Données de recherche dans Horizon 200

Celina Ramjoué

**Head of Sector
Open Access to Scientific Publications and Data
DG CONNECT
celina.ramjoue@ec.europa.eu**

**6ème journées Open Access: La science ouverte en marche
Paris, 12-14 octobre 2015**

European Commission 2014-2019

Andrus Ansip, Vice-President, Digital Single Market

Günther Oettinger, Commissioner for Digital Economy and Society

Carlos Moedas, Commissioner for Research, Science and Innovation

Digital Single Market Strategy (DSM) May 2015

DSM: An area in which individuals and businesses can access and exercise online activities under conditions of fair competition, with a high degree of consumer and data protection, irrespective of their nationality or place of residence

One of ten Juncker priorities

Importance of R&I for DSM: data-driven economy, free flow of data, open science

Copyright, including text and datamining (TDM)

Why open access?

To optimise the impact of publicly-funded research

Expected benefits of open access:

- Good for science: efficiency, verifiability, transparency
- Good for the economy: access and take-up by industry
- Good for society: broader, faster, transparent and equal access for citizens

European Commission

Commissioner view

Carlos Moedas, Commissioner for Research, Science and Innovation

"**Open science** is [...] about making sure that science serves innovation and growth. It guarantees **open access** to publicly-funded research results and the possibility of knowledge sharing [...]."

Blog post by Commissioners Oettinger and Moedas (22 June 2015): Open science for a knowledge and data-driven economy

https://ec.europa.eu/commission/2014-2019/oettinger/blog/open-science-knowledge-and-data-driven-economy_en

Günther Oettinger, Commissioner for Digital Economy and Society

Priorities of Commissioner Moedas

- **Open Innovation**
- **Open Science**
- **Open to the world**

Open Access to Research Data
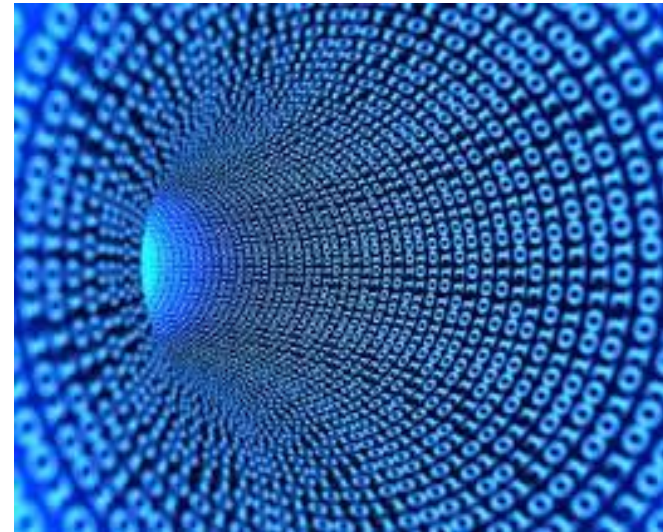
Pilot on Open Research Data in H2020

Key questions:

Which thematic areas are covered?

What data is covered?

What are the requirements?

What about data management?

Pilot on Open Research Data: Scope(1)

Areas of the 2016-2017 Work Programme participating in the Open Research Data Pilot are:

- Future and Emerging Technologies
- Research infrastructures – (new: coverage of the whole area)
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Nanotechnologies, Advanced Materials, Advanced Manufacturing and Processing, and Biotechnology: 'nanosafety' and 'modelling' topics (new)
- Societal Challenge: Food security, sustainable agriculture and forestry, marine and maritime and inland water research and the bioeconomy - selected topics as specified in the work programme (new)

Pilot on Open Research Data: Scope(2)

Continued

- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials

- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies

- Science with and for Society

- Cross-cutting activities - focus areas – part Smart and Sustainable Cities (moved from Energy WP)

Projects in other areas are encouraged to participate on a voluntary basis

Pilot on Open Research Data: Opt-out

Projects may opt out of the Pilot on Open Research Data in Horizon 2020 in a series of cases (submission stage):

- If the project will not generate / collect any data

- Conflict with obligation to protect results

- Conflict with confidentiality obligations

- Conflict with security obligations

- Conflict with rules on protection of personal data

- If the achievement of the action's main objective would be jeopardised by making specific parts of the research data openly accessible (to be explained in data management plan)

Opting out during project also possible

Being in the Pilot does not mean opening all data

Pilot on Open Research Data: requirements

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)

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)

Approach: as open as possible, as closed as necessary

Data management in Horizon 2020

- Data Management Plans (DMPs) mandatory for all projects participating in the Pilot, optional for others

    - DMPs are NOT part of the proposal evaluation

    - To be generated within first 6 months of project, updates as needed

- DMP questions:

    - What data will be collected / generated?

    - What standards will be used / how will metadata be generated?

    - What data will be exploited? What data will be shared / opened?

    - How will data be curated and preserved?

- **DMP: tool to determine what datasets can/cannot be open**

Resources

H2020 Guidelines:

Open access:
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Data management:
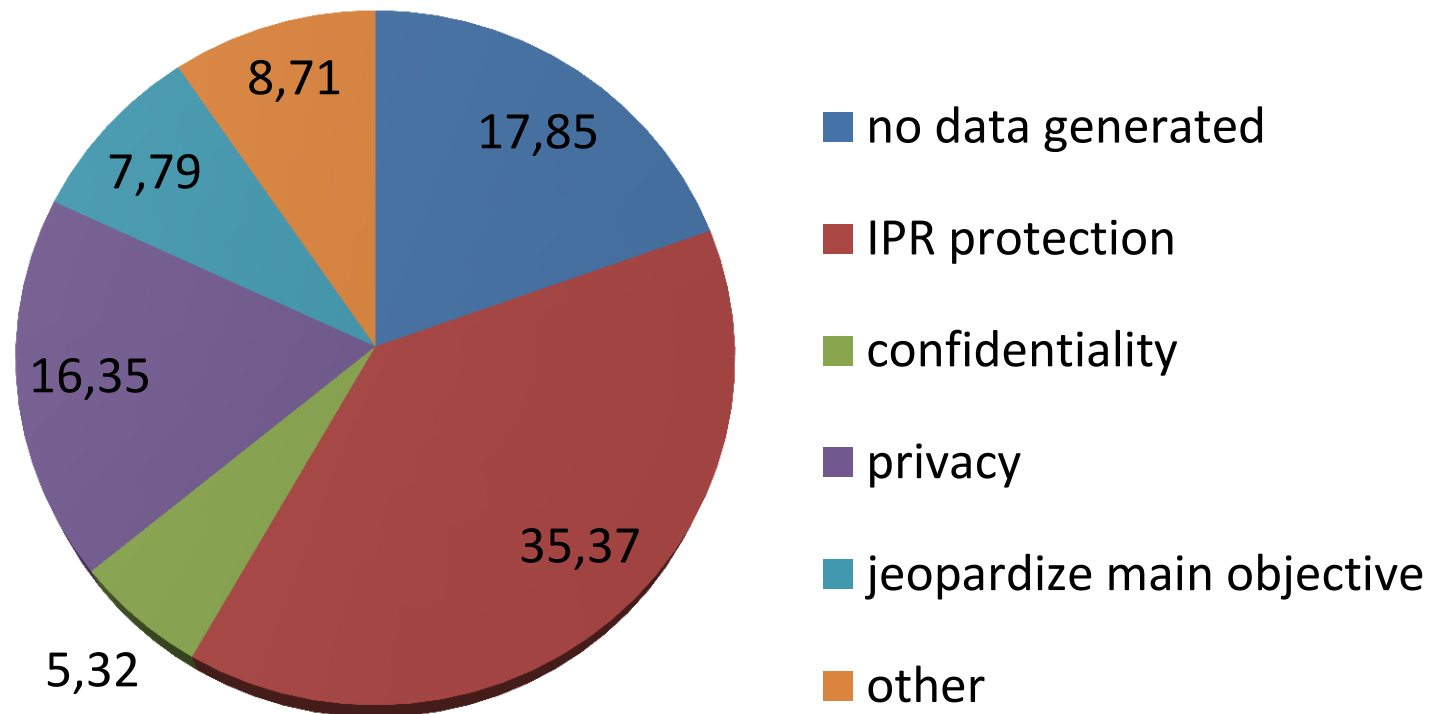http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

OpenAire - www.openaire.eu

publications and data, deposit and search, national OA

heldesks, and more!

ORD Pilot: take-up in first calls of H2020

- Basis: 3699 Horizon 2020 signed grant agreements

- Calls in core-areas: **opt out 34,6%** (149/431)

  → In other words 65,4% of projects in the core areas participate in the ORD pilot.

- Other areas: voluntary **opt in 11,9%** (409/3268)

→ Limited divergence from 2014 proposal figures but larger dataset used

→ Note that 100% participation is not feasible or even desirable (e.g. not all projects generate data)

ORD Pilot: opt-out reasons among proposals

Pie chart values: 17,85; 35,37; 5,32; 16,35; 7,79; 8,71

Legend:
- no data generated
- IPR protection
- confidentiality
- privacy
- jeopardize main objective
- other

ORD Pilot: a chance to co-shape policy

- Ambitious, yet pragmatic design of the pilot: broad scope, opt-out, voluntary participation possible

- Pilot is flexible; numerous safegards in place

- Aim: kick-starting a virtous circle, a cultural change

- Uptake of and experiences with the Pilot need to be monitored during the complete life cycle of a project: from application, to grant preparation, execution and final reporting

- **Participating in the Pilot means co-shaping European policy on opening up research data**

Text and datamining (TDM)

Text and Data Mining
What are we talking about?

TDM, Content Mining, Data Analytics…

- Not only Text but also Data (AV material, chemical formulas…)

Increased amounts of freely available or legally accessible content (in part. through <u>Open Access</u>)

- The 'mining' we talk about is not stealing material!

TDM is subject to <u>lawful access</u>

- Legal hurdles e.g. abstract minable but not the whole paper, potential cases of researchers infringing copyright and/or 'delocalising' TDM activities

<u>Multi-disciplinarity</u> of research and TDM practices

- Different fields of research ➔ hence different sources

Non-legal issues e.g. <u>technical barriers</u>

- Not resolvable by legal action

Why do we care?

Again ... optimise the impact of publicly-funded scientific research

Strong signals from researchers and innovators that we need a change in approach of data analytics (TDM)
- Risk of consequences for future of research base and growth in the EU

Policy framing:
- Digital Single Market Strategy
- Open Innovation, Open Science, Open to the world

What did the EC say?

EC Communication **"A Digital Single Market Strategy for Europe" (6 May 2015)**

- *Innovation in research **for both non-commercial and commercial purposes**, based on the use of text and data mining (e.g. copying of text and datasets in search of significant correlations or occurrences) may be hampered because of an **unclear legal framework** and **divergent approaches at national level**. The need for greater legal certainty to **enable** researchers and educational institutions to make wider use of copyright-protected material, including across borders, so that they can benefit from the potential of these technologies and from **cross-border collaboration** will be assessed, as with all parts of the copyright proposals in the light of its **impact on all interested parties**.*

What did Member States say?

[Council Conclusions](#) on open, data-intensive and networked research as a driver for faster and wider innovation (28-29 May 2015)

- *The Council of the European Union [...] NOTES the need to **promote innovation driven by text and data mining taking into account research needs**, and to consider the impact, **including financial aspects**, of the re-use of the **already legally accessible content**; and UNDERLINES the need to ensure **legal certainty and an appropriate regulatory framework** that would facilitate a **science and innovation-friendly environment** for better use of **data**;*

What did the EP say?

EP Report adopted on 7 July 2015 (MEP Julia Reda Rapporteur)

- *The European Parliament [...] 48. Stresses the **need to properly assess the enablement of automated analytical techniques for text and data** (e.g. 'text and data mining' or 'content mining') **for research purposes, provided that permission to read the work has been acquired**;*

Own initiative

- Sets common views from different (political/national) opinions

Text and datamining (TDM)

Awaiting a (copyright) proposal ...

Stay tuned!

Données de recherche dans Horizon 200

Celina Ramjoué

**Head of Sector
Open Access to Scientific Publications and Data
DG CONNECT
celina.ramjoue@ec.europa.eu**

**6ème journées Open Access: La science ouverte en marche
Paris, 12-14 octobre 2015**