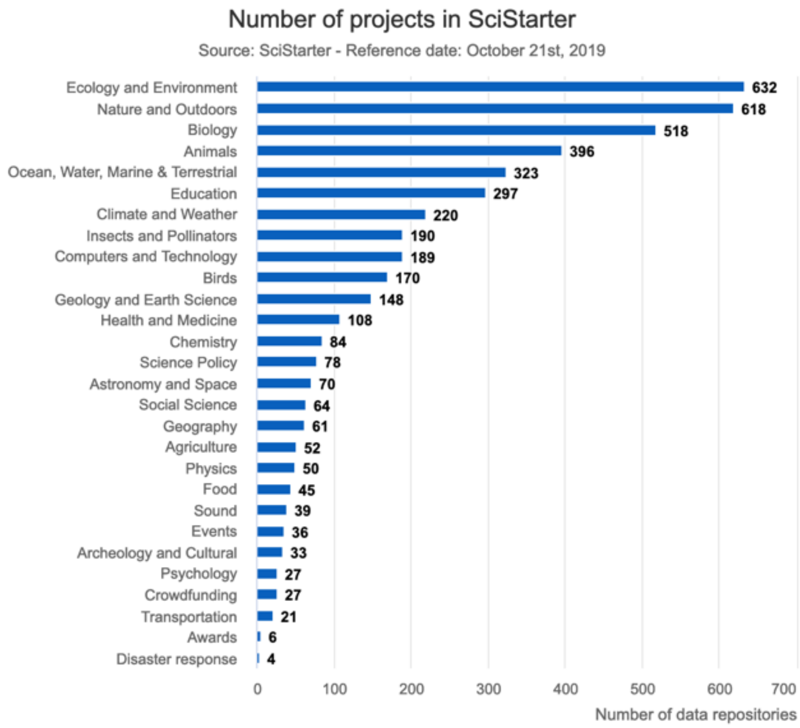
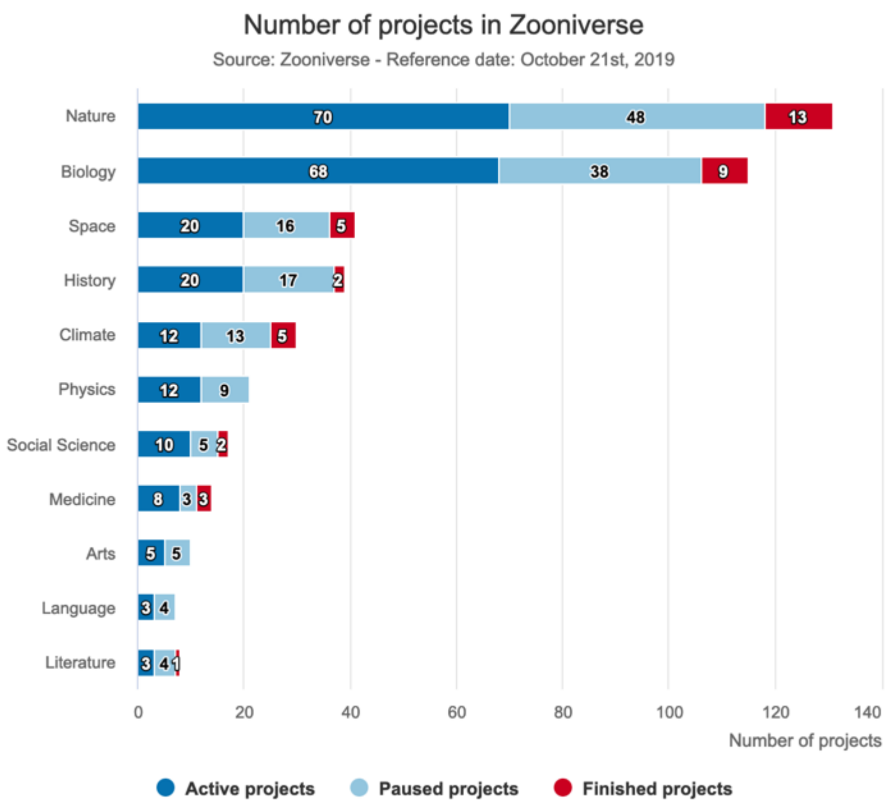


Figure 23: projects in Zooniverse by discipline



3 Drivers and barriers

3.1 Drivers

There are many different drivers at play, which can facilitate or hinder the adoption of open science. Their identification is crucial to the development of appropriate policy recommendations, which are typically built around these "levers": they aim to enhance the drivers to accelerate adoption.

While there are general trends, external to science, that are important because they reinforce or damage open science, in the framework of the 28 case studies performed, the research team attempted to identify the specific drivers or enabling factors that were behind the open science initiatives selected.

Some of the factors most commonly mentioned in prior research and strictly related to open science (non-including general trends affecting open science adoption) are summarised in Table 5.

Across the analysis of cases we have identified nine main drivers that have been clustered in three overarching categories:

1. ***Supporter***: Refers to factors that describe the support of certain stakeholders in the ecosystem to the initiative. Such support facilitates and encourages the adoption of open science behaviours by the research community through the cases. Under these categories, the factors include: Institutional support (i.e. when an institution is behind the initiative); publishers support (i.e. when a publisher or publishers are supporting in some form the initiative); funders support (i.e. the initiative has the support of a funding agency); industry support (i.e. when the initiative has private sponsors); and community support (i.e. the case is supported by the scientific community, citizens or final users in general).

2. ***Intention***: Includes factors where different design and implementation measures were taken into account, that is: whether the initiative was easily to plug in already existing measures, processes and practices (i.e. commodity); and whether it implies major or minor changes in the existing workflows (i.e. changes).

3. ***Focus***: The third category refers to how the initiative targeted the different collectives and includes: whether the initiative was designed around user needs, understanding their practices patterns and common behaviour (i.e. user-centricity); or whether the initiative involved suppliers capable of provide the product or service and support adoption or was designed taking into account or involving future suppliers (i.e. supplier)

The description of all these factors, most commonly mentioned and uncovered through the analysis of 24 cases, are summarized in Table 5 below. We have also included how they affect open science adoption through the cases.

Table 5: Coding drivers and description

Category	Drivers	Description	Effects in open science adoption
Supporter	Institutional support	Institutional supports designs when an institution is behind the initiative	Support factors enable the gestation of the cases in the first place, but also the adoption of the initiative when such support comes from one of the target groups or affects target groups behaviour.
	Publishers support	Publishers support describes when a publisher or publishers are supporting in some form the initiative	
	Funders support	Funders support designs when the initiative has the support of a funding agency	
	Industry support	Industry support designs when the initiative has private sponsors	
	Community support	Community support refers to the situation when there was a support by the scientific community, citizens or final users in general behind the initiative	
Intention	Commodity	Commodity refers to whether the initiative was easily to plug in already existing measures, processes and practices	Intention factors enhance adoption by easing the process of incorporating such open science practice toward already existing behaviours and processes.
	Changes	Change refers to whether the initiative implied major changes in the work processes of the target groups	
Focus	User centricity	User centricity refers to whether the initiative was designed around user needs, understanding their practices patterns and common behaviour	Focus factors facilitate adoption by taking into account beneficiaries and stakeholders view in the design and the development process of the initiative.
	Supplier	Supplier refers to whether the initiative involved suppliers capable of providing the product or service and support adoption or was designed taking into account or involving future suppliers	

Overview

The coding of the 24 cases developed in the framework of the Open Science Monitor uncovered not only the factors described above but also it reveals that they are unequally important to the adoption of open science practice. Across the thematic analysis of cases, and coding exercise developed across the data sources, three major groups of factors, according to their relevance, were identified. First, community support and a supplier approach appear to be the most important factors explaining the case adoption and take up. Secondly, user centricity and institutional support appears almost as much as relevant to explain the success of the open science initiatives selected. Finally, funders support and the remaining factors would explain the main residual variance for the case adoption.

As the analysis show all factors uncovered had a significant role in the cases. However, we will zoom in now in the different open science trends to understand how drivers change and impact differently depending on the trend.

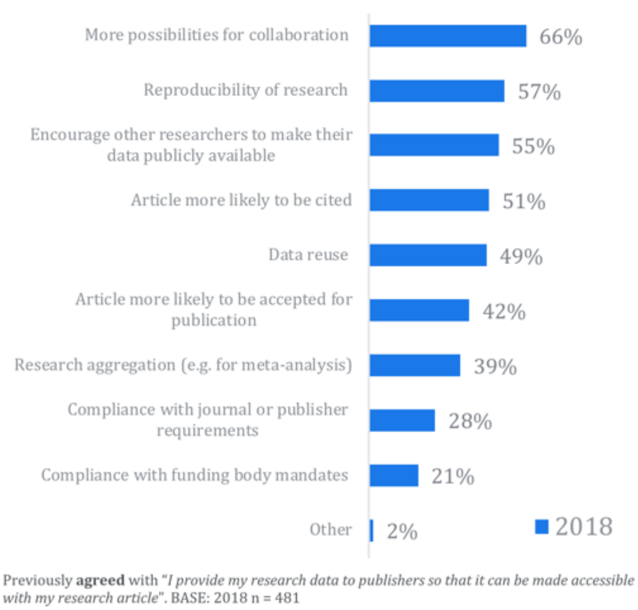
Analysis per open science trend

The analysis of cases reflected similar patterns across the diverse open science trends (i.e. open access, open data, open collaboration and citizen science). When we zoom into the different trends, we observe the following:

Open scientific data

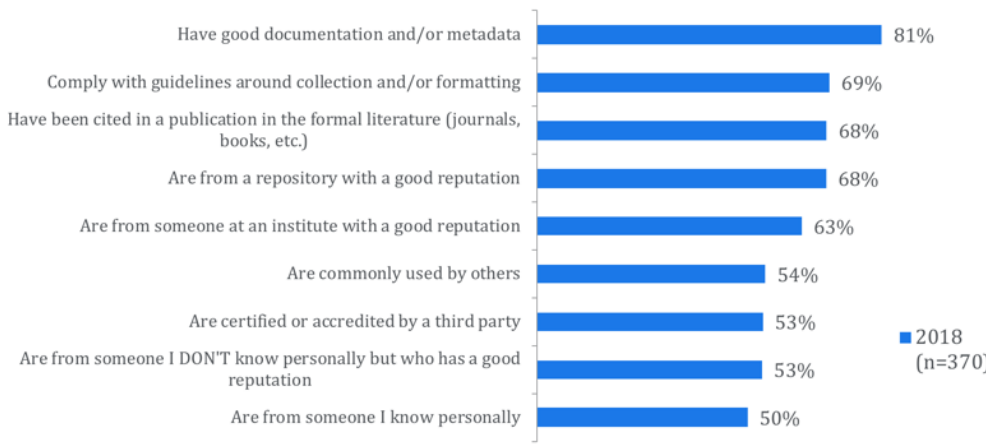
Looking at the factors driving research data sharing, those benefits related to advancing science through collaboration (66%) or reproducibility (57%) are considered to be more important to researchers than meeting journal/publisher requirements (28%) or funding body mandates (21%). The fact that sharing research data also encourages other researchers to make their data publicly available was also deemed to be of substantial importance (55%).

Figure 24: Drivers of sharing research data



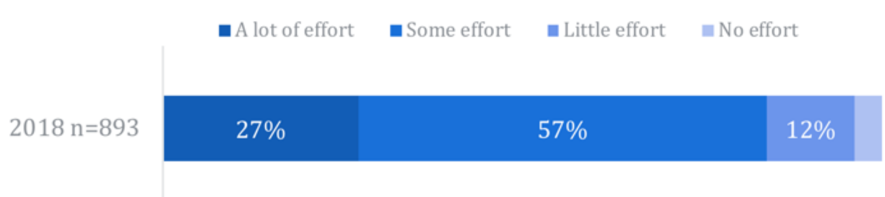
Good documentation and compliance with best practice are the most important factors for trusting and using another researcher's data. Citations in a publication in formal literature or a repository with a good reputation were also considered to be important factors. On the other hand, personal connections are less important. 44% of researchers used data that had been shared by others on their last research project.

Figure 25: How important were the following factors when deciding to use other's research data?



For more than half of researchers, "some effort" is required to make research reusable by others suggesting that a moderate amount of work has to be done to make data reusable for others. 27% of the surveyed researchers stated that "a lot of effort" was required, highlighting the complexity of data processing and formatting as a potential barrier to data sharing.

Figure 26: How would you describe the effort typically required to make your research data re-useable by others?



The major drivers present in the cases related to open scientific data show the relevance of institutional, community or funder support, user-centric design of the initiative and a supplier approach where the future exploitation is clear.

The qualitative analysis and fine-grained analysis through the cases reflect the following drivers within open scientific data:

Drivers
1. ***Context of low research and development (R&D) productivity***: Open science model is a strategy that contributes to fight lengthy, costly, low success rate, high attrition rates, and complexity in R&D processes, in particular in drug discovery processes.
2. ***New computational approaches***: The increasing value of aggregating data rely on new computational approaches that require different governance models
3. ***Responsible sharing*** of scientific data is essential to accelerate scientific discovery
4. ***Social or moral obligation***: Transparency in clinical research data can improve public health and safety; it is also considered a social obligation for the citizens' that volunteer themselves to those trials
5. ***Pressure from shareholders*** about the liability of unpublished data

1. **About fighting lengthy, costly, low success rate:** The main drivers seen in Open Targets, Pistoia Alliance and YODA are to effectively lower the barriers to R&D innovation and increase R&D productivity in biomedical research. As GSK declares: "We believe that harnessing the potential of "big data" and genome sequencing through this collaboration could help us dramatically improve our success rate for discovering new medicines," (GSK, 2017).
2. **New computational approaches and the increasing value of aggregating data in biomedical research:** The explosion of data availability in biomedical research has increasingly rendered it a data-driven science. As a result, drug discovery efforts accumulate today far more information than for previous drugs. There are vast volumes of data associated with drug discovery that is being accumulated in pharmaceutical firms. Through data aggregation, an integration of all the heterogeneous and complementary evidence available and a combination of computational approaches, the Open Targets case show how open research data is effectively facilitating scientists to systematically identify and assess all the evidence available that associates a drug with diverse diseases.
3. **About social and moral obligation:** Transparency in clinical research data can improve public health and safety. As the case of YODA reflects research data sharing is considered a social obligation for the citizens' that volunteer themselves to clinical trials. YODA is also driven by the view that patients, health care providers, and the greater life sciences industry will benefit if academia can provide independent reviews of data relevant to the potential benefits and harms of industry products, and that these analyses allow physicians and patients to inform their decisions on the most comprehensive and recent evidence available.
4. **About the pressure from shareholders about the liability of unpublished data:** One reason to believe that companies agreed to join YODA, but also other frameworks to disseminate clinical trial data, is that their shareholders are increasingly worried about the liability of unpublished data. As reported by different sources, owners of drug-companies support the broad dissemination of clinical trials result making sure that their results are fully reported (The Economist, 2015). The reason is that long-term investors

prefer to reduce the level of risk in their portfolio by having all the information about trials fully disclosed.

Open collaboration

Open collaboration is an umbrella concept that embraces: open source hardware, open code and reproducibility cases. The major drivers present in the cases related to open collaboration show the relevance of a user-centric and supplier approach behind the initiative and the support from community and institutions.

Regarding **open science hardware**, the more 'granular' drivers that were founded across the analysis of White Rabbit (WR) cases and the cross-analysis of existing open source hardware licences are (include here sources to the cases):

Drivers
1. To maximise the value of the technology for society by avoiding duplicative investments on the same hardware
2. To allow technology flexibility
3. To allow extensive peer-review of the technology design to increase technology quality and reliability
4. To allow reusability of hardware designs
5. Strong and historical commitment with open science by the initiative leader

1. **To maximise societal impact:** When CERN decided to develop WR as open source hardware to solve a problem in their accelerators, they had the intention to widely disseminate the technology to maximise their use by other organisations in research. Open source products in software are recognised for offering decreased dependency on monopoly suppliers (Bruns, 2000; Kogut and Metiu, 2001). This is analogous to hardware, and it is especially valuable for scientists. By adopting a similar approach for WR development, CERN intended to develop a platform to integrate a variety of contributions from dispersing stakeholders, while moving away from a vendor lock-in situation, where scientific infrastructures build dependencies with technology providers for highly specific technologies.

2. **To allow flexibility:** The Open source hardware approach provides the required flexibility for scientific organisations, which is an essential characteristic for scientists who need customising, never-before-seen equipment for their experimental endeavours in uncertain and modular environments. Such flexibility arguably leads to better and faster progress of science (Pearce, 2014). By developing WR in the open, CERN was seeking to pave the way for the high customisation potential of the technology, while fostering effective peer-review by a community of experts who can provide useful input, test cases, and feedback in an open space.

3. **To allow extensive peer-review of the technology design to increase technology quality and reliability:** By developing open hardware, the research infrastructure (CERN) was seeking to a dynamic peer-review of the technology by a community of experts who could provide useful input, test cases, and feedback in an open space.

4. **To allow reusability:** By adopting an open hardware approach, CERN was also expanding the possibility of reusing the designs of other electronics engineers working in experimental physics laboratories. The idea was to increase the efficacy of the technology by reducing the number of different teams working independently on similar problems and focus disperse yet complementary knowledge on a central goal.

5. **Strong and historical commitment with open science by the initiative leader:** The principle of openness is stated in CERN's founding convention, and the organisation has been a relentless pioneer in this regard since the release of the World Wide Web under an open source model back in 1994. CERN has continuously embraced the principles of open science, such as open access with the Sponsoring Consortium for Open Access Publishing in Particle Physics - SAP3 and all LHC publications have been published under Open Access conditions; open data, setting up the Open Data Portal for the LHC experiments or Zenodo, a free Open Data repository, launched by the organisation for use beyond the high-energy physics community; Invenio, an open source library management software package and their use of open source licences (Nilsen and Anelli, 2016; Murillo and Kauttu, 2018). Developing WR as an open source hardware was already in line with the philosophical *motto* of CERN research activity.

Reproducibility (including pre-registration)

Drivers
1. Awareness about 'reproducibility crisis'
2. Social impact maximization
3. Strong commitment by an organisation working to foster open science

1. **Awareness about 'reproducibility crisis':** Nature published in 2012 a study that reviewed ten years of research, where scholars found that 47 out of 53 biomedical research papers on cancer research was irreproducible (Prinz et al., 2011). Four years later, in 2016, Nature surveyed 1576 scientists, finding that more than 70% of researchers failed to reproduce another scientist's experiments and more than 50% failed to reproduce the results of their experiments. Also, a famous study was implemented in 2015 in the field of Psychology when an open, registered empirical study of reproducibility was launched and 270 researchers around the world work together and try to replicate 100 empirical studies published in the three top Psychology journals (the Reproducibility Project).

   As a result of this growing awareness, reproducible research has become a pervasive objective in the research policy agenda at different governmental levels, including funders and journal policies. In the struggle to improve the reproducibility of science, there have been different initiatives from the scientific community to fight the reproducibility crisis, which include our case: REANA, a reusable and reproducible research data analysis platform.

2. **Social impact maximization:** By doing REANA, CERN knew not only the particle-physics community would benefit, but others would also do, and that the project would benefit the vast scientific community.
3. **Strong commitment by an organisation working to foster open science:** REANA was born with the goal to fight the reproducibility crisis and as a natural next step of CERN's effort to move forward their open science efforts to the next step.

Citizen science

Finally, the major drivers present in the cases related to citizen science show the relevance of community, institutional or publisher support and user centricity. The cases assessed showed a successful adoption thanks to the support of the scientific community or organization behind it and the user-centric design that facilitated the contributions by a distributed base of citizens around the globe.

Drivers
1. Increased availability of datasets in many research disciplines
2. Motivation by project owners to produce annotated datasets
3. Usability of the technology interface

The cases such as Zooniverse show that the growth of citizen science projects is being driven by the increased availability of datasets in many research disciplines as well as the use of web-connected computer and mobile technology. On the other side, project owners are also motivated by the need to produce annotated datasets for research purposes. Finally, it is important that the citizen science projects make it easy to start new citizen science projects and contribute to them without any sophisticated technological expertise.

3.2 Barriers

There are many different drivers and barriers at play, which can facilitate or hinder the adoption of open science. Their identification is crucial to the development of appropriate policy recommendations, which are typically built around these "levers": they aim to remove the barriers and to enhance the drivers.

While there are general trends, external to science, that are important because they reinforce or damage open science, in the framework of the case studies performed, the research team attempted to identify the specific drivers or enabling factors that were behind the open science initiatives selected and the barriers or bottlenecks that the different project studied needed to face.

From the case studies, a number of recurrent barriers were anticipated. They are listed below:

Category	Barrier	Description
Micro	Lack of user orientation	When a tool, method, approach, or case as a whole was designed without taking into account or involving the future user groups.
	Lack of Adoption	When a tool, method, approach, or case as a whole is only used in a selected set of fields, user groups or institutes.
	Technological barrier	When a tool, method, approach, or case as a whole, shows that the technological aspects are too complicated for users.
	Lack of skills	When a tool, method, approach, or case as a whole requires too specific or too complicated skills of users.
	Cultural & behavioural barrier	When a tool, method, approach, or case as a whole, does not fit in the way that the research is usually carried out.
	Lack of awareness	When the existence of a tool, method, approach, or case as a whole, is only known in a small user group.
Meso	Lack of institutional support	When a tool, method, approach, or case as a whole is not supported by institutional funding.
	Lack of funders support	When a tool, method, approach, or case as a whole is not supported by funding from research funders.
	Lack of publishers' support	When a tool, method, approach, or case as a whole is not supported by publishers.
Macro	Lack of business models	When a tool, method, approach, or case as a whole is not supported by a business model.
	Lack of standards	When it is not clear whether a tool, method, approach, or case as a whole is standardized, and standards are easily available.
	Difficulty in upscaling - fragmentation	When a tool, method, approach, or case as a whole is not easy to upscale in other institutes/ user groups and hence the use remains fragmented.
	Lack of diversity	When a tool, method, approach, or case as a whole is not suitable to other individuals with different skills, backgrounds, traditions and beliefs.

For each of the case studies, the research team has assessed whether the barrier was applicable. A score from 1 to 3 was given, a 1 indicating that this barrier scored high in the case, a 2 meaning that the barrier scored medium, and a 3 when a barrier wasn't relevant at all. The scores were assigned in a group discussion, based on the case study story.

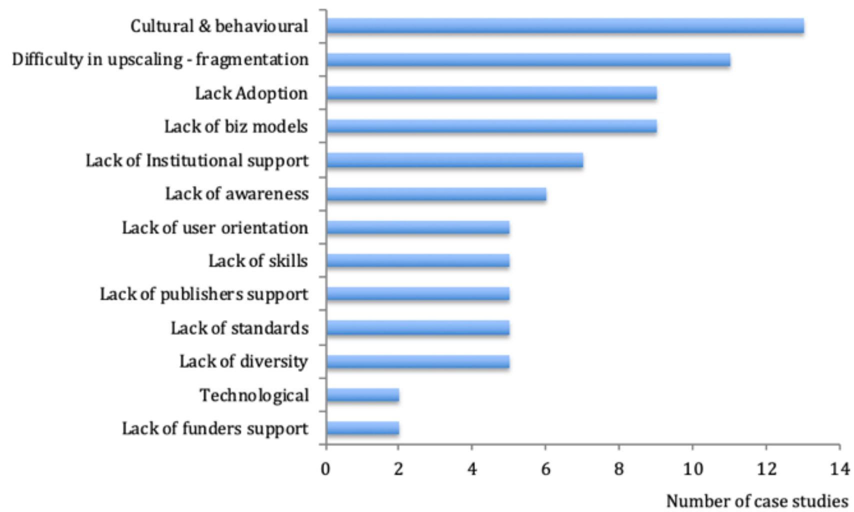
In the figure below is indicated which of the barriers was given a 1 (scored high) most often. It is clear that 'soft' barriers such as cultural and behavioural barriers are most frequent, occurring in 13 out 24 cases. The second highest barrier is difficulty in upscaling/

fragmentation (11/24), and the third one is the lack of business models (9/24), for which there are a high number of scores 2 as well.

The barrier that has the lowest score is technological aspects of a tool, method, approach or case; in only 2 out of 24 cases this barrier is scored a 1. Also, the financial support by funders (2/24) or publisher (5/24) is rarely scored as a strong barrier. Again, in the case where there was an equal number in score 1, the number of scores 2 was decisive.

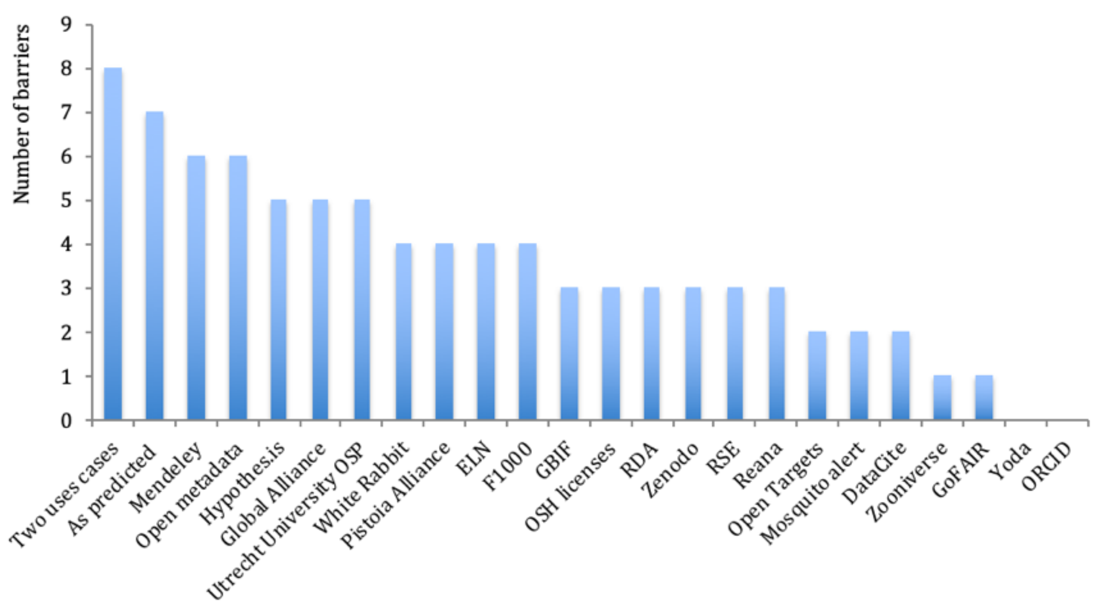
The barriers as scored here confirm earlier studies on open data practice.

Figure 27: Barriers across the case studies



The cross analysis of cases also shows that there is a broad range in the number of barriers that are experienced per case. In the figure below, the case studies are presented with the total number of high barriers (score 1). As can be seen up to 8 out of 13 barriers apply, but there are also cases which score zero high barriers.

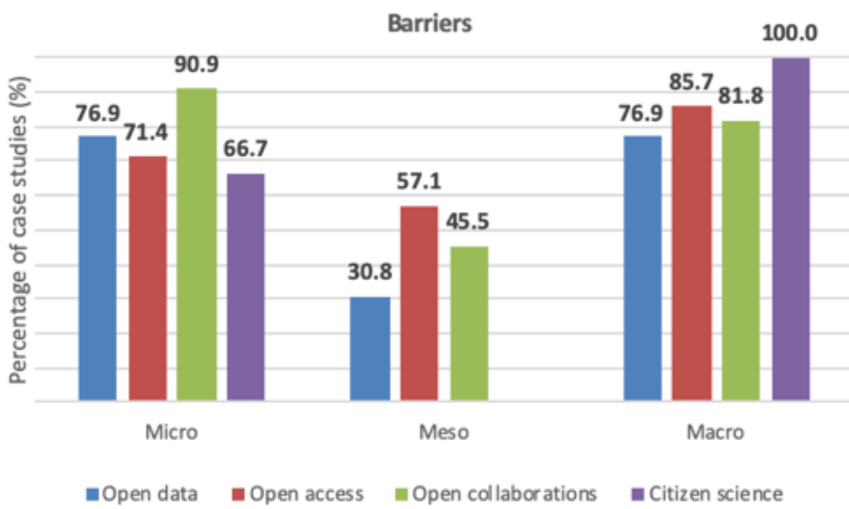
Figure 28: Number of barriers across case studies



The cases with the most barriers are *Two use cases*, *As predicted*, *Mendeley* and *Open metadata*. They cover the research phases of data-integration, pre-registration and publication. However, it seems that the barriers in the cases are not necessarily dependent on the research phase, because the cases with the lowest number of high barriers (score 1) are also in publication: YODA and ORCID with zero barriers, and Zooniverse and GOFAIR (one major barrier). From the number of barriers, it seems that broad 'technological' cases suffer from less barriers than 'use' cases. This is consistent with the high number of cultural and behavioural barriers, and the fragmentation.

In order to find patterns in the barriers, the score 1 was assessed across barrier categories and open science trends. The scores are normalized for the number of cases per trend, and the number of sub-barriers within the micro-, meso- or macro barriers. They are presented below:

Figure 29: Barrier across open science categories



The barriers are most prominent for Open collaborations (micro-level) and Citizen science (macro-level). This reflects the fact that there is strong policy support for Open access, also at meso-level, but that the actual behaviour at the micro-level is lagging behind in change. For citizen science the situation is the reverse: there is little policy support or infrastructure for citizen science, which shows to be the relatively strongest barrier. The lowest barrier scores are for Citizen science and Open data (meso-level), which both seems to be odd, since both citizen science and open data are in a premature state in HEI's. Otherwise, the barriers are more or less evenly spread.

4 Impact

This section presents the key findings on the impact of open science as emerging from the case studies. In addition, the survey provides additional insight on the impact of data sharing practices.

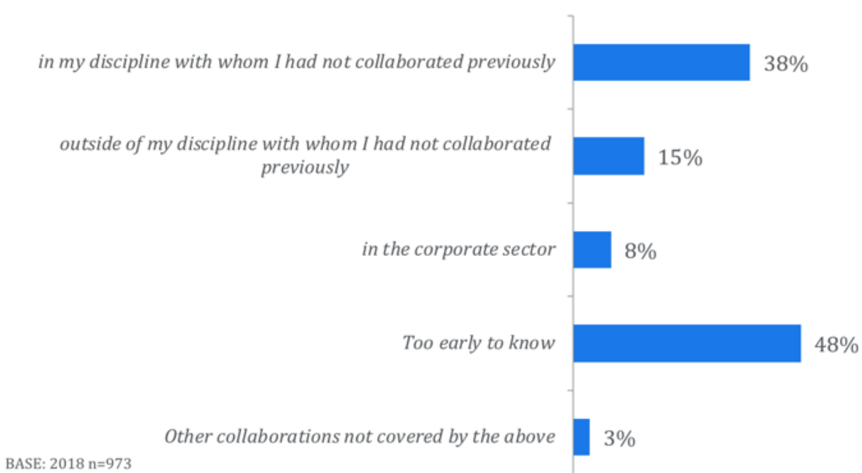
When looking at the consequences and impact of data sharing, over a third of researchers were contacted by another university/institute after sharing data. 10% were contacted by a company, highlighting the additional positive consequences of data sharing.

Figure 30: Thinking about the most recent research project on which you shared data, did individuals outside of the research team contact you concerning the data that you shared? I was contacted by researchers from

Furthermore, over a third of researchers believe that sharing data promoted new collaborations with researchers in their discipline.

Figure 31: Still thinking about the most recent research project on which you shared data, do you believe any of the following happened as a consequence of sharing the data?

Collaboration with [academic] researchers…



4.1 Scientific impact

There is substantial evidence that open science is correlated with better science, mainly by removing mistakes at an earlier stage, enhancing the productivity of science through data reuse and reducing the costs of science through collaboration. The main different positive effects of implementing an open science approach captured in the literature so far include:

- *Faster discovery:* Fast track to discovery thanks to external contributions and comments. Knowledge mining of open access literature can automatically classify, analyse and reason on existing literature, making new discoveries and connections that currently happen very seldom and unexpectedly.
- *Increased efficiency:* Open science increases the efficiency of the research system because it helps to reduce the duplication of costs, as well as costs stemming from the creation, transfer and re-use of data. It allows more research to be created by the same data and thus boosts return on publicly funded research (OECD, 2015; Lyon, 2009; Whyte & Pryor, 2010).
- *Greater productivity:* Contributions from volunteers to datasets.
- *More robust findings:* Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results (Wicherts et al., 2011). In the absence of open code, computational science as practiced today does not generate reliable knowledge but "breezy demos" (Ioannidis 2005).
- *Increased transparency and replication of studies:* Open science raises transparency and quality in the validation of research results and multiplies the opportunities for replicability and validation of scientific findings (Franceschet & Constantini, 2010; OECD, 2015; Fecher et al. 2015). By making information on methods, protocols, and data easier to peer review, and by strengthening the scope of the material published (including negative results), open science offers greater scrutiny as it allows for more accurate verification of research results (Lyon, 2009; Whyte & Pryor, 2010).
- *Increased collaboration and scientific interdisciplinarity:* The implementation of an open approach in science enhances collaboration across institutional, national and disciplinary boundaries, and it fosters the exchange of information and expert knowledge (OECD, 2015; Whyte & Pryor, 2010). Scientists are re-scaling the level of contribution in the scientific process by cutting it into small pieces and enabling micro-contributions on a macro-scale. Hyper-specialisation is foreseen by scientists in many cases and they look for micro-contributions from 'citizen scientists', which could lead to a further gamification of Science (Von Ahn & Dabbish, 2008; Szkuta & Osimo, 2016).
- *Minimise publication biases:* Minimise publication biases if studies that report negative or no effects are not published in traditional venues, but are more likely to be published on novel publication platforms (Boulton et al., 2012).
- *New research fields and enhanced opportunities:* Big Data is helping to create brand new fields of science: computational chemistry, biology, economics, engineering, mechanics, neuroscience, and geophysics. Open science enables sharing of complex models and simulations based on large-scale open data analysis (Lyon, 2009; Boulton et al., 2012).

Modern computers can identify highly complex, unperceived relationships. Thus, technology is supporting traditional ways of doing science, and enables enquiry by constructing hypotheses after identifying relationships in the datasets.

- *New emerging instruments to evaluate science:* Over the last ten years, with the adoption of web 2.0 tools by scientists, new instruments are emerging to assess the quality of science outputs (Herman et al., 2015). Today, reputation (which is key to scholarly research) is intrinsically linked to the model of journal publication. The impact of research is measured by citations and patents. Recently, there have been experiments with new evaluation and reputation models, such as data citation and data journals, altmetrics and post-publication reviews (Szkuta and Osimo, 2016).

Regarding the potential negative impacts, the literature suggests that open science may also pose various threats. Most of the concerns are related to research quality: With the openness of the research process there are some concerns about the effectiveness and practicality of quality checks. There is a widespread acknowledgement that assessing the quality of such vast volumes and ranges of scientific materials is challenging (Whyte & Pryor, 2010).

The main questions that persist after the revision of the different studies that Open Science Monitor sought to address through the selection and analysis of a set of case studies across open science trends are:

- How does open science impact the quality of research?
- Under which mechanisms does open science increase the efficiency, effectiveness and productivity of research?
- How new forms of evaluation of science affect scientific performance?
- How open science impacts multidisciplinary research?

4.1.1 Cross analysis of cases

According to the assessment of the cases, in 19 out 24 cases the scientific benefit of the tool, method, approach or case is considered high by a score 1. Only five cases score a 2 - Pistoia Alliance, Two use cases, DataCite, RDA and Open metadata. These cases cover a broad range of research phases so again there is no specific phase where open science is critical for scientific impact.

Figure 32: Scientific impacts as listed in the cross analysis of policies

	Faster discovery	Increased efficiency	Greater productivity	More robust findings	Increased transparency and replication of studies	Increased collaboration and scientific interdisciplinari	Minimise publication biases	New research fields and enhanced opportunities	New emerging instruments to evaluate science
GBIF	x		x	x	x	x			
Hypothes.is	x	x			x	x		x	
Mendeley		x			x				x
White Rabbit				x		x		x	x
OSH licences									

	Faster discovery	Increased efficiency	Greater productivity	More robust findings	Increased transparency and replication of studies	Increased collaboration and scientific interdisciplinari	Minimise publication biases	New research fields and enhanced opportunities	New emerging instruments to evaluate science
Open targets	x	x	x						
Zooniverse				x	x	x		x	
Mosquito alert	x	x			x				
Pistoia alliance	x	x		x		x			x
Global Alliance									
Two uses cases									
Utrecht OSP		x			x	x			x
RDA						x			x
ELN									
Zenodo									
GOFAIR		x			x				
As predicted									
RSE									
F1000									
YODA	x	x			x				
Open Metadata							x		x
REANA									
ORCID					x		x		x
DataCite									

4.1.2 Qualitative insight from cases

The open data case studies were all bottom-up initiatives in the broader biomedical field in which Pistoia Alliance and Open Targets were industry-driven; and YODA was led by an academic organisation. Therefore, it is no surprise that the scientific impact of these cases relates to faster discovery and increased efficiency. This may provide an answer to the second key question.

All case studies show that they aim to bridge practices that are not necessarily taken up in 'normal' scientific practices. It takes time, effort and funding to set up infrastructures or develop tools that have primarily a bridging capacity. In all cases, industry is taking part and is bringing in funding. Large sustainable infrastructures such as CERN provide environments in which open science initiatives flourish benefitting the overall scientific community.

In the current cases, except from YODA, there seems to be low impact on minimising publication biases. This may have to do with the fact that they are all bridging activities that do not operate in the core of the scientific publication arena, where most science is taking place.

Open scientific hardware is an emerging formula being explored by research infrastructures that produce experimental technologies. The example of **White Rabbit** (WR) provides relevant insights in how the publication of even a first version of the White Rabbit designs, resulted in the uptake by the scientific industry, and other scientific infrastructures and facilities. At present, according to last update in the WR repository, 30 organisations have implemented WR and 15 others are current evaluating if the technology can fit their purposes. In addition, nine R&D projects have applied White Rabbit technology for their scientific purposes.

Examples of this scientific application include GSI where WR became the timing system of the Facility for Antiproton and Ion Research (FAIR), an international accelerator based in Germany of an estimated investment of $2 billion. Also, the Cubic Kilometre Neutrino Telescope (KM3Net), a European research infrastructure located at the bottom of the Mediterranean Sea, uses WR in order to synchronize the detector units.

Open Targets is an innovative, large-scale, public-private collaboration on pre-competitive research that provides comprehensive and up to date data for drug target identification and prioritisation. Open Targets integrates publicly available information and data relevant to targets and diseases in the Open Targets Platform. It performs high throughput experimental projects that generate target-centred data in physiologically relevant systems to understand causal relationships between targets and diseases in three therapeutic areas: Oncology, Immunology, and Neurodegeneration. A cornerstone of this public-private collaboration from the beginning is an agreement among the organisations that all data and resources generated within Open Targets should be made available rapidly in the public domain to the entire scientific community. The Open Targets Platform is an open access "Google"-type search engine that extensively searches assess and integrates the huge quantity of genetic and biological data available. The platform supports two main workflows: the first is target-centric; the second is a disease-centric workflow.

The explosion of data availability in biomedical research has increasingly rendered it a data-driven science. There are extensive volumes of data associated with drug discovery that is being accumulated in pharmaceutical firms. Drug repositioning is one strategy in drug development that seeks to expand the indication space for a successful drug or find a new indication for a drug that was not successful in the clinical trials. Open Targets facilitates identification of potential repositioning opportunities. Through the Open Target Platform, scientists are able to systematically identify and assess all the evidence available that associates a drug with diverse diseases. Basically, the aggregation and integration of all the heterogeneous and complementary evidence available in the platform help prioritise and analyse potential drug repositioning. As reported by Khaldakar et al. (2017), Open Targets has been able to uncover 2,540 potential new indications for 791 existing drug targets. Among these 2,540 new indications, 1,366 are for rare diseases where the target is associated with more common diseases.

Regarding the use of the Open Targets Platform, 900 unique IP addresses access the Open Targets Platform every week. Some metrics available about Open Target platform usage, from April 2016- March 2017, reveal that the platform is used substantially. The metrics are also aligned with the qualitative feedback reported by the Open Targets team from platform users.

The scientific impact of the evidence available in the platform is substantial. Associations between drug target and disease are the main focus for both new drug development and drug repositioning. Scientists seek evidence supporting target-disease associations that can be stored in structured databases and integrated to obtain a comprehensive assessment in target validation studies (Kafkas et al., 2017).

The fundamental conclusion from the Open Targets case study is that the inherent tension between the goals of scientific openness and commercial exploitation does not necessarily imply incompatibility, but a need to identify sophisticated solutions that adequately balance the divergent interests at different phases of scientific processes.

The **Yale University Open Data Access Project (YODA)** was launched by the Center for Outcomes Research and Evaluation (CORE) from Yale University of Medicine in 2011, to address the problem of unpublished, and selectively published, clinical trial research data. YODA has the goal to increase access and availability of clinical research data while promoting the re-use of such data to create new knowledge. YODA has been acting as an independent, unbiased bridge between researchers and clinical data from pharmaceutical companies and consumer businesses. To do so, YODA has developed a *model* to make data available to scientists, which is mediated through the YODA team. YODA facilitates access to clinical trial data made available by third party Data Holders to promote research that supports scientific endeavours and public health. YODA partnered with Medtronic and Johnson & Johnson to share its data.

Metrics on all data requests, including response times shows that currently, there are 90 clinical trials listed for a request. Different types of stakeholders have requested YODA for accessing clinical trials data, including governmental institutions, industry, and academia. The majority of requests for accessing clinical data trial come from academia (~90% of requests), and usually the requests are accepted.

Despite governments and regulatory agencies efforts towards transparency of clinical research data, a great deal of clinical trial data is partially or completely undisclosed. The first lesson is that regulatory enforcement can play a critical role in increasing disclosure. The framework from YODA can also be applied for other fields that expand medical research, where data is considered sensitive and where stakeholders share significant concerns that prevent them from disclosing relevant data to accelerate scientific discovery. To be re-used for scientific goals, clinical research data has to be disclosed with complementary and comprehensive sources of clinical research, which has a level of evidence granularity that is exceeds journal publications; annotated case report forms; dataset specifications; protocols with any amendment and reporting and analysis plan; amongst others.

YODA has carried out active dissemination about the project and potentialities of re-using clinical data research. Sharing the data is insufficient to accelerate its re-use; active dissemination and engagement of the research community are needed to foster analysis of already existing data.

Pistoia Alliance is an *industry-driven* partnership created by large pharmaceutical companies (GlaxoSmithKline (GSK), AstraZeneca, Pfizer, and Novartis) in 2009 without government intervention to cooperate and share resources in the pre-competitive phase of drug development. The Alliance has been expanding since then accepting applications from both private and public who share the ideology of sharing data and resources of drug discovery research pipeline. Pistoia started to generate *mutual standards* in industry, ontologies and web services that are made available under an open-source framework to academic institutions, vendors and companies, to facilitate data sharing, data representation, text-mining activities and improve the R&D efficiency of the drug discovery and development process. With many projects underway since its foundation, covering from Internet of Things (IoT) to 'ontologies mapping,' the alliance pools resources from many companies, and openly shares the output of its work with the outside research community. The alliance has designed different types of membership in order to engage a diverse set of interest in enhancing life science R&D.

When Pistoia Alliance started its activity in 2009, the partnership set out to work around four major technology pilots: Sequence services; SESL (Scientifically Enriched Scientific Literature); Electronic Lab Notebook (ELN); VSI (Vocabulary Standards Initiative).

Pistoia Alliance has a project portfolio, which is continuously evolving with projects entering the pipeline periodically. In the case study, one use case (project) was analysed to show the impact of the collaboration: *Hierarchical Editing for Large Molecules (HELM)*, which is a global standard notation that is machine-readable for large molecules originated at Pistoia Alliance[4]. HELM is now a FAIR standard. The project team included members from 23 different organisations in the life science space integrating a large and diverse ecosystem. The HELM technology was officially released into the open source community consisting of a GitHub repository containing the source code of the toolkit and editor, along with a web site (www.openhelm.org) containing the notation language specification, a free applet version of the editor, training videos and user guides. HELM notation became an open standard, published openly and for free. HELM is increasingly being adopted through the companies and academic organisations working on biopharma, including also solution providers and regulatory agents (Pistoia Alliance, 2018).

The uptake of HELM has been rapid and includes today a vast range of organisations as content providers, informatics vendors, and life science companies. The benefits of using HELM include facilitating the work of scientists willing to employ computational manipulations and calculations with complex molecules. HELM makes easier researcher tasks. By doing this, HELM facilitates the integration of private and public data easier using the standard.

4.2 Industrial impact

Moving innovations from scientific discovery to the commercialisation of new products and services involves numerous stakeholders and challenges to overcome (e.g., Carayannis and Campbell, 2009). At one end of the spectrum there is a heavy concentration of government investment in basic research; and at the other, in the commercial marketplace, there is a much higher level of industry investment in direct product development (National Science Foundation, 2017). Between both of them, there is the path for potential innovation but firms and industry players need to invest resources to develop them to a stage where their commercial potential can be exploited. Successfully going this process requires complex ties and relationships through the innovation spectrum, which usually involves contractual relationships with non-disclosure-agreements (NDAs), Intellectual Property issues and

[4] About HELM: www.openhelm.org

other formal approaches for collaboration that guarantee that the interests of the different stakeholders are protected.

The sociologies of scientific communities are often unique and the pursuing of openness and free sharing in the scientific process can be at odds with the institutional logic of the business world. On the one hand, the norms of Science, which were famously described by the sociologist Robert K. Merton (1973), highlight the cooperative character of inquiry, emphasising that knowledge growth stems from a collaborative process where full disclosure of findings and methods are fundamental. On the other side, for-profit seeking firms follow the rationale of private investment models that seek to protect R&D investments to extract rents from the resulting new products and services in the market. For instance, the traditional business model for most commercial hardware companies is based on keeping the details of the design secret to be able to maximise the margins (Pearce, 2017).

In the current context, this historical 'tension' coming from the sociology of scientific community and business world, which has been extensively studied by scholars especially since the '80s after the Bayh-Dole Act, needs to be re-examined in the light of the recent events boosting open science, where policy-makers and the scientific community itself have started placing greater emphasis on the transparency and public scrutiny of scientific processes (OECD, 2015). Open science mandates across the scientific community stress the efforts by scientists and governments to make the primary outputs of publicly funded research results (both data and publications) publicly accessible in a digital format, without any (or minimal) restriction in order to accelerate knowledge growth, while enhancing transparency and collaboration.

The academic and grey literature assessing the impact of applying an open science approach highlights as ***positive impacts*** the i*ncreased collaboration and innovation.* The implementation of an open approach in science enhances collaboration across institutional, national and disciplinary boundaries, and it fosters the sharing of information and expert knowledge (OECD, 2015; NESTA, 2010). By facilitating access to research data and outputs, open science fosters knowledge spillovers and innovation across different economic sectors. It facilitates collaboration among research organisations and businesses, enhances firms' absorptive capabilities (Cohen and Levinthal, 1990; Fabrizio, 2009; Sorenson and Fleming, 2004) and makes for a swifter path from research to innovation by cutting delays in the re-use of scientific results, including data sets (OECD, 2015; NESTA, 2010). Also, the increase *access to results* by improving public access to research outcomes and data has been highlighted as positive impact of open science towards business, which fosters spillovers from science and research (OECD, 2015). As an example, the use of data from PubMedCentral, the US National Institutes of Health's repository, shows that 25% of the unique daily users are from universities, while 17% are from companies and 40% from individual citizens (UNESCO, 2012). A recent study on R&D-intensive SMEs in Denmark shows that 48% of the firms consider research results essential to their businesses, and over 65% report barriers in accessing research outcomes (Houghton *et al*., 2011).

On the other hand, other scholars suggest that the forces of *sharing* scientific knowledge and *protection* of commercial interests run in opposition. The commercialisation of scientific outputs usually requires a significant investment that companies are willing to bear if they can protect the innovation from imitation or unfair competition. Some scholars highlight the potential risks for firms engaging in open knowledge systems, with risks arising from the opposed institutional rationales of the worlds of science and technology (Jong & Slavova, 2014). For instance, Perkman & Schildt (2015) show that participation in Open Data partnerships with universities may jeopardise a company's attempts to capture value from research.

4.2.1 Cross analysis of cases

The diverse case studies carried out in the framework of the Open Science Monitor provide an understanding of the ***effects*** of open science trends on innovation, explores the ***drivers of such effects*** and in particular assess ***under which conditions*** open scientific practices are compatible with the market logic. We discuss the three of them in the following sections.

First, regarding the major **effects** that the analysis of the thirty cases studies reflect are the following. We describe the effects by employing the examples of the cases that score the highest in terms of industry impact (i.e. White Rabbit; Comparative analysis of Open source hardware licences; Open Targets; Pistoia Alliance; Re-use of public data; Open metadata; and GOFAIR). However, all cases have informed the present assessment, thus we invite readers to read them for further information.

1. Re-use of scientific outputs by firms in R&D processes

Opening up scientific data (open scientific data), the designs of scientific experimental tools that later find applications in different industry settings (open source hardware), code that can be re-used in data analytics processes (open code) and making available the processes through which scientists reach conclusions (open notebooks), amongst others, show in the cases positive knowledge spillovers towards businesses.

For example, in open scientific data, the Open Targets case reflected that not only partners within open targets consortium benefitted of the data published in open targets platform for better identifying and validating targets, but also external stakeholders (see Two use-cases of Re-use of open data) take advantage from the open data. Open Targets receive 1000 visits average per week; but also, we see an industry emerging of intermediate players supporting pharmaceutical, biotech and companies in biomedical research to extract value by integrating open scientific data in their workflows and internal information systems. Pistoia Alliance case also reflected the benefits from open scientific data and the need for industry to implement FAIR-ification of data and adopt data standards so that data interoperability across systems is not a barrier to grasp such benefits.

Additionally, the case of White Rabbit show for open source hardware, the positive effects of opening up designs and schematics about scientific hardware. White Rabbit, which was first conceived as a solution for the synchronisation of CERN's distributed network, find applications later on in a wide range of scientific but also industry settings such as financial services (Frankfurt stock exchange), telecommunication operators (e.g. Vodafone) but also Smart grid, air-traffic control or automation vehicles. If the sponsor organization, CERN in such case, would have chosen the alternative situation, that is to outsource the development of White Rabbit technology to a vendor and protect the design with IP instead of developing as an open source hardware, all the other scientific infrastructures that were in need and re-used such design, would have had to incur the same investment for developing something that may be not as good as this one that leverage collective skills and expertise. The comparative analysis of open source hardware licences sheds light in how businesses may benefit from open source hardware licences by setting governance rules that allow proprietary developments and improvements to emerge.

2. Acceleration of industrial R&D processes: Efficiency and Efficacy gains

The growing complexity in R&D processes is increasingly requiring collaborative methods across different stakeholders (including competitors) to try to leverage complementary knowledge and skills to accelerate R&D processes. This is precisely what the case of Open Targets reflected where companies have been able to gain efficiency in the early stages of drug discovery process by sharing and openly disclosing pre-competitive data and knowledge regarding targets. They have also been able, by employing such collaborative resources, to identify better targets and thus to improve efficacy.

Other cases such as GOFAIR seeks precisely to accelerate R&D processes by making the scientific data available to be re-used by different players starting by themselves. Cases such as Pistoia Alliance or GA4Health showed us the importance of setting up and widely agreeing on standards in order to make possible such re-use of data and knowledge within and across organizations, which turns in efficiencies in R&D processes for pharma and biotech industry.

3. New products and services emerging around open science

Open science can also lead to new products and services commercialised by a set of companies. The example of two use cases of companies employing public scientific data to build semantic platforms and analytics that they sell to pharmaceutical industry and other players in R&D biomedical research; consultancies that sell their services to other players in the ecosystem to work around platforms of public data such as open targets which help to integrate public data in their workflows.

Beyond data, the case of White Rabbit also reflected how companies were able to re-use and sell White Rabbit switches and nodes and their services to different organizations in multiple industrial settings. Currently four vendors offer White Rabbit technology, according to the information in the open source hardware repository. Beyond the cases selected, there are different examples that show how open science is generating an industry with different products and services around public resources to accelerate re-use and integrate what is available in internal processes.

4. Increase quality of R&D processes

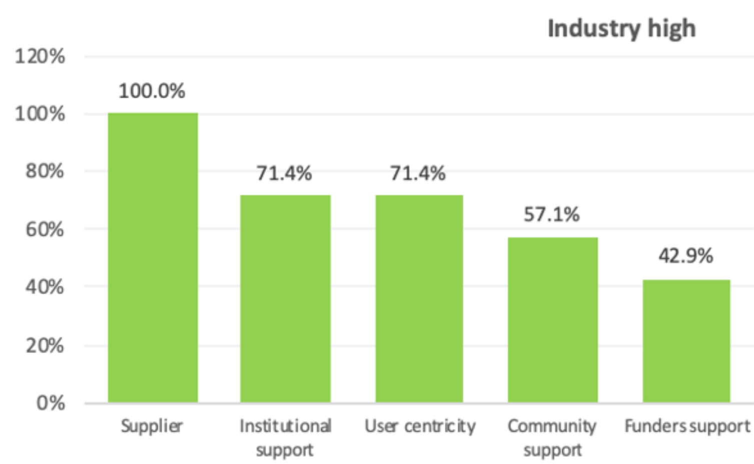
The cases such as YODA or REANA who have reproducibility as their *raison d'être* moved R&D processes towards increase quality and overall reliability. REANA who started as CERN project early find an application in the life sciences domain and YODA provides good use cases about how the complete and full availability of clinical trials helped pharmaceutical companies to scientifically crowd source a better understanding of their results. Open Targets also describes how collaborative efforts in pre-competitive research improve the quality of the early stages of drug discovery process; and White Rabbit, clearly reflected how the open source approach allowed to pool together different types of expertise and to crowd source the identification of bugs in the technology which lead to better quality of WR overall.

5. **New cooperative strategies in industry R&D processes**

The cases in the monitor describe emerging cooperative strategies and new governance models that are being built around open scientific outputs. Some examples include: the creative governance models put in place in YODA to govern the access and control towards data between companies who are data holders and the scientific community; open targets also set up a set of governance boards and rules to allow both the disclosure of relevant information to cooperate in a pre-competitive space while allowing protecting relevant assets from competition; the evolution of the licences in White Rabbit which governed the collaboration across contributors was also insightful in seeing how CERN set up a licence which go through a set of debates around the community who finally agree upon some basic rules of the game.

Second, going through the thematic analysis carried out across the thirty cases and the coding exercise implemented by the Open Science Monitor research team, we uncover the ***drivers*** of the cases that score the highest in terms of industry impact. Not surprisingly we see that the cases with high industry impact also score the highest in terms of supplier approach. User centricity approach appears relevant, which is logic when thinking about the re-use of data and knowledge, which means taking seriously users' workflows so that the benefits of public disclosure are grasped. Finally, all those cases received the support from institutions, community and funders

Figure 33: Top five drivers found in case studies with high industry impact



4.2.2 Qualitative insights

While companies increasingly seek to cooperate with other organisations outside firm boundaries to tap into ideas for new products and services (e.g. Chesbrough, 2003), mechanisms that motivate innovators to reveal the processes and the outcome of their investments openly are not evident.

The results of the case studies are coherent with what previous literature described. Open science benefits innovation by enabling a swifter path from science to new products and services in the market. However, as the case studies show, this can only happen if different mechanisms are put in place in order to make compatible both goals of widely sharing the scientific outputs while letting companies protect their private investments in the R&D process that they are being part of. Those mechanisms vary depending on the open science trend:

Open source hardware

The example of WR provides relevant insights on how it is possible to go one step further from the model of open source software to other, more capital-intensive technologies that require different incentives, dynamics, and conditions, combining open revelation and technology commercialisation. The open source approach possesses an "essential tension" about appropriating the returns from an innovation versus gaining adoption of that innovation (West, 2003).

From the White Rabbit case, we learned that the initiative would not even "take off" **unless** the right incentives were in place so that suppliers and other stakeholders would contribute their knowledge to an open source scheme where both R&D processes and outcomes were open to everybody.

WR identified the following *mechanisms* that needed to be put in place in order to stimulate private R&D in an open source hardware scheme, that benefited a vast number of scientific infrastructures (higher than 30 research organisations using White Rabbit), but also the technology found later a commercial application in the market (e.g. in the financial sector, telecommunications, smart grid, IoT, air traffic control etc.):

Fundamental mechanisms identified in the case include:

- *New legal framework:* CERN Knowledge Transfer Group along with its engineers provided a new legal framework, the Open Hardware Licence, inspired by the "Tucson Amateur Packet Radio" licence (TAPR[5]), to foster knowledge sharing among the diverse organisations (its first version was released on March 2011).
- *Compromise between proprietary and open hardware:* The licence made compatible, after a special request by multiple firms and agreement of the organisations being part of the WR development, that contributors were not obliged to disclose improvements to the technology. Under the Open Hardware licence, organisations were not obliged to notify the changes to upstream licensors. Thus, while contributing to the open source core technologies of WR, firms were able to develop in parallel their proprietary hardware based on WR technology.
- *Standardisation:* It was crucial for stimulating both R&D revelation by companies taking part in WR development to provide some stability around the technology. Going through a standardisation process was envisaged to additionally increase the stability, viability, and credibility of the technology by soliciting the feedback of experts to fine-tune the technology further. Firms were willing to put private investment and help developing WR if they could anticipate some stability over the technology, meaning that the technical specifications would not change over time. This is why CERN, in collaboration

[5] Description of TAPR in Open Hardware Repository: https://www.oshwa.org/research/brief-history-of-open-source-hardware-organizations-and-definitions/

with other research infrastructures and companies (part of WR development), decided to start a standardisation process of WR technologies.

- *Arbitrage and orchestration of exchanges by an academic partner:* CERN not only had to engage new companies capable of providing WR technology proactively but also it had to govern the exchanges, being an arbiter of differences when needed (Serrano, 2016). The different organisations involved in WR development accepted the arbitrage of CERN in the different exchanges and its leadership.
- High engagement and close interactions with academic partners in order to provide *learning and reputational gains* by companies developing the hardware. Firms were envisaging an upgrade of their technical skills by closely collaborating with highly qualified engineers at CERN. Additionally, organisations engaged in the development expected to be differentiated from future market competitors by being considered a tier-1 partner of CERN. The branding and reputation of such collaboration were expected to signal the quality of the organisation while extending the companies' capacity to attract clients.

Open scientific data

Open Targets, Yoda and Pistoia Alliance cases showed that open research data benefits industry and accelerates innovation processes if different mechanisms are put in place in order to provide the right incentives and conditions for companies to both share data, and to re-use third party data. From the three cases the following main *mechanisms* were identified:

- Agreement among industry players of *what* data and resources are considered pre-competitive and what is *competitive***.** Pre-competitive data is susceptible to being shared, while competitive information and knowledge are going to be held secret simultaneously. As an example, the collaboration within Open Targets required a clear delineation between what was agreed to be shared amongst organisations and what remained closed for the participants. The 'discontinuity' in the line going from cooperation to competition needs to be demarcated to generate a comfortable environment for corporations to collaborate and share data to accelerate the early stages of drug discovery while preserving their assets to compete amongst each other in later stages of the development process. Firms in Open Targets agreed to collaborate and disclose data and knowledge related to phases 1 (target identification) and phase 2 (target validation) to succeed in systematically find the best targets to safely and effectively treat disease. However, firms do not reveal the knowledge and proprietary data that they are going to use to identify the multiple molecules active against the potential targets, nor other data useful in later stages of the drug development process (i.e. phase 3 -lead optimisation- and onwards).

- The fundamental need for *data standards* to effectively share research data. As GOFAIR, Pistoia Alliance, GA4Health, RDA and other cases show, without common data standards widely adopted by academic and industry organisations, the benefits of open research data cannot be fully grasped by firms. In order to actually facilitate data exchanges, and thus tap on the benefits of artificial intelligence, machine learning and any computational approach relying upon the aggregation of high volumes of quality data, the industry needs to agree and widely adopt standards collectively. Pistoia Alliance case showed that standards are fundamental in biomedical research (e.g. Unified Data Model) but are also resource intensive because they require hours of distributed teams working together to agree on common tech specifications and solutions. Once the standard is out, if it successfully implemented by a wide range of players, it can dramatically reduce inefficiencies and accelerate information sharing within and outside organisational boundaries.

- *Arbitrage* by an academic partner. For instance, in the case of Yoda, the initiative led by the Centre for Outcomes Research and Evaluation (CORE) from Yale University of Medicine, developed a model to make data available to scientists, which is mediated through YODA team. YODA facilitates access to clinical trial data made available by third party Data Holders to promote research that supports scientific endeavours and public health. YODA does not support nor facilitate access to data for commercial purposes, litigation, or any goal that is not purely scientific. The model works as follows: Data holders give access to their clinical research data to YODA. The YODA team follows a strict protocol to grant access to these data to researchers. Another example was provided by Open Targets, which showed a model to pool together distributed capabilities and resources from the different organisations and companies through *matchmaking* exercises to match people from distributed teams working on a joint research idea; *merging* companies individual investments devoted to early drug discovery in a joint research agenda; and other additional approaches that helped orchestrate the distributed resources (data, capabilities and expertise) with a prominent role of the academic partners in the governance of exchanges.

- High engagement by academic partners to increase *learning gains* and possibilities for research data re-use. Both Pistoia Alliance and Open Targets cases described how face-to-face exchanges and close interactions are still crucial and need to be combined with an interactive platform that facilitates data and resources exchanges (e.g. IP3 in the case of Pistoia Alliance and Open Targets Platform).

  Technological platform approaches need to be combined with close cooperation among the organisations that generate the data and those willing to re-use the data. As the case of Open Targets described, sharing scientific data in the public domain is a necessary but insufficient requirement for being able to re-use such data for drug development purposes. Companies taking part of Open Targets consortium agree to private investments – both cash and in-kind – in order to be close to the scientists that generate the data to be able to understand the data, how it was generated, and how to interpret it.

- *Time lag* and *'premium' features in data platforms* for companies sharing the research data. The case of Open Targets describes a model where companies generating the data agree to publicly share all scientific data produced in the framework of open targets collaboration. However, while partners in the consortium have access to the research data from minute one, the data is open to the public through the open target's platform after publication. This time difference gives competitive advantage towards firms inside Open Targets consortium. On the other side, the functionality of the Open Targets platform by the public and partners differ. For instance, partners within the consortium can aggregate private data in the platform (e.g., compound libraries) to accelerate further stages of drug discovery.

General patterns across trends

The analysis of the results of the case studies implemented agrees with the literature that claims that open science benefits innovation by enabling a swifter path from science to new products and services in the market. However, the case studies also show, that this effect can only happen ***if*** different mechanisms are put in place. Table 6 below summarises those mechanisms also described above basically for the two open science trends that reflected higher industry impact: open scientific data and open science hardware. Those two trends which concentrated most of the cases in this respect.

Table 6: Analytical summary of mechanisms enabling the positive impact of open science to industry

	Open scientific hardware	Open Scientific Data
Governance	New legal framework: See comparative analysis for different governance models in OSH. There is a need for licences that enable the non-disclosure of improvements or the compromise between proprietary and open hardware.	There is a need for a prior agreement by companies and academic partners about what is going to be disclosed (pre-competitive) and what falls under the competitive arena: what data/resources are not subject to IP; what are subject to IP.
Standards	Standardisation process to increase the stability, viability, and credibility of the technology developed openly to allow peripheral developments to emerge and thus new proprietary to be commercialized around the open technology.	Data standards for industry and academic partners to enable data sharing and re-use but also integration within information systems inside organizations.
Arbitrage by non-industry firm	Arbitrage and orchestration of exchanges by the academic partner, as a trusted party to govern the exchanges amongst stakeholders and guarantee, in some cases, re-use for that do not damage competitive advantage.	Arbitrage by the academic partner between data holders and data (re)users putting in place protocols that provide trust to commercial partners and third-parties.
Learning gains through interaction	High engagement and close interactions by the academic partner for companies to extract learning gains from the interaction.	High engagement and close interactions with the academic partners for learning gains and to increase possibilities of research data re-use.

	Open scientific hardware	Open Scientific Data
Competitive advantage towards contributors	Favouring industrial contributors towards free-riders through branding and reputational gains.	Time lag and premium services for contributors.

Finally, the study team invites readers to go through the different case studies in order to get more insights about the different mechanisms, boundary conditions and contextual characteristics that made the case studies be selected in the first place as insightful examples of successful open science implementations.

4.3 Societal impact

Open science certainly has a visible impact on society as an indirect effect of the scientific and industrial impacts analysed above. Enhanced, faster scientific discovery – such as those identified through Open Targets – benefit all parts of society. Early detection of side effects – provided by YODA – lowers the costs of development and accelerates the detection of side effects.

Secondly, industrial innovation is a determinant of growth and employment. The development of new products, such as White Rabbit, has created a new market opportunity, and increased the productivity of the companies adopting these innovations – from finance to telecom.

Beside these well-known benefits towards society through scientific and industrial impact, there are other relevant aspects to be considered. Firstly, open science can contribute to greater public understanding of science. Open science has the potential to enhance public engagement and the understanding of science principles and practice by raising awareness, pro-active participation and citizens' direct contribution to research (Boulton et al., 2012; Kowalczyk & Shankar, 2010). The openness of the scientific process fosters new actors' access to the research process (Nowotny, 2001). The increasing importance of citizen science is a force for disruptive change, engaging amateurs and the general public in science. Those inclusive and participatory approaches to boost human capacity and capability from professionals, amateurs, volunteers and citizens to assist in collecting, curating and preserving the growing scientific record (Lyon, 2009).

Citizen science is also a method for formal and informal science education and public understanding of science, which should not be underestimated (Christian et al., n/a.). Greater citizen engagement can lead to greater participation in scientific experiments and data collection (OECD, 2015; Boulton et al., 2012; Whyte & Pryor, 2010). Making research data publicly available can advance people's understanding of science and citizen science initiatives (Kowalczyk & Shankar, 2010). The Global Mosquito Alert Project has brought together key citizen science monitoring initiatives to work together to help understand the geographic spread of mosquitos who are capable of carrying disease, in turn helping health professionals and citizens to manage any potential risks.[6] Zooniverse, a world-leading

[6] Jon Switters, David Osimo, Citizen Science in the Surveillance and Monitoring of Mosquito-Borne Diseases, Open Science Monitor Case Study (Brussels: European Commission, 2019)

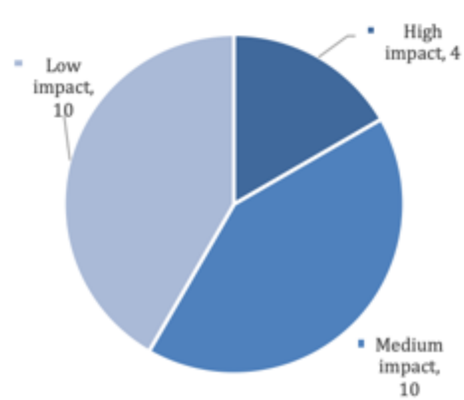
platform for online crowdsourced research, has more than 100 active projects (in 2019) where volunteers participate in crowdsourced scientific research, classifying large datasets across multiple knowledge domains (such astronomy, ecology, cell biology, humanities, climate science, etc., with significant overlap between areas). The Zooniverse research brings forward new discoveries, datasets useful to the wider research community, and publications. Hypothes.is developed a tool to enable anyone to annotate anywhere, and opened up and standardised the annotation's technologies and practices. It aimed to make it easy for users to leave annotations across the entire web. And, the open-source nature of the platform allows anyone (developers, publishers, academic institutions, researchers, and individuals) who use it to create their own annotation reader or writer. The nature of the tool also facilitates the collection of annotations in a structured way, allowing to build the overall picture of the information knowledge.

On the other hand, increased research data sharing can constitute a challenge to societal values, mainly by increasing the risk for infringement of data protection. There is a fundamental tension between the scientific value of data built through correlation of disparate datasets variables and the need for preserving anonymity. Several studies have shown that personal data security cannot be safeguarded by anonymisation. Companies that release data openly for research purposes have encountered problems related to the re-identifiability of subjects. Datasets that contain information on individuals support inferences regarding the probability of other information about people (Denning & Schlörer, 1980; Lyon, 2009). Conversely, the recent emphasis around the implementation of the General Data Protection Regulation is forcing scientists and research institutes to adopt more stringent data management practices that can be consistent with increased data sharing: for instance, by limiting storage on own computers and requiring scientists to use dedicated law-compliant repositories (Stark, 2019). This could turn out to force the adoption of appropriate data management strategies (rather the informal ones illustrated in section 2.2) that could ultimately lead to wider reusability of data.

4.3.1 Cross analysis of cases

When performing a cross-analysis of the case studies, few of them show high impact on society (see Figure 1).

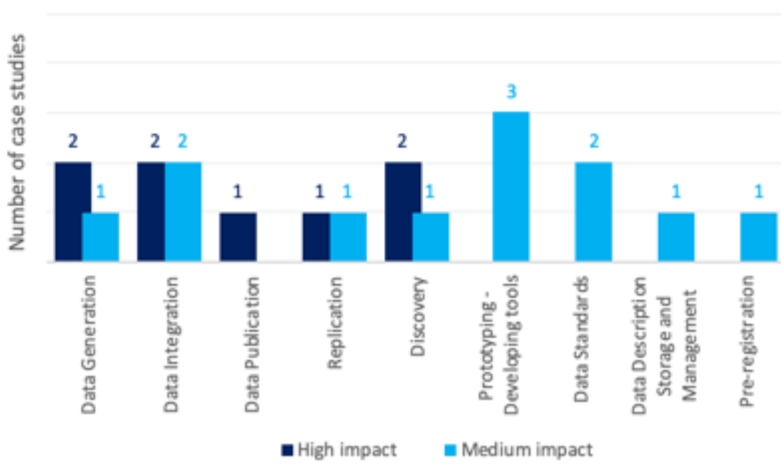
Figure 34: Impact on society of the analysed case studies



According to the assessment of the cases, there are only four case studies (of the total of 24 analysed) with high impact on society: Open Targets, Yoda, Global Biodiversity Information Facility (GBIF) and Zooniverse. For other 10 cases, they show a medium level of impact on society: White Rabbit, Pistoia Alliance, REANA, OSH licences, As predicted, Two use cases, Global Alliance, GOFAIR, Hypothes.is and Mosquito alert. The rest of 10 remaining case studies show a rather low impact on society.

From research lifecycle side, the cases with high and medium impact are found in data generation (Open Targets, Mosquito alert), data integration (Open Targets, YODA, Two use cases), data publication (YODA), replication (YODA, REANA), discovery (Global Biodiversity Information Facility, Hypothes.is), prototyping and developing tools (White rabbit, Open Hardware licences and Hypothes.is), data description storage and management (GOFAIR), pre-registration (As predicted) and data standards (Pistoia Alliance and Global Alliance).

Figure 35: Case studies with high and medium impact on society, by research phases



It is important to mention that the case studies with high and medium impact on society are not restricted to only one open science category (open data, open access, open collaboration or citizen science). Four out of the 14 case studies with high and medium impact on society are classified in more than one open science category (see Table 8).

Table 7: Case studies with high and medium impact on society, by open science category

	Open science category	
	Classified in one area	Classified in more than one area*
Open collaboration	2 (*14.3%*)	5 (*27.8%*)
Open data	7 (*50.0%*)	9 (*50.0%*)
Open access	-	1 (5.*6%*)
Citizen science	1 (*7.1%*)	3 (16.*7%*)
Total case studies (considered)	**14**	**18**

Note: *The case studies with high and medium impact on society classified in more than one open science category are counted multiple times.

Open data is most frequent category the case studies are associated to, regardless of single or multiple coverage – half of the case studies include the open data principle. It is followed by open collaboration and citizen science. When it comes to open access, only one case study includes it, together with open collaboration (*As predicted*).

5 Policy analysis

This section provides the integrated policy analysis of the study. The first part presents the overview provided by the data contained in the Sherpa database, about funders and journals policies on open access and open data (Sherpa-Juliet and Sherpa-Romeo database). The second part presents the in depth qualitative cross analysis of case studies.

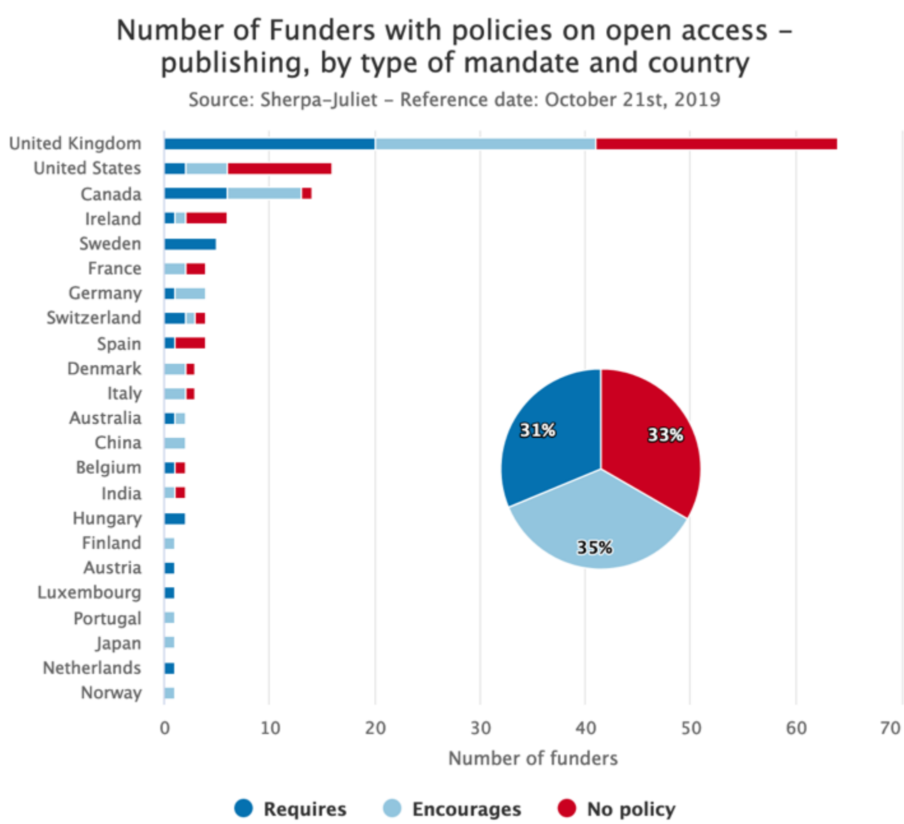
5.1 Statistical overview

The statistical overview is divided in open access to publications and open data. It includes data on funders and journal policies.

5.1.1 Open access policies

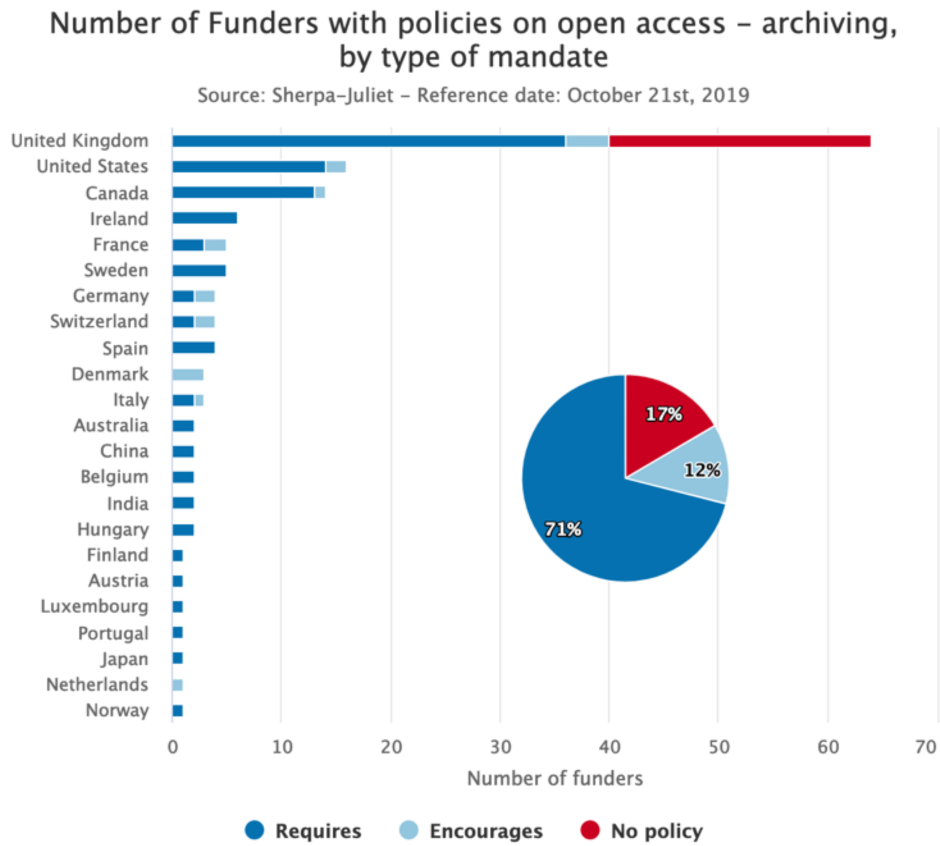
One of the main factors affecting open access to publication is the **mandates from the research funders**. Data shows that open access publication is enforced by research funders to a much lesser extent than open access archiving - out of 145 institutions identified, 33% has no open access policy in place, whereas 35% encourages it and only 31% requires it.

Figure 36: Funder's policies on open acess publishing by country



Concerning number of research funders regarding mandate on open access archiving the vast majority (71%) requires open access archiving and further 12% encourages it, only 17% of institutions (all from UK) do not have any policy regarding open access archiving.

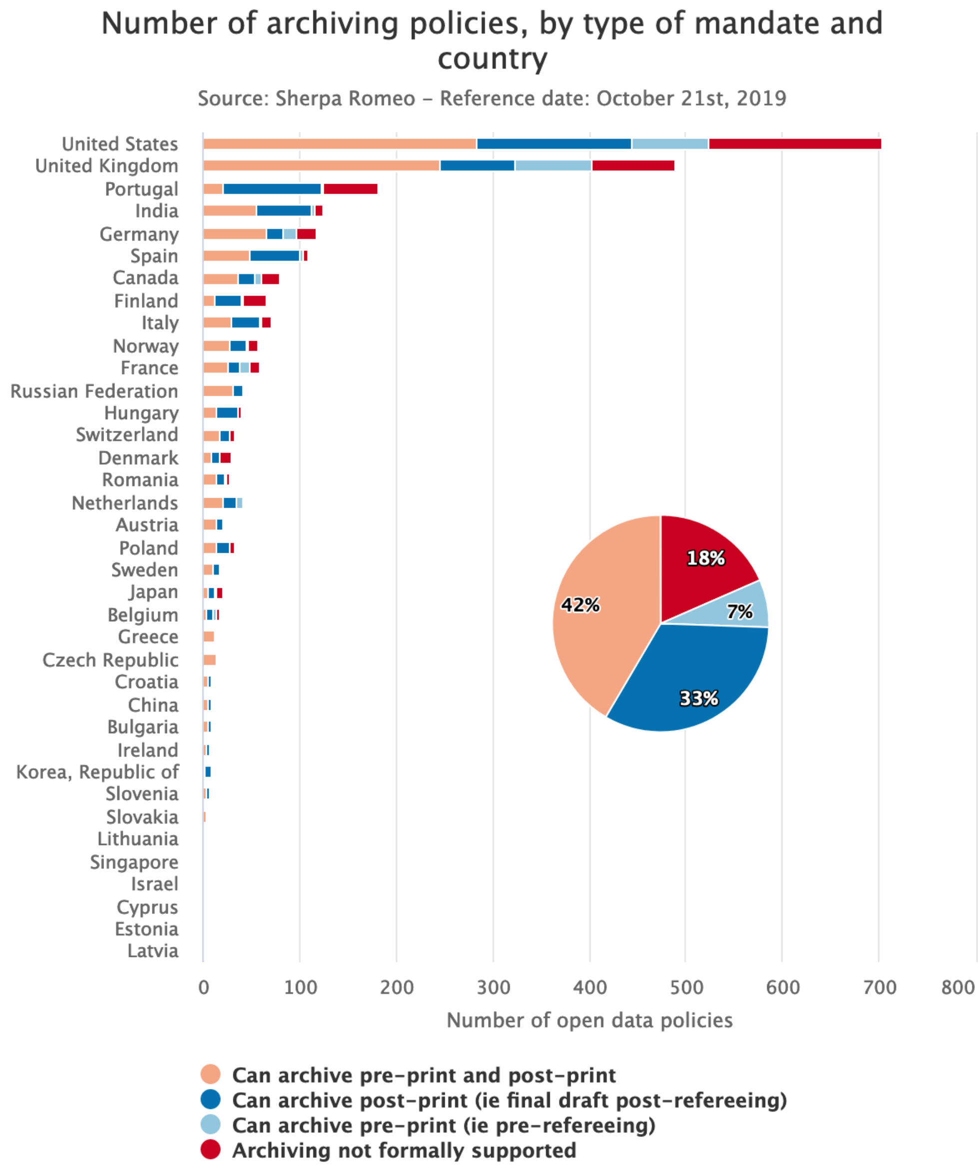
Figure 37: Funders' policies on open access archiving



Policies by **research journals** are also important in setting the boundaries for researchers' choices. The vast majority supports some sort of archiving (82%): 42% supports archiving of pre-print and post-print, 33% can archive post-print and further 7% can archive pre-print. Only in 18% of journals archiving is not formally supported.

US research journals stand for the majority of those journals where archiving is not supported, followed by United Kingdom and Portugal. These are also the countries with the largest amount of the research journals. However, from all top five countries (India and Germany as top 4 and top 5) only in Portugal more research journals support post-print only rather than post-print and pre-print together, the latter being and overall predominant trend in most of the countries.

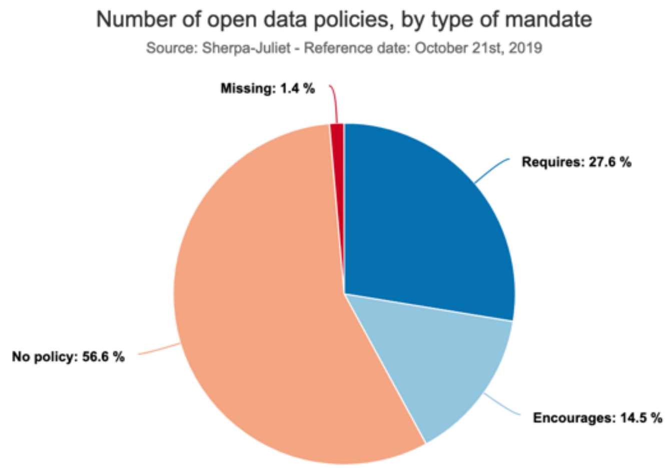
Figure 38: Archiving polices by type of mandate



5.1.2 Open data policies

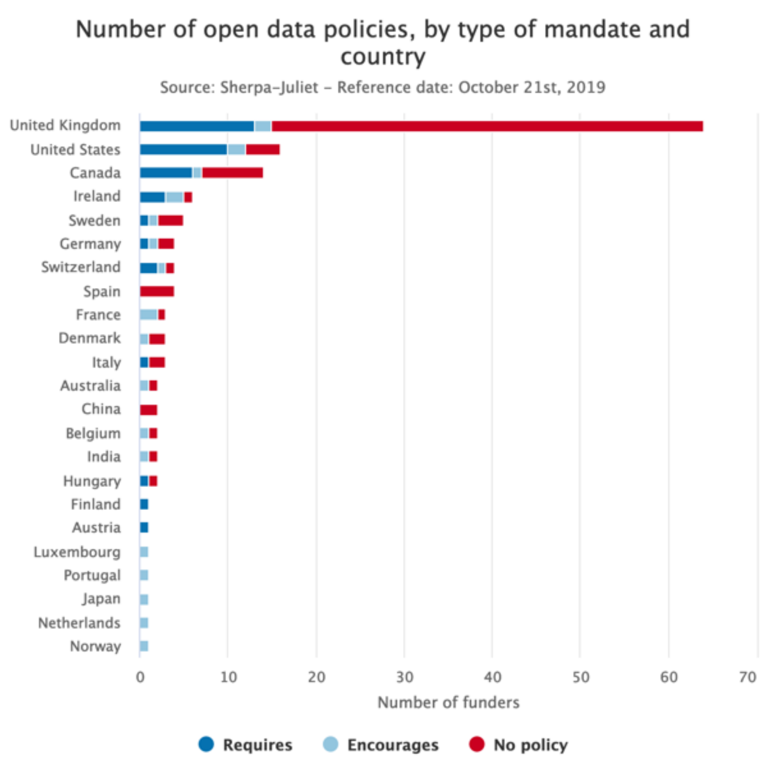
Yet, concerning **policies of research funders,** more than a half (56.6%) do not have any open data policy in relations to data archiving and only 27.6% requires it.

Figure 39: Open data policies by type of mandate



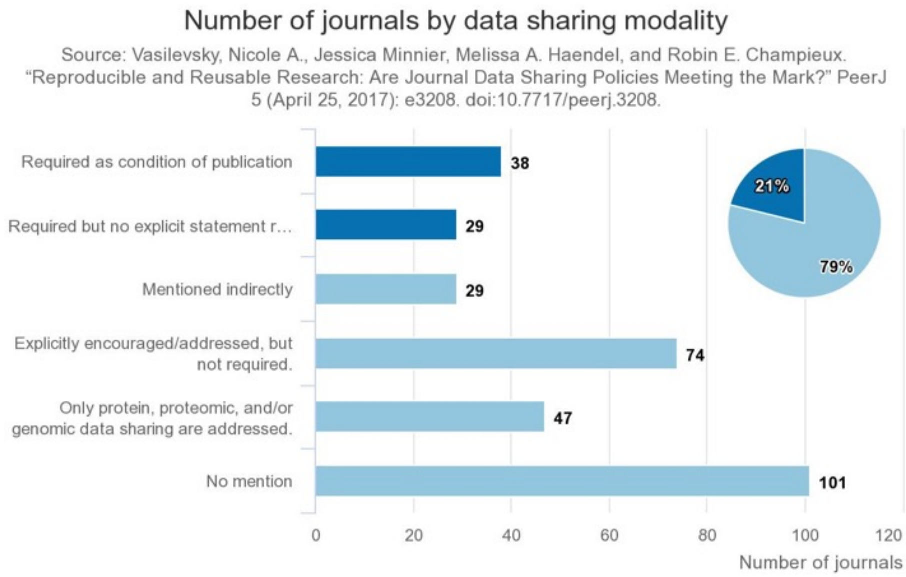
As in the case of open access, these statistics are largely influenced by United Kingdom, a country with the highest number of research funders mapped in the Sherpa-Juliet database. Almost 76% out of 64 research funding institutions in the UK does not have open data policy. Furthermore, in many other countries, research funding institutions either do not have open data policy (e.g. Spain, China, Luxembourg) or there are more institutions that do not have an open data policy than those that encourages it or requires it (e.g. France, Belgium, Denmark, Italy, Sweden).

Figure 40: Open data policies by country



Regarding **journals** and its data sharing modality, according to the latest available data, only 21% of journals requires data sharing and 32% do not addresses data sharing at all. This indicator is a one-off analysis carried out in 2017.

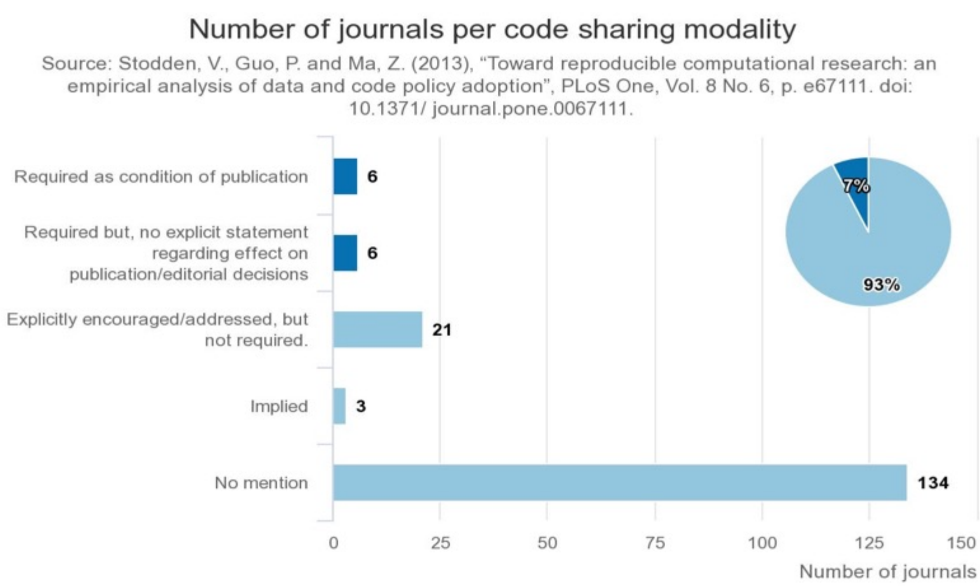
Figure 41: Journals by data sharing modality



5.1.3 Other policies

There is little data on journals' policies concerning open code. Data from 2013 show that journals´ open code policy is almost non-existent. Out of 170 journals analysed 79% of them did not address code sharing. In fact, codes sharing was explicitly required only in case of 12 journals (see Figure below).

Figure 42: Journals by code sharing modality



5.2 In depth cross-analysis of policy cases

This cross-analysis presents the systematic investigation of the national implementation of open science policies in the Netherlands (NL) and Finland (FI), and open research policies in the United Kingdom (UK). Firstly, the purpose of this report is to facilitate discussions on the implications and the policy recommendations made in the European open science policy context. Secondly, this study synthesises the national policies conducted by the Open Science Monitor (OSM) thus far, by analysing the main bottlenecks and the policy measures gathered across the three countries. Finally, the result of this evaluation is the first step in a broader analytical task of monitoring open science development in Europe – what open science strategy (at European level) is needed, why, and to which extent? From there, this study will conclude with potential insights that can guide future analysis on open science.

5.2.1 Objectives

Across the national policy case studies in the NL, FI and UK, two key drivers are guiding the narrative of the policies: (1) the moral imperative that publicly funded research generates knowledge that belongs to everyone, thus must be shared to and accessed by the public immediately, and (2) the economic gains of easily and widely accessible research, which generates productivity and competitiveness in their respective national research and innovation systems. Since all three countries are keen to be known as the leaders in open science and research, Table 8 presents a comparable overview of their scientific, societal and industrial drivers within each of the national policy context.

Table 8: Overview of scientific, societal and industrial drivers across NL and FI Open Science and UK Open Research Policies

Driver	NL	FI	UK
Science	Open science improves the transparency, verifiability, efficiency, reproducibility, and sustainability of **research processes**	Openness will promote transparency, make the **research process** more effective and allow researchers to deal with complex, multidisciplinary questions	Open research improves efficiency in the **research process**, transparency, accountability, and public engagement with research
Industry	Open science increases innovative capacity as individuals benefit more readily from publicly available information and use it in combination with their own knowledge and experiences **to**	Open science leads to surprising discoveries and creative insights - fostering the research system in Finland towards **better competitiveness and higher transparency and innovation**	Open research creates closer linkages between research and innovation, benefits products and services, encourages the **commercialisation of research** and

Driver	NL	FI	UK
	develop novel products and processes		stimulates economic growth
Society	**Everyone outside the scientific community may benefit from open science** because they can readily access and use scientific information	The circulation and promotion of research data outside academia contributes to the greater societal impact by **increasing scientific literacy among citizens.** Transparency **increases the credibility of science for the citizens**	Open research **promotes public understanding of science**, evidence-based practices, and citizen-science initiatives

Source: OSM case studies

Similar drivers were identified in all three cases, although the open science policies were introduced in different years for each of the countries. For the Netherlands, it was 2016 as led by the Dutch Ministry of Education, Culture and Science, and for Finland it was 2014, as initiated by the Finnish Ministry of Education and Culture. Meanwhile in the UK, it was termed open research and declared by the Minister of State for Universities, Science, Research and Innovation in 2013. The moral and economic drivers for open science and open research in these three countries were expressed clearly, which hardly leave any room for objection.

5.2.2 Provisions

Public policy plays an important role in influencing researcher behaviour. In recent years, several institutional adjustments to the general governance framework of science have taken place:

- The introduction of competitive funding mechanisms, as opposed to bulk funding in the European higher education sector. Publications output and impact factors became common performance indicators for quality-related research in higher education institutions. This has increased competition and reinforced the importance of assessment metrics in the context of funding distribution.
- The increased policy emphasis on innovation, to boost the competitiveness of European industry. Industrial impact, characterised by collaboration with private companies, generation of patent and intellectual property rights (IPR), were implemented into the broader European Commission policies, to better convert scientific research into

improved goods, services and processes for the market for the welfare and wellbeing of society.

These two science policy trends certainly create a tension with the specific open science policy goals of openness and collaboration, increasingly adopted by funding agencies and national institutions. Table 9 presents the observed tension through the evidence gathered from the national open science policies of NL and FI, as well as open research policies in the UK in three key features: open access, open data, and research evaluation and reward systems.

Table 9: Provisions addressed in the NL and FI Open Science and UK Open Research Policies

Feature	NL	FI	UK
Open access	Hybrid journals will no longer be paid for. Challenge to turn Green OA route to an accepted means of publishing.	A lack of common procedures on open publishing. Low publication in OA Journals and the costs for OA publishing remains expensive for most of the evaluated publishers.	70% of OA publishing is going to hybrid journals. The average APC for hybrid payment by UK universities were increased by 19%, between 2013 and 2016.
Open data	All universities had their own online repository, which is linked to a national portal – NARCIS. Although there is a slight increase of open datasets over the years (except for 2011 and 2013), most datasets remained closed or restricted. The National Coordination Point Research Data Management (LCRDM) will conduct a study to establish the qualitative and quantitative need for data stewards and research software engineers, the content of their role, the training required, and the rewarding of their work.	Strongly data-driven with a focus on the development of research data architecture: management of ownership, metadata, and IPR. The initiative was planned until 2017, with no further strategy or plan developed for the following years.	Patchy open data provisions: research councils are either (1) silent on the matter, (2) provide direct support for the work of data centres and services, or (3) rely on university or third-party repositories to preserve and provide access to research data.
Research evaluation and reward systems	Negative evidence for open science reward in performance agreements between the ministry and universities, and tenure track policy between universities and individual researchers.	The evaluation framework for openness was developed to incentivise research organisations with respect to their policies on and implementation of open science practices.	OA publications, resources and infrastructure that support open research are considered in the quality of research outputs and research environment of the

Feature	NL	FI	UK
			Research Excellence Framework.

Source: OSM case studies

Finally, the General Data Protection Regulation (GDPR) was enforced by the EU on the 25th of May 2018. While the GDPR does strengthen data protection and privacy of EU persons, it also complicates the proprietary aspect of data as institutions and individuals are still trying to understand how changes in privacy laws will impact their work. GDPR thus creates further tension between open data and privacy and security, as illustrated in section 4.3.

5.2.3 Lessons learnt

The UK is leading in the number of total OA publications, followed by NL and FI. Given that the UK has the highest number of research funder mandates concerning OA publication, not to mention that the Research Excellence Framework (REF) and research funding depends on it, it is unsurprising that the UK is leading in Green OA publications. Contrary to the preference for Gold OA as announced by the UK Science Minister, David Willets in 2013, the Gold route is still far from being a desirable and acceptable means of publishing for UK researchers. In the Netherlands, a strong focus in Green OA was observed in general, since all universities had their own online repository, which makes it easier for researchers to self-archive their publications.

The current policies in NL, FI and UK demonstrate converging policies, with minor differences in the drivers on open science. The focus of all three countries are mainly on open access publishing from the top-down, where funders and governing bodies are leading progress in open science instead of bottom-up, research communities led initiatives.

It is worth noting that the funding bodies of all three countries in the present report, are members of 'cOAlition S', an international consortium of research funders to accelerate immediate Open Access publishing. Clearly, their involvement in Plan S demonstrated the country's commitment in realising the moral imperative and economic benefits as outlined in the 'Drivers' section.

The United Kingdom is seen as the strongest in open science and is in fact, dominated by impact-oriented practices that are sanctioned and imposed by research funders. This has resulted in a divergence in open access ambitions and actual publication practices, with a rise in publications in hybrid journals, and an immediate decrease in gold OA publications. The cost of open science opening continues to take centre stage in the open science policy arena, while the key issues of open data, research evaluation and reward system are hardly addressed by policymakers in all three countries.

6 Policy conclusions

The large research effort leads to many possible policy reflections, in different areas. For simplicity, we divide them in those concerning open science policy, and those concerning future monitoring efforts.

1.1. Conclusions on the future of open science

1. Data shows clearly that open science is there to stay.

All trends clearly indicate progress: from open access to publications (considering the impact of embargo period on statistics) to open data (with limited progress) to open API, open hardware and citizen science. Every indicator and every trend show progress, even if slow.

2. Open science is a wide-ranging set of trends: it is more than open access to publication and open research data.

This perhaps obvious statement needs to be repeated. In the policy debate, it appears that open access attracts all the attention and the "expenditure" in terms of policy capital. After that, the topic of open research data in policy terms is emerging strongly, led mostly by the policy debate about the data economy. Yet all the other trends are growing strongly despite the lack of policy attention. At a time when Microsoft acquires Github, the largest repository of open source software, for about 7 billion euros, few journals, funders and regulators have yet introduced policies concerning open scientific software, and the developments are mostly bottom up.

3. The progress detected by the indicators does not justify any complacency.

Data shows a slow progress. For instance, data sharing among researchers show limited progress in the two years measured. Progress on open access to publications in unequal, with a decline in green open access that we currently attribute to embargoes – but that needs to be monitored in the future. Overall, it is clear that open science is not the default option – but rather chosen by a minority of scientists both in terms of open access and open research data. In other words, open science requires continuous policy attention.

4. The policy attention needs to be focussed in particular on getting the incentive systems right.

Open science requires changes by all stakeholders involved and can't be achieved only through regulation. It is clear that we see a proliferation of policy initiatives at national and European level, but when we look at behavioural changes by scientists the importance of funders' policies is limited – it is much more important to act through the community of peers. Ultimately, open science requires a substantial effort of governance, a continuous tilting of the playing field to ensure optimal results. More concretely, initiatives such as the European Open Science Cloud require a continuous effort in governance even AFTER the launch. The launch is the beginning, not the end of the policy process. All the successful, large scale open science initiatives described in the case studies show that the only way to scale up is through continuous adaption and experimentation of new modalities to involved different stakeholders. The reason for this is clear:

5. Open science does not appear as an ideological radical choice, but as a rational decision based on the inherent advantages.

Most large-scale projects have balanced closed and open methods, depending on the specific challenges and opportunities. Some methods work under specific conditions and depending on the stakeholders' interests. There is no one size fits all approach to open science. It is clear that both scientists and companies are using different degrees of openness depending on the circumstances. Researchers share data mostly within their research group or projects partners, rather than openly. Datasets in repositories have different levels of openness. Research centres develop new licences to accommodate for the needs of different stakeholders.

6. Scientific progress is the ultimate goal – and the impact is visible.

What is clear from the cases analysed is that the achievement of scientific progress is the main priority, and the openness varies in order to maximize that goal. Equally in the corporate sector, getting new products faster to market is a priority with respect to openness in itself. In other words, when open science scales, it is a means not a goal.

Some of the highest profile scientific and commercial endeavour have incorporated elements of open science. And it's not only basic science where open practices are consolidated. Open science practices are now used in commercial endeavours, such as the development of high precision time measurement or the discovery of new pharmaceutical products. Despite the attention on open access to publications, it is in open research data and open collaboration that we see greater industrial impact. Open practices and commercial goals can go hand in hand, but only under specific conditions. There is a trade-off, not an incompatibility.

7. Open science policy should look beyond open science.

What is clear is that open science is part of a wider trend towards data driven and large-scale collaboration, for instance through the adoption of APIs, the diffusion of data portability practices, servitisation and more. Policies such as the data economy, the digital agenda and industrial policy should be aligned with the open science agenda.

Ultimately, open science policy should not be driven by ideology but by concrete needs. Only in this way can the participation of all relevant stakeholders be ensured. To achieve this, the community building work that the EC has implemented so far is a necessary foundation, but it is only the beginning. Open science practices require systemic change that takes several years to take place. And good metrics are a fundamental part of this effort.

1.2. Conclusions on the future of the Open Science Monitor

1. The existence of an Open Science Monitor is necessary to successful policy implementation

The transition to open science is a complex, multi-stakeholder organic transition, not a top down decision-making process. In such a context, monitoring is even more important, because it helps all stakeholders build a common assessment of the situation and to align everyone's efforts. Obviously, a monitor does not *per se* generate consensus and common action, but its absence makes achieving those goals impossible. National case studies such as the Netherlands showed that the absence of monitoring mechanisms for specific policy areas such as open data was a major barrier to progress. Monitoring open science is not a simple straightforward task: it requires clear methodological choices and robust methods. As such, it cannot rely purely on third party metrics, which often differ a lot in terms of coverage and definitions, including between member states. Moreover, the sensitivity of the issue, as shown by the conflict about the existing Open Science Monitor, requires strong central knowledge presidium to justify and defend methodological decisions. Put simply, European open science policy requires a European open science monitor.

2. The next Open Science Monitor should reduce the robustness gap between open access and other trends

While open science is recognized as a multifaceted trends, some trends are much better covered than others when it comes to indicators. There are major differences between the availability and robustness of data on open access to publications, open data and open collaboration. Data about open access to publication are easily available, although there are constant discussions about the modalities of measurement. Indicators about open data sharing have been provided through ad hoc survey, in view of its relevance. But for the other trends we are at loss of reliable indicator and we can only use proxies, mainly related to the adoption of existing services or existing collections that are necessarily partial. The other trends are the "Cinderella" of open science in terms of data availability. For the future, we need therefore to increase the attention to measuring reliably the different trends. This could be done in the first instance by extending the survey from open research data to other important trends, starting from open scientific software; secondly, through ad hoc sample based bibliometric analysis to assess the availability of data and code associated to articles.

3. Multiple sources and methods should continue to be used

Indicators produced by the monitor should rely on the widest set of sources and methods. The only criterion for choice should remain reliability and validity of indicators. A good example is how the present Monitor extended the sources for open access metrics to include Unpaywall alongside Scopus. This integration of different sources allowed a more fine-grained, timely and extended coverage. In addition to data integration, data triangulation is also necessary to reinforce the legitimacy of the findings. The present Monitor compared the results with other sources, such as Web of Science for open access and the "State of open data" survey by Figshare, recognizing explicitly the similarities and contradictions. In other words, the future Open Science Monitor should be as broad as possible and provide a critical meta-analysis of other surveys and studies.

4. The key focus should be adoption

Open science is maturing, and so should do the indicators. Supply side indicators such as the mere availability of tools, the number of data repositories, or the existence of policies should be secondary to the other priority: measuring the adoption of open science practices by all stakeholders. Adoption should include not only the number or percentage of users, but the intensity or frequency of adoption. In other words, the indicators should not only measure the percentage of scientists that publish in open access, share data and code, but whether they do it systematically on all papers or occasionally. When measuring adoption, bibliometric indicators appear more effective than surveys, insofar they allow for objectives measurement of behaviour rather than self-reported.

5. Bibliometrics first, survey second

There is little doubt that measuring adoption is better through observation rather than self-reporting. Asking researchers whether they adopt open science practices is more prone to biased results than gathering bibliometric data about whether they effective publish data and code alongside an article. Indeed, it is more difficult to measure data and code sharing, than it is to measure publication in open access, but it is still doable. For instance, sample based bibliometric have been used by many individual studies – while open access to publications is measured on the universe of publications. Finally, survey still have an important role to play, in particular when it comes to assessing drivers and barriers. Instead, adoption metrics from specific services are of limited value: indicators such as the number of citizen science or open hardware projects can provide a suggestion of growth in usage, but do not allow for assessing adoption growth in the overall scientific community as a whole. Finally, case studies remain unique in providing insight about a better understanding of how open science works and the impact it achieves. They are very important in this early and fast stage of evolution. In summary, we suggest moving towards greater usage of bibliometric indicators for assessing adoption, and reducing metrics provided by specific services, while dedicating surveys and case studies specifically to drivers and impacts, as shown in the table below.

Table 10: High level methodological approach, past (x) and future (grey shade)

	Drivers	Open access	Open data	Open collaboration	Impact
Bibliometric (universe)		x			
Bibliometric (sample)					
Survey	x		x		x
Service metrics		x	x	x	
Cases				x	x

6. The next open science monitor should be as open as possible – and as closed as necessary

There have been some strong criticisms by part of the open science community towards the present Monitor because of the usage of proprietary data. This has led to a serious internal review of the methodology and significant improvement, including an expansion of the original sources. The methodology was then discussed in a high-profile expert workshop. The results confirmed the choice to use some proprietary data, because they provide high quality additional information (namely about affiliation). The final choice should be pragmatic and driven by the objective, which is to provide reliable high-quality data points in support to policy. As of today, there is still room for proprietary data in this context. Using

only open data would lead to lower quality output. Still, there are important options to be used when using proprietary data, such as allowing researchers to access the data on demand for replication and research purposes. And this is exactly what has been done.
On a similar note, the process should continue to be radically open, but this does not mean that decisions should be taken by the community. There is a need for strong central leadership in making the appropriate methodological choices. As in the case of the present monitor, the open process included allowing everyone to post comments on the methodology (visible to everyone), and provided systematic feedback to each individual comment about the choices made. The choices are made centrally, but after open consultation, expert review and providing clear feedback on the rationale behind refusing or accepting suggestions. In other words, both the data and the process should be as open as possible and as closed as necessary – but always transparent. The open science monitor should embrace the spirit of open science – but as we have seen in the analysis, this does not mean excluding at all cost any form of closed process and output. Just as open science, the open science monitor requires careful governance of trade-offs.

7. The scope of the work, and the community addressed, should be science as a whole, not open science

The open science monitor has to cover the developments across science – not merely tracking those who practice open science. It should not provide indicators only about the adopters but observe the transition in the overall scientific process. This is why it should no longer prioritize data provided by open science services, such as the number of users of a services, or the number of projects in an open repository. It should instead focus more heavily on measuring the adoption of open science practices by the whole scientific community – such as the percentage of articles or researchers that include the adoption of open science trends.
Equally, the users and the stakeholders of the Monitor are not the open science community, but the scientific community as a whole. This is as difficult as it is important, because the open science community is very active and dedicated – and vocal on social media. Ultimately, the legitimacy of the Monitor comes from the attention and impact in the overall scientific community, not from the approval of open science advocates. The Open Science Monitor should design its services around the needs of the scientific community – and proactively reach out through articles, presentations and communication as a priority, rather than engage mainly with the vocal open science community.

7 References

Andreoli-Versbach, P. and Mueller-Langer, F. (2014) Open access to data: An ideal professed but not practised. Research Policy, 43(9), pp.1621-1633.

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., and Roberge, G. Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013, Report to the European Commission, April 2014

Arnold, E. (2005). What the Evaluation Record tells us about Framework Programme Performance. Management, (January).

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z. (2007) Dbpedia: A nucleus for a web of open data. The semantic web (pp. 722-735). Springer Berlin Heidelberg.

Baker, M (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604): 452–454. DOI: https://doi. org/10.1038/533452a

Bannister, F. (2007). The curse of the benchmark: an assessment of the validity and value of e-government comparisons. International Review of Administrative Sciences, 73(2), 171–188. http://doi.org/10.1177/0020852307077959

Berghmans, S., Cousijn, H., Deakin, G., Meijer, I., Mulligan, A., Plume, A., … Waltman, L. (2017). Open data : The researcher perspective, 48 p. http://doi.org/10.17632/bwrnfb4bvh.1

Blumenthal, D., E. G. Campbell, M. Gokhale, R. Yucel, B. Clarridge, S. Hilgartner, and N. A. Holtzman (2006). Data Withholding in Genetics and the Other Life Sciences: Prevalences and Predictors. Academic Medicine, 81(2), 137-145.

Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Laurie, G., ... and Walport, M. (2012) Science as an open enterprise. The Royal Society.

Bücheler, T. and Sieg, J. H. (2011). Understanding science 2.0: crowdsourcing and open innovation in the scientific method. Procedia Computer Science, 7, 327-329.

Burgelman, J. C., Osimo, D. and Bogdanowicz, M. (2010) Science 2.0 (change will happen….). First Monday, 15(7).

Burgelman, J.-C., Osimo, D., & Bogdanowicz, M. (2010). Science 2.0 (change will happen…). First Monday, 15(7).

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A. and Blumenthal, D. (2002) Data withholding in academic genetics: evidence from a national survey. Jama, 287(4), 473-480.

Carlson, S and Anderson, B 2007 What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. Journal of Computer-Mediated Communication, 12(2): 635–651. DOI: https://doi.org/10.1111/ j.1083-6101.2007.00342.x

Carrara, W., Nieuwenhuis, M., & Vollers, H. (2016). Open Data Maturity in Europe 2016. Brussels.

Chokshi, D. A., Parker, M. and Kwiatkowski, D. P. (2006). Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. Bulletin of the World Health Organization, 84(5), 382-387.

CODATA-ICSTI Task Group on Data Citation Standards Practices 2013 Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. Data Science Journal, 12: CIDCR1–CIDCR75. DOI: https://doi.org/10.2481/dsj.OSOM13-043

Collison, P., & Nielsen, M. (2018, November 16). Science Is Getting Less Bang for Its Buck.

Retrieved from https://www.theatlantic.com/science/archive/2018/11/diminishing-returns-science/575665/
Costas, R., Zahedi, Z., Wouters, P. (2015a). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. Journal of the Association for Information Science and Technology, 66(10), 2003–2019. https://doi.org/10.1002/asi
Costas, R., Zahedi, Z., Wouters, P. (2015b). The thematic orientation of publications mentioned on social media: large-scale disciplinary comparison of social media metrics with citatios. Aslib Journal of Information Management, 67(3), 260–288. https://doi.org/10.1108/AJIM-12-2014-0173
Crouzier, T. (2015). Science Ecosystem 2.0: how will change occur. Acta biomaterialia, 20, 51-59.
Dasgupta, P. and David, P. A. (1994) Toward a new economics of science. Research Policy, 23, 487–521.
David, P. A. (2004) Can "open science" be protected from the evolving regime of IPR protections?. Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft, 9-34.
David, P. A. (2004) Patronage, Reputation, and Common Agency Contracting in the Scientific Revolution: From Keeping’Nature’s Secrets’ to the Institutionalization of’Open Science’. SIEPR Policy paper No. 03-039.
Denning, D. E. and Schlörer, J. (1980). A fast procedure for finding a tracker in a statistical database. ACM Transactions on Database Systems (TODS), 5(1), 88-102.
Dosi, G., Llerena, P. and Labini, M. S. (2006) The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called ‘European Paradox’. Research policy, 35(10), 1450-1464.
Drummond, C 2009 Replicability is not Reproducibility: Nor is it Good Science. In: ICML ’09 Proceedings of the 26th Annual International Conference on Machine Learning (The fourth workshop on evaluation methods for machine learning). Montreal, Quebec, Canada. Retrieved from: http://cogprints.org/7691/.
Eisenhardt, K. M. (1989). Building Theories from Case Study Research. Academy of Management Review, 14(4), 532–550. http://doi.org/10.5465/AMR.1989.4308385
Eisenhardt, K. M., & Graebner, M. E. (2014). THEORY BUILDING FROM CASES : OPPORTUNITIES AND CHALLENGES diverse. Academy of Management Journal, 50(1), 25–32. http://doi.org/Article
European Commission (2009) Preparing Europe for a new Renaissance: A strategic view of the European research area. First report of the European Research Area Board — 2009. Brussels: European Commission.
European Commission (2010) Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth: Communication from the Commission. Brussels: European Commission.
European Commission (2012) Recommendation on access to and preservation of scientific information. Brussels: European Commission
European Commission (2014) Innovation Union Scoreboards. Brussels: European Commission DG Enterprise
European Commission Expert Group on FAIR Data (November, 2018). Turning Fair into Reality - Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Commission Directorate General for Research and Innovation.

European Commission High Level Expert Group on Scientific Data 2010 Riding the wave: How Europe can gain from the rising tide of scientific data (Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission). European Union. Retrieved from: https://www. fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data.

Fang, F. C., & Casadevall, A. (2015). Competitive Science: Is Competition Ruining Science? Infection and Immunity , 83(4), 1229–1233. http://doi.org/10.1128/IAI.02939-14

Fecher, B., Friesike, S. and Hebing, M. (2015) What drives academic data sharing? PLoS One, 10(2), e0118053.

Feller, J. and Fitzgerald, B. (2002). Understanding open source software development (pp. 143-159). London: Addison-Wesley.

Franceschet, M. and Costantini, A. (2010) The effect of scholar collaboration on impact and quality of academic papers. Journal of informetrics, 4(4), 540-553.

Friesike, S., Widenmayer, B., Gassmann, O. and Schildhauer, T. (2015) Opening science: towards an agenda of open science in academia and industry. The Journal of Technology Transfer, 40(4), pp.581-601.

Furman, J. L., Porter, M. E. and Stern, S. (2002) The determinants of national innovative capacity. Research policy, 31(6), 899-933.

Gittelman, M. and Kogut, B. (2003) Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. Management Science 49, 366–382.

Goodman, A A, Pepe, A, Blocker, A W, Borgman, C L, Cranmer, K, Crosas, M, et al. 2014 Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Computational Biology, 10(4): e1003542. DOI: https://doi.org/10.1371/journal.pcbi.1003542

Guston, D. H. (2000) Retiring the social contract for science. Issues in Science and Technology, 16(4), 32-36

Haeussler, C. (2011) Information-sharing in academia and the industry: A comparative study. Research Policy, 40(1), pp.105-122.

Haeussler, C., Jiang, L., Thursby, J. and Thursby, M. (2014) Specific and general information sharing among competing academic researchers. Research policy, 43(3), pp.465-475.

Haustein, S., Costas, R., Larivière, V. (2015). Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. PloS ONE. https://doi.org/10.1371/journal.pone.0120495

Herman, E., Jamali, H. R., Nicholas, D., Osimo, D., Porcu, F. and Pujol, L. (2015). Analysis of Emerging Reputation and Funding Mechanisms in the Context of Open Science 2.0 (No. JRC94952). Institute for Prospective Technological Studies, Joint Research Centre.

Hey, A. J. and Trefethen, A. E. (2003). The data deluge: An e-science perspective

Houghton, J., Swan, A. and Brown, S. (2011) Access to Research and Technical Information in Denmark: Report to The Danish Agency for Science, Technology and Innovation (FI) and Denmark's Electronic Research Library (DEFF. Danish Agency for Science, Technology and Innovation.

Jong, S. and Slavova, K. (2014) When publications lead to products: The open science conundrum in new product development. Research Policy, 43(4), pp.645-654.

Jonkers, K., & Zacharewicz, T. (2016). Research Performance Based Funding Systems: a Comparative Assessment. European Commission. http://doi.org/10.2791/659483

Jubb, M., Plume, A., Oeben, S., Brammer, L., Johnson, R., Bütün, C., & Pinfield, S. (2017).

Monitoring the transition to open access.
Jubb, M., Plume, A., Oeben, S., Brammer, L., Johnson, R., Bütün, C., & Pinfield, S. (2017). Monitoring the transition to open access.
Kowalczyk, S. and Shankar, K. (2011) Data sharing in the sciences. Annual review of information science and technology, 45(1), 247-294.
Lakhani, K. R. and Panetta, J. A. (2007) The principles of distributed innovation. Innovations, 2(3), 97-112.
Lawson, S (2017). Report on offset agreements: evaluating current Jisc Collections deals. Year 2 – evaluating 2016 deals.
Liu, C.C. and Stuart, T. (2014). Positions and rewards: The allocation of resources within a science-based entrepreneurial firm. Research Policy, 43(7), pp.1134-1143.
Lundvall, B. A. (1992) National systems of innovation: An analytical framework. London: Pinter.
Lyon, L. (2009). Open Science at Web-Scale : Optimising Participation and Predictive Potential Consultative Report. Information Systems Journal, 29(November), 1–45.
Margolis, R, Derr, L, Dunn, M, Huerta, M, Larkin, J, Sheehan, J, et al. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. Journal of the American Medical Informatics Association, 21(6): 957–958. DOI: https://doi.org/10.1136/ amiajnl-2014-002974
Maxwell, J. A. (2012). Qualitative Research Design: An Interactive Approach: An Interactive Approach.
Mayer-Schonberger, V and Cukier, K (2013) Big Data: A Revolution That Will Transform How We Live, Work, and Think. Boston: Houghton Mifflin Harcourt.
McKiernan, E.C., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B.A., Ram, K., Soderberg, C.K. and Spies, J.R. (2016) How open science helps researchers succeed. Elife, 5, p.e16800.
Murray, F., and S. Stern (2007). "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis," Journal of Economic Behavior & Organization, 63(4), 648–687.
Nagel, E., & Kaplan, A. (1965). The Conduct of Inquiry: Methodology for Behavioral Science. American Sociological Review (Vol. 30). http://doi.org/10.2307/2091353
National Institutes of Health (2015) Commons Home Page | Data Science at NIH. Retrieved from: https:// datascience.nih.gov/commons
National Institutes of Health (2016) NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources. Retrieved from: https://grants.nih.gov/policy/sharing.htm
National Science Foundation (2011) Award and Administration Guide: Data Sharing Policy. Retrieved from: https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4
Nielsen, M. (2012). Reinventing discovery: the new era of networked science. Princeton University Press.
OECD (2015) Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris.
Olensky, M., Schmidt, M., & Van Eck, N.J. (2016). Evaluation of the Citation Matching Algorithms of CWTS and iFQ in Comparison to the Web of Science. Journal of the Association for Information Science and Technology, 67(10), 2550-2564.

doi:10.1002/asi.23590.
Olson, G.M., Zimmerman, A. and Bos, N. (2008) Scientific collaboration on the Internet. The MIT Press.
Open Knowledge Foundation (2015) Open Definition: Defining Open in Open Data, Open Content and Open Knowledge. Retrieved from: http://opendefinition.org/od/.
Orelli, B (2016) The Business of Genomic Data. O'Reilly Media. Retrieved from: http://www.oreilly.com/data/ free/the-business-of-genomic-data.csp.
Organisation for Economic Co-operation and Development (2007) OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: Organisation for Economic Co-Operation and Development, p. 24. Retrieved from: http://www.oecd.org/dataoecd/9/61/38500813.pdf.
Palmer, C L, Weber, N M and Cragin, M H (2011) The Analytic Potential of Scientific Data: Understand¬ing Re-use Value. Proceedings of the American Society for Information Science and Technology, 48(1):
Pasquetto, I V, Sands, A E, Darch, P T and Borgman, C L (2016) Open Data in Scientific Settings: From Policy to Practice. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Presented at the 2016 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: ACM, pp. 1585–1596. DOI: https://doi.org/10.1145/2858036.2858543
Pasquetto, I., Randles, B., & Borgman, C. (2017). On the Reuse of Scientific Data. Data Science Journal, 16.
Perkmann, M. and Schildt, H., (2015). Open data partnerships between firms and universities: The role of boundary organizations. Research Policy, 44(5), pp.1133-1143.
Pilat, D. and Fukasaku, Y. (2007). OECD principles and guidelines for access to research data from public funding. Data Science Journal, 6, OD4-OD11.
Piwowar, H. A., Vision, T. J. and Whitlock, M. C. (2011). Data archiving is a good investment. Nature, 473 (7347), 285-285.
Porter, M. E. (1990) The competitive advantage of nations. Harvard Business Review, 68(2), 73-93.
Powell, A. (2012). Democratizing production through open source knowledge: from open software to open hardware. Media, Culture & Society, 34(6), 691-708.
Powell, A. B. (2015). Open culture and innovation: integrating knowledge across boundaries. Media, Culture & Society,
Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A. and Asgari-Targhi, M. (2010) Adoption and use of Web 2.0 in scholarly communications. Philos Transact A Math Phys Eng Sci, 368(1926), 4039–4056.
Raymond, E. (1999) The cathedral and the bazaar. Knowledge, Technology & Policy, 12(3), 23-49.
Robinson-García, N., Torres-Salinas, D., Zahedi, Z., Costas, R. (2014). New data, new possibilities: exploring the insides of Altmetric.com. El Profesional de La Informacion, 23(4), 359–366. https://doi.org/10.3145/epi.2014.jul.03
Savage, C. J. and Vickers, A. J. (2009) Empirical study of data sharing by authors publishing in PLoS journals. PLoS ONE, 4(9), 9–11.
Science, Digital; Hahnel, Mark; Fane, Briony; Treadway, Jon; Baynes, Grace; Wilkinson, Ross; et al. (2018): The State of Open Data Report 2018. figshare. Paper. https://figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058

Shreeve J (2005) The Genome War: How Craig Venter Tried to Capture the Code of Life and Save the World. Ballantine Books: New York.

Sorenson, O. and Fleming, L. (2004) Science and the diffusion of knowledge. Research policy, 33(10), pp.1615-1634.

Stodden, V. (2010). The Scientific Method in Practice: Reproducibility in the Computational Sciences.

Szkuta, K. and Osimo, D. (2016) Rebooting science? Implications of science 2.0 main trends for scientific method and research institutions. Foresight, 18(3).

Szkuta, K., & Osimo, D. (2016). Rebooting science? Implications of science 2.0 main trends for scientific method and research institutions. Foresight, 18(3). http://doi.org/10.1108/FS-06-2014-0040

Taubenberger, J. K., Reid, A. H., Lourens, R. M., Wang, R., Jin, G. and Fanning, T. G. (2005) Characterization of the 1918 influenza virus polymerase genes. Nature, 437.

Tenopir, C, Allard, S, Douglass, K, Aydinoglu, A U, Wu, L, Read, E, et al. 2011 Data Sharing by Scientists: Practices and Perceptions. PLoS ONE, 6(6): e21101. DOI: https://doi.org/10.1371/journal.pone.0021101

The MoRRI indicators report, Monitoring Report D4.3, Pre-publication version (February 2018): http://www.technopolis-group.com/report/evolution-responsible-research-inovation-europe-morri-indicators-report-d4-3/

UE. (2015). Better Regulation Guidelines. Guidelines European Commission. http://doi.org/10.1016/S0022-3913(12)00047-9

Van Leeuwen, T., Meijer, I., Yegros, A., Costas, R. (2017). Developing indicators on Open Access by combining evidence from divese data sources. STI 2017 Conference. Paris (France)

Von Hippel, E. (2009) Democratizing innovation: the evolving phenomenon of user innovation. International Journal of Innovation Science, 1(1), 29-40.

Walsh, J.P. and Huang, H. (2014) Local context, academic entrepreneurship and open science: Publication secrecy and commercial activity among Japanese and US scientists. Research Policy, 43(2), pp.245-260.

Waltman, L., Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. Journal of the Association for Information Science and Technology, 65(3), 433–445. https://doi.org/10.1002/asi

Whyte, A., and Pryor, G. (2010) Open to all? Case studies of openness in Research. United Kingdom: Nesta

Wilkinson, M D, Dumontier, M, Aalbersberg, Ij J, Appleton, G, Axton, M, Baak, A, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3: 160018. DOI: https://doi.org/10.1038/sdata.2016.18

Wilsdon, J. R., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., & Wouters, P. (2017). Next-generation metrics: responsible metrics and evaluation for open science.

Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D.R., Los, W., Marinucci, M., Romary, L., Van de Sompel, H. and Vigen, J., (2010) Riding the wave: How Europe can gain from the rising tide of scientific data. European Union.

Wouters, P., Costas, R. (2012). Users , narcissism and control – tracking the impact of scholarly publications in the 21 st century. (M. Van Berchum K. Russell, Eds.), Image Rochester NY. SURFfoundation. Retrieved from http://www.surffoundation.nl/en/publicaties/Pages/Users_narcissism_control.aspx

Yin, R. K. (2009). Case study research : design and methods. Applied social research methods series ; (Vol. 5.). http://doi.org/10.1097/FCH.0b013e31822dda9e

Zahedi, Z., Costas, R., Wouters, P. (2014). How well developed are altmetrics? cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications. Scientometrics. https://doi.org/10.1007/s11192-014-1264-0

Zahedi, Z., Fenner, M., Costas, R. (2014). How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIS. In altmetrics14: expanding impacts and metrics (pp. 5–6). Bloomintong, Indiana, USA: An ACM Web Science Conference 2014 Altmetrics Workshop. https://doi.org/10.6084/m9.figshare.1041821