



DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION (RTD)

Brussels, April 2016 (update: July 2016)

rtd.ddg1.a.6(2016)2395458

BACKGROUND NOTE ON

OPEN ACCESS TO SCIENTIFIC PUBLICATIONS AND OPEN RESEARCH DATA

SUMMARY

- Open access (OA) can be defined as the practice of providing on-line access to scientific information that is free of charge to the user and that is re-usable. A distinction is usually made between OA to scientific peer reviewed **publications** and **research data**.
- The Open Science Agenda defines two high level ambitions for OA:
 - By 2020, all peer reviewed scientific publications are freely accessible
 - By 2020, FAIR data sharing is the default for scientific research
- In FP7 the Commission has been running a pilot for OA to scientific peer reviewed publications. In Horizon 2020 OA is now an obligation for all scientific, peer reviewed publications, allowing both 'gold' and 'green' open access.
- Furthermore, the Commission is running a flexible pilot on OA to research data ('ORD pilot'), which will be extended to cover all thematic areas of Horizon 2020 in 2017.
- As concerns the uptake of open access in the Member States (MS):
 - For open access to publications we see MS supporting either Gold, Green or both forms of open access. In many MS OA remains essentially a bottom-up movement, driven by stakeholders such as research institutes and funders, universities and libraries. Therefore, mostly soft measures have been implemented to promote open access. In particular, the proposed transition to full scale gold open access, based on a paper from Max Planck, has received considerable attention, although questions on the scalability of the deal remain to be answered.
 - OA to research **data** is much less developed across EU countries than OA to research publications, with a few MS very proactively driving the debate forward.

1. INTRODUCTION & BACKGROUND INFORMATION

Systemic changes are currently taking place in the way the science and research system functions. These changes –referred to as 'open science' – entail a shift towards a more open, collaborative and networked way of doing research. Open access is a key part of this phenomenon.¹

Open access (OA) can be defined as the practice of providing on-line access to scientific information that is free of charge to the user and that is re-usable.² In the context of R&D, open access to 'scientific information' refers to two main categories:

- a) Peer-reviewed scientific publications (primarily research *articles* published in academic journals)
- b) Scientific research *data*: data underlying publications and/or other data (such as curated but unpublished datasets or raw data).

Open access has the potential to improve scientific research (through improved reproducibility), involve citizens and society as well as accelerate innovation (see below and annex 1 for evidence).

The level of ambitions for the two open access related items of open science were defined as

- By 2020, all peer reviewed scientific publications are freely accessible
- By 2020, FAIR Data sharing is the default for scientific research

1.1. Scientific publications: from traditional to open access business models

Historically, access to scientific information has to a significant extent been based on **subscriptions** to scientific journals. For several centuries learned societies were in the forefront of publishing such scientific journals, the most prominent example being the Royal Society in the UK – an early not for-profit organisation. However, since the 20th century most scientific journals have been run by a variety of scientific publishers (sometimes on behalf of learned societies). A 2006 study on the publishing market in Europe found that while a huge number of small publishers exist, most of the market is concentrated in the hands of a few big players.³

¹ See <http://scienceintransition.eu/> and <http://ec.europa.eu/research/openscience/index.cfm> Other components of open science include (but are not limited to) alternative metrics, new ways of peer review and the evaluation of research careers.

² Legally binding definitions of 'open access' and 'access' in this context do not exist, but authoritative definitions of open access can be found in key political declarations on this subject, such as the Budapest Declaration (2002) and the Berlin Declaration (2003). These definitions describe 'access' in the context of open access as including not only basic elements such as the right to read, download and print, but also the right to copy, distribute, search, link, crawl, and mine.

³ See European Commission (2006) [The economic and technical evolution of the scientific publication markets in Europe](#) and more recently Vincent Larivière, Stefanie Haustein, Philippe Mongeon (2015) The Oligopoly of

In recent decades, journal prices have been rising much faster than the Consumer Price Index while the funds available to the libraries have remained static or have declined in real terms. As a result, academic and research libraries (even Harvard) have regularly cancelled serial subscriptions to accommodate price increases of the remaining current subscriptions. This so called 'serials crisis' was also the starting point for a bottom up movement of academics to promote open access in the early 2000s. Over the last years, an increasing number of governments, research funding bodies and research performing institutions world-wide have therefore developed open access policies to improve the access to the scientific publications resulting from the research they fund. Two main business models for open access to publications have been developed:

- A. **Self-archiving** (also referred to as 'green' open access) means that the published article or the final peer-reviewed manuscript is archived (deposited) by the author - or a representative - in an online repository before, alongside or after its publication. Repository software usually allows authors to delay access to the article ('embargo period'). Some publishers request embargo periods, arguing that these protect the value of the journal subscriptions they sell.
- B. **Open access publishing** (also referred to as 'gold' open access) means that an article is immediately provided in open access mode when published. In this model, the payment of publication costs is shifted away from readers (paying via subscriptions) to the author, often through a one-off charge, a so called 'Article (sometimes 'Author') Processing Charge' (APCs). These can usually be borne by the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

Additionally, new and alternative open access publishing models beyond 'gold' and 'green' are now emerging. One example is SCOAP3, an initiative in particle physics, which provides funding for all journals in this area to be converted to open access. Other relevant initiatives are LingOA, eLife and F1000. Some radical thinkers even abandon the journal as the dominant framework for publishing in the future driven by the availability of digital technologies, the ever increasing multidisciplinary interest of scientists and that the fact that scientists predominantly read scientific papers per se (through the internet and search engines) rather than complete journals.

1.2. Open access to research data

Open access to research data refers to the right to access and re-use research data⁴. While open access to scientific publications has been implemented for a decade and is increasing in terms of acceptance and use⁵, open access to research data is more recent. Opening up research data has in general the potential to not only improve scientific research and involve society but also to significantly contribute to economic growth (through open innovation) and is therefore relevant for the Digital Single Market. A recent study demonstrated that the European Bioinformatics Institute of the European Molecular Biology Laboratory generates a benefit to users and their funders of around 1.3 billion euros per year by making scientific information freely available to the global life science community. This is equivalent to more than 20 times the direct operational cost of the institute.⁶ Another example given in this context refers to the Copernicus earth observation satellites. Their data will be freely available and could generate up to 30 billion € and 50.000 new jobs by 2050.⁷ Furthermore, research data is valuable for open innovation, in particular as regards innovative start-ups, e.g. in the context of the app economy as well as the use of data in the context of the re-industrialisation of Europe ('industry 4.0').

However, open access is only one part of making data findable, accessible, interoperable and re-usable (FAIR) and therefore needs to be addressed in the wider context of 'open science', which includes supporting the community in a bottom-up fashion (the activities of the Research Data Alliance) and forward thinking on further infrastructure needs (*inter alia* including a European open science cloud).

1.3. Open access and IPR

It is sometimes alleged that open access is in conflict with Intellectual Property Rights (IPR), e.g. patenting. This is not the case. Researchers need to make the decision on whether they want to commercially exploit the results of their research or not – in the first case they will decide to protect their IPR, in the second case they should go down the open access route. Graph 1 provides a decision tree in this context.

⁴ 'Research data' refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.

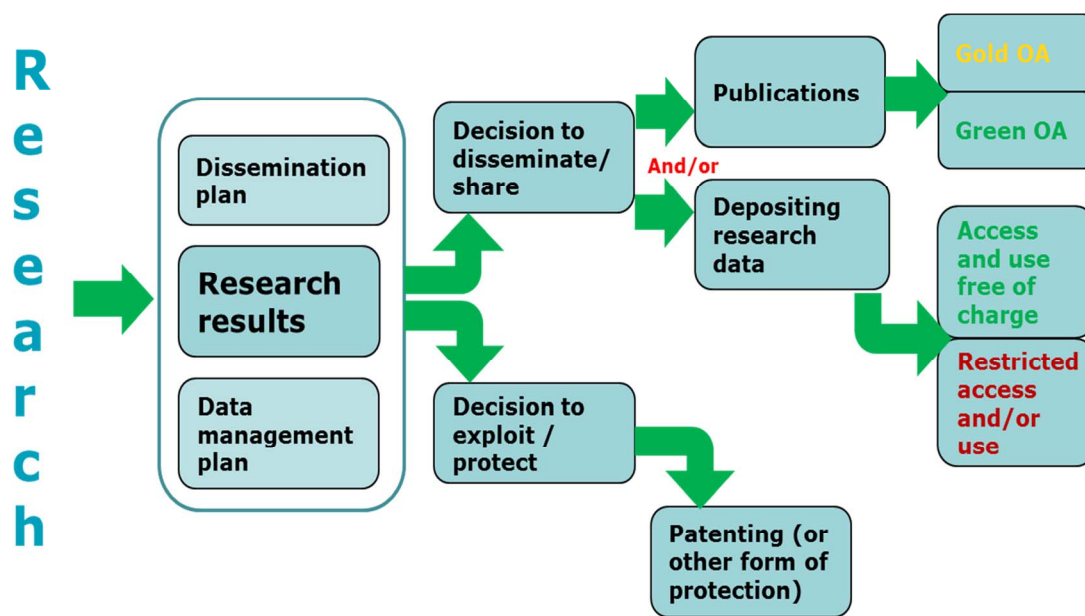
⁵ More than 50% of research publications from 2012 were open access in 2014, according to data collected by Science Metrix (counting both gold, green and other open access).

See <http://science-metrix.com/en/publications/reports>

⁶ See Beagrie et. al (2016) The Value and Impact of the European Bioinformatics Institute.
<http://www.beagrie.com/static/resource/EBI-impact-summary.pdf>

⁷ For instance, the Latvia-based SME 'ThermoCERT' uses Copernicus data as the basis of tools for monitoring the thermal efficiency of buildings. See RDA Europe (2014) The Data Harvest: How sharing research data can yield knowledge, jobs and growth. P.13

Graph 1: Open access to scientific publications and research data in the wider context of dissemination and exploitation



Source: developed jointly between DG RTD, DG CNECT, IPR helpdesk

It is also important to note that

- i. Open access publications go through the same peer review process as non-open access publications;
- ii. Open access does not entail an obligation to publish: it is up to researchers whether they want to publish some results or not.

2. STATE OF PLAY

The Commission has been leading by example in integrating open access in its own multiannual framework programmes as well as promoting open access with Member States and with stakeholders. A key document in this regard was the 2012 Recommendation on *access to and preservation of scientific information* (C(2012) 4890 final), which covers open access to publications and data as well as preservation and curation and changing the culture of research.

2.1. Open access in the Framework Programmes

2.1.1 Open access to scientific publications in FP7

The Commission already promoted open access to scientific peer reviewed publications in FP7. For green open access it is running a pilot action in 7 areas of FP7⁸, mandating green open access on a 'best effort' basis (embargo period 6 months for natural sciences and 12 months for social sciences and humanities). Furthermore, the Commission also supports open access in all areas of FP7 by making relevant costs for gold open access (APCs) eligible for reimbursement as part of the overall project grant.

From the total of 171,258 scientific peer reviewed publications created with FP7 funding (at the time of writing) 92,826 are OA, 3,216 are restricted (i.e., OA but with a more restrictive license or restricted to specific groups), while 315 are still in embargo. This translates to an **open access rate of 54,2% for all scientific peer reviewed publications created during the lifetime of FP7 so far.**⁹

It should be also noted, however, that not all FP7 projects have yet ended and that, furthermore, publications are also published after the end of a project. The figure of 54% is therefore provisional. **Furthermore, looking at individual years, open access publications have significantly increased in recent years, reaching 67% in 2014.**¹⁰

This could be seen as an indicator of the growing support towards open access in the scientific community. The Commission is therefore on track to reach 60% open access by 2016, a target stipulated for the whole of Europe in the 2012 Recommendation on Access to Scientific Information.

Open access to research data was not addressed in FP7.¹¹

⁸ Energy, Environment, Health, Information & Communication Technologies [only cognitive systems, interaction & robotics], Research Infrastructures [only e-Infrastructures], Science in Society, Socioeconomic Sciences & Humanities

⁹ Data from the OpenAIRE project: OpenAIRE and its successor projects inter alia assist in monitoring FP and Horizon 2020 research outputs. OpenAIRE computes statistics based on data retrieved from a range of sources: OA repositories and journals (literature and data), CRIS systems, library databases and end-user feedbacks. They also employ a variety of deduplication, cleaning, and text-mining processes on the metadata, as well as on the actual content.

¹⁰ Based on publications which can be clearly attributed to a specific year.

¹¹ Except for voluntary guidelines by the ERC which addressed primary data.

2.1.2 Open access to scientific publications and research data in Horizon 2020

Based on the pilot in FP7 (see above) in Horizon 2020 the Commission has made open access to peer-reviewed scientific publications mandatory.

- For open access publishing (gold open access), researchers can publish in open access journals, or in journals that sell subscriptions and also offer the possibility of making individual articles openly accessible (hybrid journals). In that case, publishers often charge an article processing charge (APC). These costs are eligible for reimbursement during the duration of the action as part of the Horizon 2020 grant. For APCs incurred after the end of the grant agreement, a mechanism for reimbursing some of these costs is being piloted and implemented through the OpenAIRE project. Note that in case of gold open access publishing, a copy must also be deposited in a repository.
- For self-archiving (green open access), researchers deposit the final peer-reviewed manuscript in a repository of their choice. In this case, they must ensure open access to the publication within six months of publication (12 months in case of the social sciences and humanities).

While it is too early for data on Horizon 2020 scientific peer reviewed publications to be available, the services will continue to monitor the uptake of open access to scientific publications in Horizon 2020.

As for open access to **research data** the Commission committed itself to running a flexible pilot in Horizon 2020 ('ORD pilot'). The pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects, taking into account the need to balance openness and protection of scientific information, commercialisation and IPR, privacy concerns, security as well as data management and preservation questions.

In the Work Programmes 2014-2016, this ORD pilot concerned selected areas of Horizon 2020. However, in the Communication 'a European Cloud Initiative – Building a competitive data and knowledge economy in Europe' the Commission commits itself to "make open research data the default option, while ensuring opt-outs, for all projects of the Horizon 2020 programme" as of 2017. As of the revised version of the Work Programme 2017, the ORD pilot has therefore been extended to all thematic areas of the Horizon 2020 Research and Innovation Programme.

While open access to research data thereby becomes applicable by default in Horizon 2020, the Commission also recognises that there are good reasons to keep some or even all research data generated in a project closed. It has, therefore, provided robust opt-out possibilities at any stage, that is in the application phase, during the grant agreement preparation (GAP) phase and after the signature of the grant agreement.

The ORD pilot applies primarily to the data needed to validate the results presented in scientific publications. Other data can also be provided by the beneficiaries on a voluntary basis, as stated in their Data Management Plans. Costs associated with open access to research data, can be claimed as **eligible costs** of any Horizon 2020 grant. It should be noted that participation in the ORD pilot is not part of the evaluation of proposals: in other words, proposals are not evaluated more favourably because they are part of the pilot and are not penalised for opting out of the pilot.

For the uptake of the pilot in 2014 and 2015 (when its scope was more restricted), figures show an **opt-out rate of 34.6%** in the core areas of the pilot. In other words 65.4% of projects in the core areas participate in the ORD pilot. The most important reasons for opt-outs were (i) IPR concerns (37%), (ii) projects which do not expect to generate data (18%); or (iii) over privacy concerns (18%). Outside the core areas, **11.9%** of projects make use of the voluntary **opt-in** possibility.

Table 1 Open access in FP7 and Horizon 2020: a comparison

Issue	FP7	Horizon 2020
<i>Scientific publications</i>		
Thematic coverage	Seven areas in FP7	All areas of Horizon 2020
Nature of the OA requirement	Best effort (special clause 39 MGA)	Obligation (article 29.2 MGA)
Allowed embargo for green OA	6-12 months	6-12 months
Costs for gold open access	Eligible during the duration of the project	Eligible during the duration of the project and a pilot action to fund some costs afterwards.
Copyright provisions	Not specified	Authors "encouraged" to retain copyright and grant licences to publishers (e.g. creative commons) instead, but not a hard obligation.
<i>Research data</i>		
Data management	Not specified	Best effort to deposit data but not to make open (data management) Data management plans obligatory for actions participating in the pilot scheme, optional for others

Open access to research data	Not specified	Pilot scheme for making research data open, extended to all areas of Horizon 2020 by 2017 – but option to opt-out at any stage
------------------------------	---------------	--

Generally, the development of open access policies for publications and data in Horizon 2020 proceeded organically in interaction with the Council and the European Parliament. A variety of stakeholders were consulted, including industry, and their views were taken into account, in particular as regards the design of the Open Research Data Pilot.

2.2. Open Access in the Member States

2.2.1. Open Access to Scientific Publications

Following up on the Commission's Recommendation on access to and preservation of scientific information of (C(2012) 4890 final), each EU Member State has nominated a National Point of Reference, with the task of reporting on the implementation of open access in the Member States. Most EU Member States reported a national preference for one of the two types of open access, either the Green (self-archiving) or the Gold (open access publishing) model. Preference for the Green model is found in Belgium, Cyprus, Denmark, Estonia, Greece, Ireland, Lithuania, Malta, Norway, Portugal, Slovakia and Spain. Those expressing a preference for the Gold model are Hungary, the Netherlands, Romania, Sweden and the United Kingdom. Other Member States support both models equally, such as Germany, France, Croatia, Italy, Luxembourg, Poland and Finland. However, the expressed preferences for one of the two models are not pure models in which only one route is followed. Instead, there is generally a system of predominance of one model with the possibility of using the other model, so a mixture of both routes results. While few Member States have a national law requiring open access to publications, a mandate put in place by law is not necessarily stronger or more effective than a mandate put in place by a single institution or funder. For example, an open access mandate is strong as it ties open access to possible withdrawal of funds in the case of non-compliance, or to the evaluation of researchers' careers.¹²

2.2.2 Open Access to Research Data

The data from the NPR report suggests that policies on open access to research data are less developed across EU countries than policies and strategies on open access to research publications. However, individual Member State feedback shows a general acknowledgement of the importance of open research data and of policies, strategies and actions addressed at fostering the collection, curation, preservation and re-use of research data. This is also reflected in the March 2015 Competitiveness Council which called for action to remove

¹² Full report available at http://ec.europa.eu/research/openscience/pdf/openaccess/npr_report.pdf

obstacles not only to wide access to publicly funded research publications but also to underlying data.

Based on the self-reporting of the EU Member States and participating associated countries, the following classification is proposed.

- Very little or no open access to research data policies in place and no plan for a more developed policy in the near future: Cyprus, Latvia, Luxembourg, Malta, Poland.
- Very little or no open access to research data policies in place, but some plans in place or under development: Austria, Belgium, Croatia, Czech Republic, Estonia, Hungary, Italy, Portugal, Romania, Slovakia, Sweden, Turkey.
- Open access policies/institutional strategies or subject-based initiatives for research data already in place: Denmark, Finland, France, Germany, Ireland, Lithuania, the Netherlands, Norway, Slovenia, the United Kingdom.

The May 2016 Council Conclusions on ‘the transition towards an Open Science system’ call for a transition to open access to scientific peer reviewed publications as the default by 2020. They also support the optimal re-use of data with the underlying principle of "as open as possible as closed as necessary" and welcome the intention of the Commission to make research data produced by Horizon 2020 open by default, whilst recognising the right of opting out.

3. CONTACT

E-Mail: RTD-open-access@ec.europa.eu

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

Twitter: [@OpenAccessEC](https://twitter.com/OpenAccessEC)

4. FURTHER INFORMATION

European Commission (2016) Open innovation, open science, open to the world. A vision for Europe <http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbKI0416263/>

Horizon 2020 Guidance for Applicants on open access to scientific publications and research data http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

Council Conclusions on the transition towards an Open Science system

<http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>

European Commission (2015) Dataset on the Uptake of ORD pilot on a call-by-call basis

<https://data.europa.eu/euodp/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020>



4. ANNEXES

- Annex 1: Overview of the benefits of open access
- Annex 2: The Open Access to Research Data Pilot: coverage in the WP 2016

ANNEX 1: OVERVIEW OF THE BENEFITS OF OPEN ACCESS

Generally, fuller and wider access to scientific publications and research data can help to:

(i) improve scientific research:

Scientific articles that are available in the open are between 26% and 64% more cited on average, which clearly shows that open access to scientific publications improves the circulation of knowledge.¹³

As for open access to research data, the wider accessibility of results could help to increase the reproducibility of research, which can be as low as 10-30% in key areas, such as cancer research.¹⁴

Additionally, open access information can cross-fertilize interdisciplinary research, with is essential for addressing the grand challenges of our time.¹⁵

(ii) accelerate innovation: the potential benefits of opening up research information are clearly recognised in the European Commission's investment plan for Europe where it is stated that in order to "boost research and innovation, EU competitiveness would benefit from fewer barriers to knowledge transfer, open access to scientific research and greater mobility of researchers."¹⁶

a) In this context, open access to publications can help to overcome the barriers that innovative companies, in particular SMEs, face in accessing the results of research funded by the public purse. It has been estimated that switching to open access could results in savings of around £400 million for the UK, 133 million € for the Netherlands and 80€ million Euros a year for Denmark.¹⁷

b) Data is becoming increasingly important for all aspects of the European economy.¹⁸ A specific example of the economic impact of opening up *research* data are the Copernicus earth observation satellites.¹⁹ Their data will be freely available and could generate up to 30 billion € and 50.000 new jobs by 2050. The European Bioinformatics Institute of the European Molecular Biology Laboratory generates a benefit to users and their funders of around 1.3 billion euros per year by making scientific information freely available to the

¹³ Archambault, et.al (2014) Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013 p.22. http://www.science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf

¹⁴ See the *Economist*, 'How Science goes wrong' (issue of 19 October 2013).

¹⁵ This is inter alia reflected in the title of the 2014 Finish roadmap for open access entitled 'Open science and research leads to surprising discoveries and creative insights'.

(Com 2014 903 final) p.16.

¹⁷ Houghton et. al (2009) Economic implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits. Swan, A. (2010) Modelling scholarly communication options: costs and benefits for universities.

¹⁸ See Buchholtz, S. Bukowski, M., Śniegocki A. (2014) Big and open data in Europe: A growth engine or a missed opportunity?" Warsaw Institute of Economic Studies / Demos Europe http://www.bigopendata.eu/wp-content/uploads/2014/01/bod_europe_2020_full_report_singlepage.pdf.

¹⁹ RDA Europe (2014) The Data Harvest: How sharing research data can yield knowledge, jobs and growth. P.13

global life science community. This is equivalent to more than 20 times the direct operational cost of the institute.²⁰ Finally, the Human Genome Project has made its data open access and has spawned the creation of a whole new industry.

(iii) involve citizens and society: Making research openly available is potentially beneficial not only for the individual citizen but also for NGOs and other non-for profit organisations, which often cannot afford subscriptions to a large number of academic journals but for whom academic research is nevertheless very important, e.g. in the field of climate change. A recent study has shown that health NGO staff utilize more research in the course of their work as a result of increasing open access to research.²¹

²⁰ See Beagrie et. al (2016) The Value and Impact of the European Bioinformatics Institute.
<http://www.beagrie.com/static/resource/EBI-impact-summary.pdf>

²¹ Moorhead et al (2015) 'In an Age of Open Access to Research Policies: Physician and Public Health NGO Staff Research Use and Policy Awareness' Plos1.
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0129708>

ANNEX 2: THE OPEN ACCESS TO RESEARCH DATA PILOT: COVERAGE IN THE WP 2016

The main characteristics of the pilot (opt-out/opt-in approach) have been retained in the Work Programme 2016. As regards the thematic coverage of the pilot, the core areas have been updated and slightly expanded based on the feedback from the thematic directorates and units. The table below presents the updated list of the core areas, as presented in the general introduction of the work programme. As can be seen, the pilot has been extended in three areas (marked 'yes') in the table.

As of the revised version of the Work Programme 2017, the ORD pilot has been extended to all thematic areas of the Horizon 2020 Research and Innovation Programme.

Table: the ORD Pilot in the Work Programme 2016 – updated thematic coverage

<i>Work Programme 2016</i>	<i>Extended coverage compared to WP 2014-15?</i>	<i>Comment</i>
Future and Emerging Technologies	No	WP already completely covered
Research infrastructures (including e-Infrastructures)	Yes	in WP 2014-2015, only the e-infrastructure part was covered, now the whole WP is covered
Leadership in enabling and industrial technologies – Information and Communication Technologies	No	WP already completely covered
Nanotechnologies, Advanced Materials, Advanced Manufacturing and Processing, and Biotechnology: 'nanosafety' and 'modelling' topics.	Yes	New in WP 2016 – note that only 'nanosafety' and 'modelling' topics participate, not the whole WP
Societal Challenge: Food security, sustainable agriculture and forestry, marine and maritime and inland water research and	Yes	New in WP 2016 – note that only selected calls participate, not the whole WP

the bioeconomy - selected topics in the calls H2020-SFS-2016/2017, H2020-BG-2016/2017, H2020-RUR-2016/2017 and H2020-BB-2016/2017, as specified in the work programme		
Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials	No	
Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies	No	WP already completely covered
Science with and for Society	No	WP already completely covered
Cross-cutting activities - focus areas – part Smart and Sustainable Cities	N/A	No change in content coverage. However, Smart and Sustainable Cities was listed under the Secure, Clean and Efficient Energy WP in 2014-2015, now it is listed as a cross cutting activity