



# **Open Science policy and infrastructure support in the European Commission**

## **Joint COAR-SPARC Conference**

**Porto, 15 April 2015**

**Jarkko Siren  
European Commission  
DG CONNECT  
eInfrastructure**

*Author's views do not commit the European Commission*

# Open Science in the EC

## *Expanding scope*

- OA to peer-reviewed scientific publications
  - H2020 Grant Agreement obligation / FP7 pilot / OpenAIRE
- OA to research data / data sharing / data management
  - H2020 pilot / EUDAT / OpenAIRE
- Open Science / Science 2.0
  - open consultation (closed 30.09.2014) / WP2016-17?

**A Reinforced European Research Area Partnership for Excellence and Growth, COM(2012) 392 – July 2012**

**Towards better access to scientific information: boosting the benefits of public investments in research, COM(2012) 401 final - July 2012**

**Commission, Recommendation on access and preservation of scientific information, C(2012) 4890 final – July 2012**

## Horizon 2020

- **Open Access to Scientific Publications**
- **Pilot on research data: Data Management Plan**



# What is Horizon 2020



- **Initial Commission proposal for a €80 billion research and innovation funding programme (2014-2020); now just over €70 billion**
- **A core part of Europe 2020, Innovation Union & European Research Area:**
  - Responding to the economic crisis to invest in future jobs and growth
  - Addressing people's concerns about their livelihoods, safety and environment
  - Strengthening the EU's global position in research, innovation and technology

# OA to publications mandate in H2020

**Each beneficiary must ensure OA to all peer-reviewed scientific publications relating to its results:**

- **Deposit** published version or final peer-reviewed manuscript in a **repository** of the researchers choice
- Ensure OA on publication or at the latest within 6/12 months
- Ensure OA to the bibliographic metadata that identify the deposited publication, via the repository
- **Aim to deposit at the same time the research data needed to validate the results ("underlying data")**

# **Pilot on Open Research Data in H2020**

**It is pilot: a way to define future action.**

## **Types of data concerned:**

- Data needed to validate the results presented in scientific publications ("underlying data")
- Other data as specified in Data Management Plan (=up to projects)
- Data Management Plans (DMPs) mandatory for all projects participating in the pilot

# Pilot on Open Research Data in H2020

## Beneficiaries participating in the Pilot will:

- Deposit this data in a research data repository of their choice
- Take measures to make it possible to access, mine, exploit, reproduce and disseminate free of charge
- Provide information about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (where possible, provide the tools and instruments themselves)

# **Pilot on Open Research Data in H2020**

- **20% of Horizon 2020.**
- **Voluntary opt in and conditional opt out possible.**
- **Access and reuse of individual data sets can be restricted (reasons detailed in the DMP).**



# Looking beyond 2015: open questions

- How to support H2020 objectives through infrastructures
  - Societal Challenges
  - Innovation, jobs and growth
- Sustainability of (global) Scientific Data Infrastructure (funders collaboration?)
- Infrastructures that support wide reuse (e.g. TDM), long term preservation
- Etc.

# Staying Competitive in Science

- Large scale collaborations becoming the norm
  - *often global*
  - *virtual research communities*
  - *access to rare/remote resources*
- Data-intensive science and innovation
  - *Use and manage exponentially growing sets of data*
- Experimentation in silico, simulation
  - *Use of high-performance computing*

## 3 guiding principles

- **Data and Computing e-infrastructures go together**
  - **Research Infrastructures and e-Infrastructures go together**
- **Research Data –European Policy Framework**
- **Research Data Alliance**
  - **Report “Data Harvest” (follow-up of “Riding the Wave”)**

## Data has been and remains key to science

Need for expensive instruments is something that increasingly conditions scientific production (need for more powerful telescopes, light sources, research boats, geological probes etc)

Intrinsic to the ambition that European researchers remain at the vanguard of scientific discovery

But there is something about research data:

**information opens new possibilities for science**

# Research logic machines

Now **research data** is stored in digital form. Easier to be processed by "**logic machines**" programmed with complex models able to dig into the data

Logic machines are made of **human scientific knowledge and creativity, software** and the underlying **hardware**

Scientist notebooks can now be **linked** to a huge amount of other **data resources** (including scientific papers), **computers** with unprecedented capacity, eventually connected to **global networks**

# Digital scholarly record

**Publication, data, software, etc. repositories** have the potential to become the foundational element of the scholarly record.

+ identifier infrastructures + registries

This digital record should start with basic, registered, linked and validated **research claims**. All higher level **services** and **products** can be added on top.

Necessary conditions:

- Authentication and authorisation infrastructure
- Open Access to publications (SA license?)
- Protection of author's rights

## Useful definitions

**Data:** digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings

(not include lab notebooks, preliminary analysis, drafts of scientific papers, plans for future research, peer review reports, communication with peers, physical objects, lab specimens)

*[c.f. White House Memo on "Increasing Access to the Results of Federally Funded Scientific Research" paraphrasing OCDE definition]*

**Data infrastructures:** services, applications, tools, knowledge and policies for research data to be discoverable, understandable, accessible, preserved and curated... and available 24/7



## Issues to be addressed (e-infrastructure)

The EC in coordination with EU Member States is looking after research data as an infrastructure

As a valuable and a strategic resource, research data opens at least three key issues to be addressed<sup>(\*)</sup>:

- **How data can be networked**
- **How to envision and set up data governance on a global scale**
- **How the EU can play a leading role in helping start and steer this global trend**

*(\*) Fred Friend, Jean-Claude Guédon Herbert van de Sompel  
“Beyond Sharing and Re-using: Toward Global Data Networking”*



- RDA community focuses on building **social, organizational and technical infrastructure** to
  - **reduce barriers to data sharing and exchange**
  - **accelerate the development of coordinated global data infrastructure**



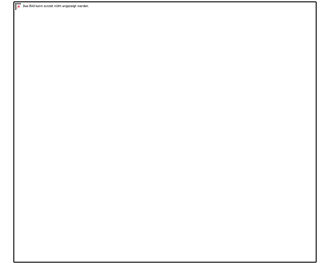
## CREATE → ADOPT → USE

### RDA Working Group Infrastructure Deliverables are:

- **Focused pieces of adopted code, policy, infrastructure, standards, or best practices** that enable data to be shared and exchanged
- **“Harvestable” efforts** for which 12-18 months of work can eliminate a roadblock for a substantial community
- **Efforts that have substantive applicability** to “chunks” of the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today** while more long-term or far-reaching solutions are appropriately discussed in other venues

## take five

**5 principles** describing the benefits of a global research data infrastructure (G8+O6)



Publicly funded research data is:

**Discoverable** – IDs, Descriptive Metadata, ...

**Accessible** – Acknowledgment, License, Terms of Use, Intellectual Property, Legal ...

**Understandable** – Semantics, Analysis, Quality, Language translation ....

**Manageable** – Responsibility, Costs, Preservation ...

**People** (Usable) - Workforce, Cultural, Training, ...



# **The Data Harvest Report**

**How sharing research data can  
yield knowledge, jobs and grow**

**A RDA Europe Report**



# OA in Horizon 2020: where to look

- Regulation establishing Horizon 2020 (article 18)
- Specific Programme (preamble 1.3)
- Rules for Participation (article 43)
- Work Programme 2014-15 (Introduction 1.5 and relevant areas)
- Model Grant Agreement (articles 6.2.D.3, 29.2 and 29.3)
- Annotated Model Grant Agreement (reference to Guidelines below)
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
- Guidelines on Data Management in Horizon 2020
- Source for all documents: Participant Portal (reference documents)  
<http://ec.europa.eu/research/participants/portal/desktop/en/home.html>
- [www.openaire.eu](http://www.openaire.eu) (NOADs pages, H2020 toolkit)



**Jarkko Siren**  
**[jarkko.siren\(at\) ec.europa.eu](mailto:jarkko.siren(at) ec.europa.eu)**

**Thank you!**