THE ROLE OF RESEARCH DATA & OPEN SCIENCE

Jean-Claude Burgelman
Eva Podgorsek, Daniel Spichtinger
DG RTD - European Commission

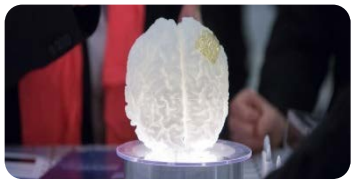
New roles in open science and data stewardship

University Venice, e-infrastructures Austria, Elsevier

Venice, 25 November 2016

CONTENT

Data-driven Science in Europe



DDS as part of a system change: Open Science

Creating a Data Culture: EC Policies on Open Access to Publications & Research data

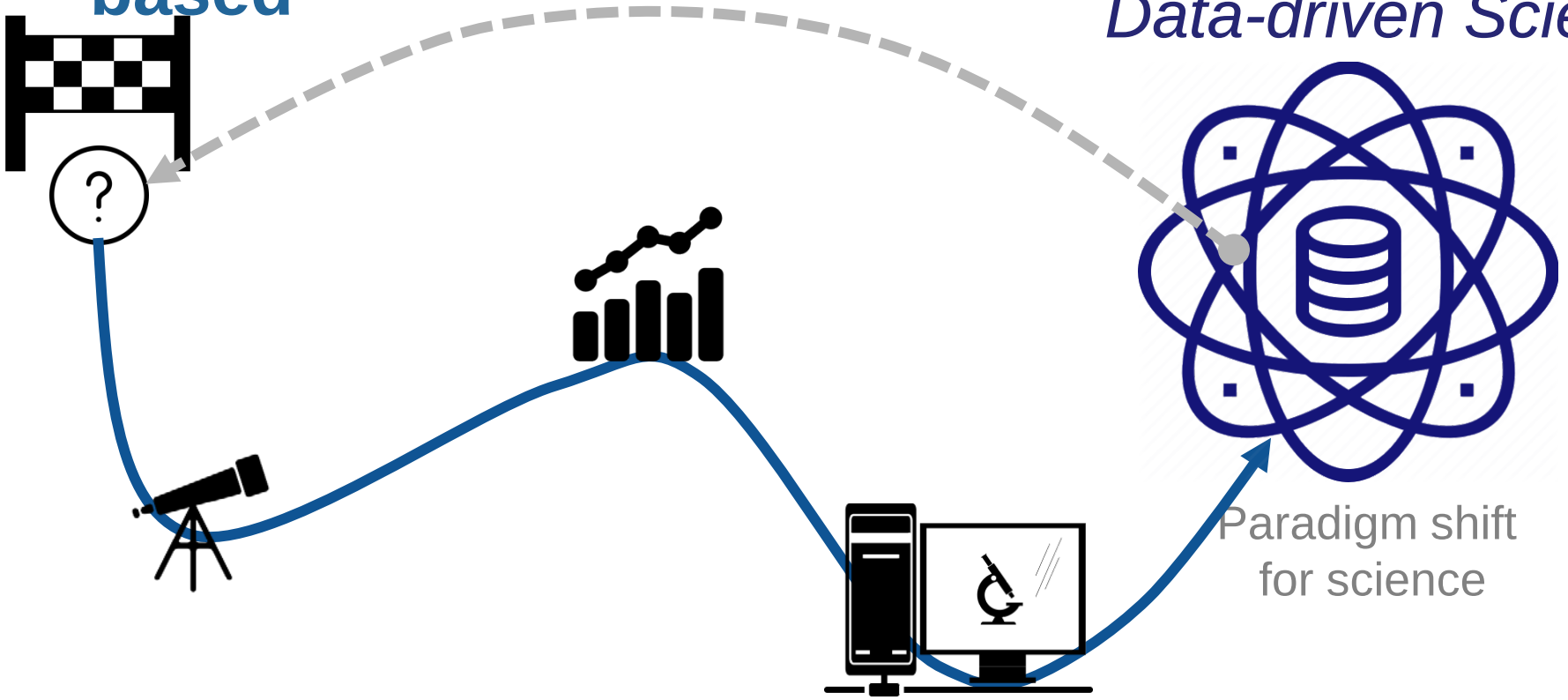
Creating a Data Culture:
The European Open Science Cloud

DATA-DRIVEN SCIENCE IN EUROPE

It's the data, stupid!

- Every 2 Days we create as much info as we did up to 2003.
- The daily data create by SKA, needs 2 million y's to playback on an i-pod…
- And this is only the beginning: with the Internet of Everything (> IoT) everything becomes a data point and hence a potential empirical basis to a scientific problem

Science has always been data based



A 4th paradigm for science?

Due to data abundance: Greater role for inductive, not only hypothesis driven science: "Here's the evidence, now what is the hypothesis?"

	Manual	**Computational**
Deductive	2nd paradigm: theoretical (*Newton)*	3rd paradigm: computational (*Von Neumann*)
Inductive	1st paradigm: empirical (*Bacon*)	4th paradigm: data-intensive (*Venter, DNA sequencing*)

Why is DDS on the Policy agenda?

- Data-driven science is the application of Big Data in Science
- It challenges the established ways of organising scientific work (a new modus operandi emerging)

For policy it means that we are dealing with two categories:

- Data related policy issues
- Computer infrastructure, software and skill policy issues

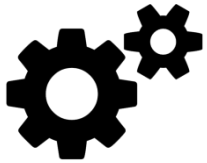
Data related policy isues

- Data generated for research purposes


Open Access to data


Preservation of data


Discoverability of data


Interoperability of data

- Data not generated for research purposes
  - public administration data, health & medical records information, internet/ apps/ social media data etc.
  - Special case: text-mining

Computer, infrastructure, software and skill policies

- Processing and hosting needs: performance at an precedented scale required
- New software to handle big data in volume, variety and velocity
- New skills required: DMP, curation, stewardship etc. PLUS visualisation etc.

→ emerging new field: data science

DDS AS PART OF A SYSTEM CHANGE: OPEN SCIENCE

Top priority

A systemic change in the modus operandi of science and research

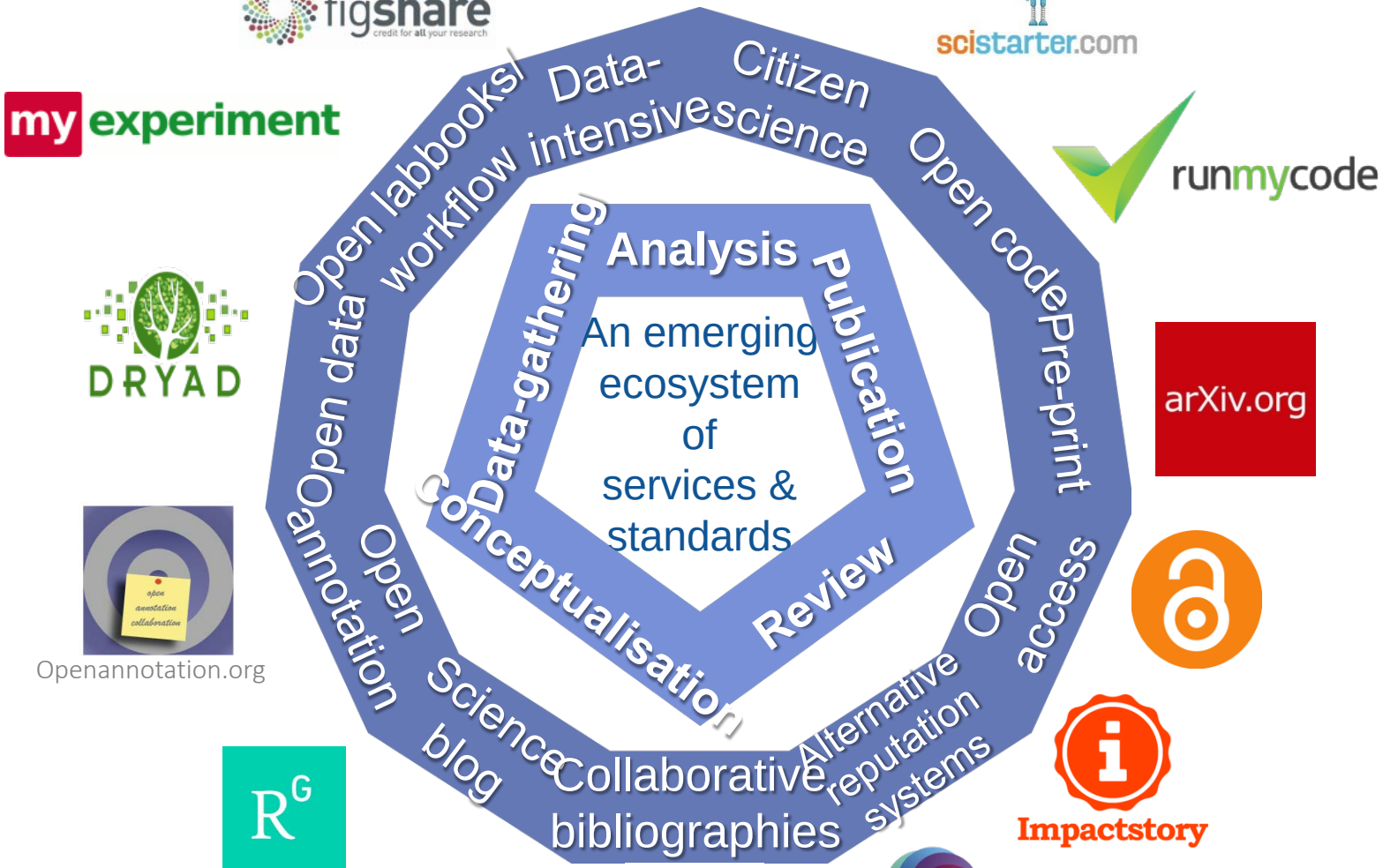
Affecting the whole research cycle and its stakeholders

Commissioner Carlos Moedas
Open Science Presidency Conference
Amsterdam, 4 April 2016

A new approach to the research process

- Systemic change to the way science is organised & research is carried out
- Based on cooperative work and new ways of diffusing and sharing knowledge using digital technologies and new collaborative tools
- It affects virtually all components of doing science & research, from conceptual work to publishing, from empirical research to data-analysis.
- Shifting focus from "publishing as fast as possible" to "sharing knowledge as early as possible"

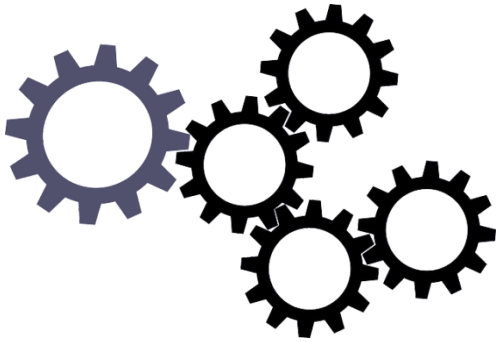
It's real!


figshare
credit for all your research
scistarter.com
my experiment
runmycode
DRYAD
arXiv.org
Openannotation.org
Impactstory
Altmetric
ACADEMIA
MENDELEY
Open labbooks/ workflow
Data-intensive
Citizen science
Open code
Pre-print
Open access
Alternative reputation systems
Collaborative bibliographies
Science blog
Open annotation
Open data
Data-gathering
Analysis
Publication
Review
Conceptualisation
An emerging ecosystem of services & standards

Irreversible

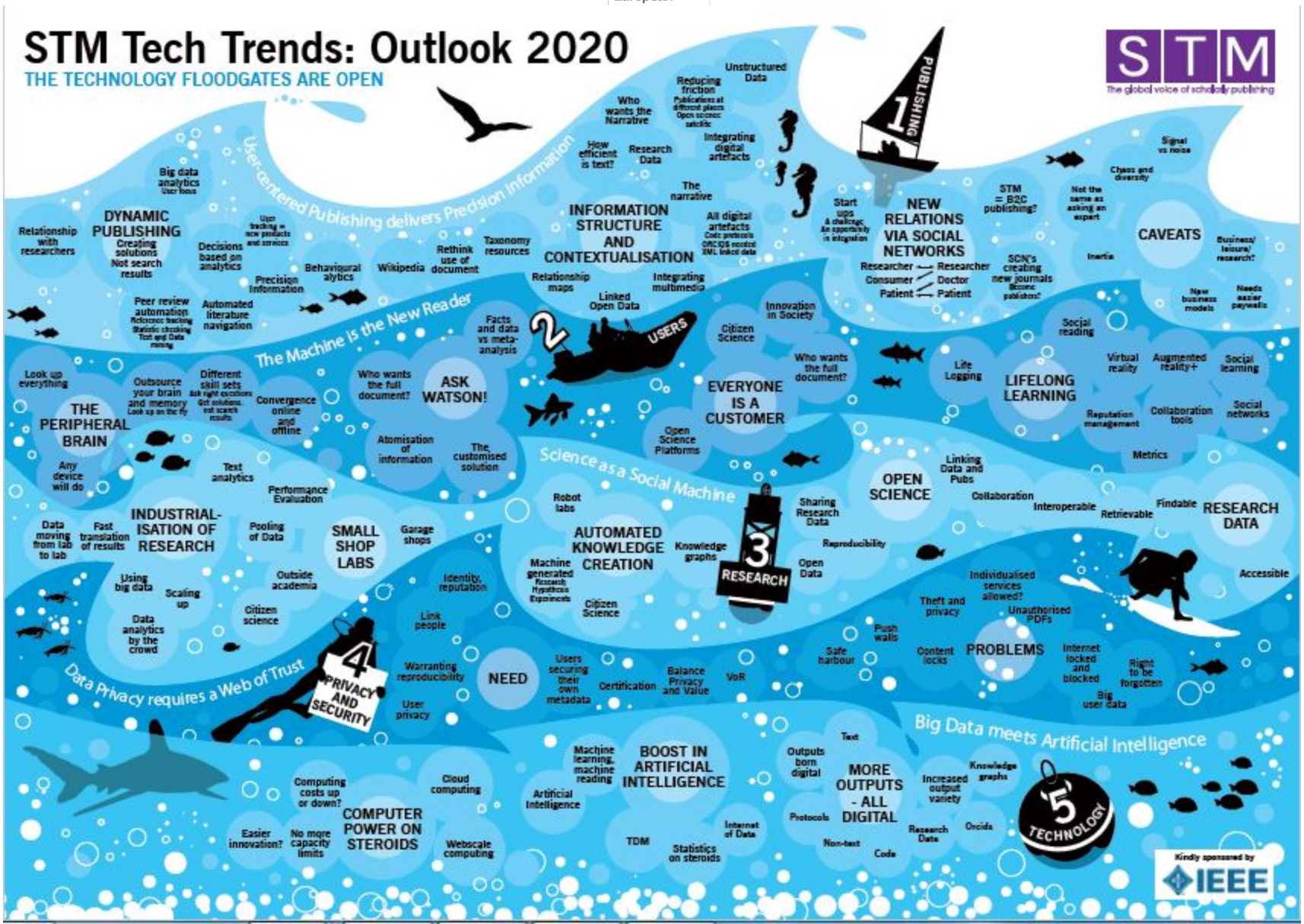
- Digital technologies
- Exponential growth of data
- More researchers and research institutions
- Increase in the scientific production
- Grand Challenges
- Expectations and involvement of citizens
- Accountability, responsiveness and transparency
- Digital "natives"

Not happening in isolation

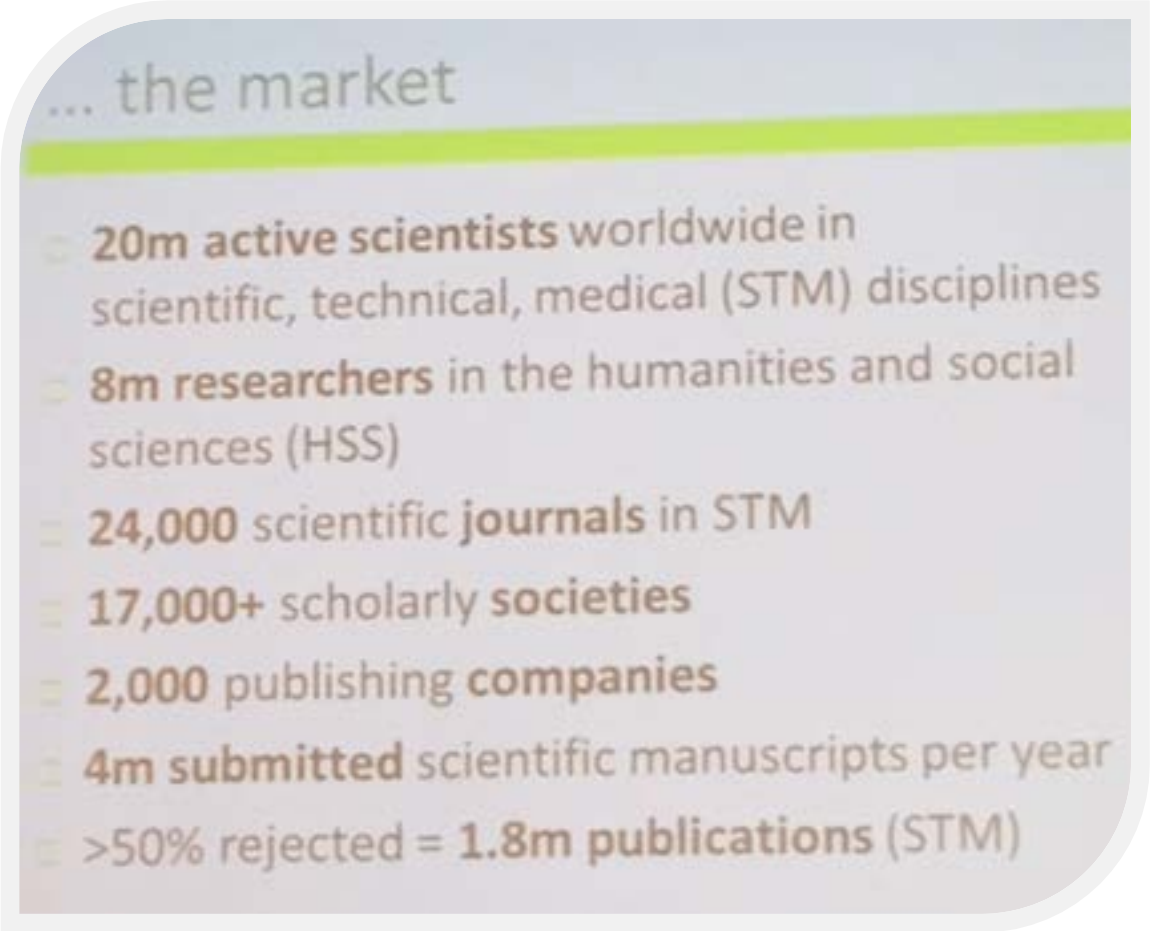


- Open source software
- MOOCs
- Collaborative knowledge production
- Creative commons
- Open innovation
- The sharing/collaborative economy
- Web 2.0

SHARED ANALYSIS



A lot at stake



In total in 2014, LERU members alone had an economic impact across Europe of:

€71.2 billion GVA

900,000 jobs

Study by **BiGGAR economics**

Opportunities for all Europe

Better value for money by strengthening the productivity of the European science and research system through the uptake of results by businesses, in particular SMEs that may not have the resources to pay for access to research results

More transparency, openness and networked collaboration leading to a higher degree of responsiveness of the research community to societal challenges

A sound science and society relationship: more openness may also lead to more trustworthy science from the point of view of the citizen and civil society organisations (NGOs)

Big and open data are estimated to add 1.9% of EU-28 GDP by 2020.

Opportunities for Researchers

Wider dissemination & sharing of the results

Easier Involvement in more **interdisciplinary research**

More visibility and credit for those collecting & sharing underlying research data

Involvement in international networks

New and better calibrated career paths (data scientists, start-ups, altmetrics)

CREATING A DATA CULTURE: THE EUROPEAN COMMISSION'S POLICIES ON OPEN ACCESS TO PUBLICATIONS & RESEARCH DATA

Our Objective

Optimise the impact of publicly-funded scientific research

- *At European level (FP7 & Horizon 2020)*
- *At Member State level*

Leading by example on Open Access (Horizon 2020)

- Open access to scientific publications resulting from publicly funded research under Horizon 2020 shall be ensured [...].
- Open access to research data resulting from publicly funded research under Horizon 2020 shall be promoted. [...]."

OA to publications
FP7 to Horizon 2020

Pilot → Underlying

FP7

- **Green** OA pilot in 7 areas of FP7 with 'best effort' stipulation
- Allowed embargoes: 6/12m **Gold** OA costs eligible for reimbursement as part of the project budget while the project runs

*Outcome: 54% of all scientific peer reviewed publications produced during the lifetime of FP7 open access (no data yet for Horizon 2020).**

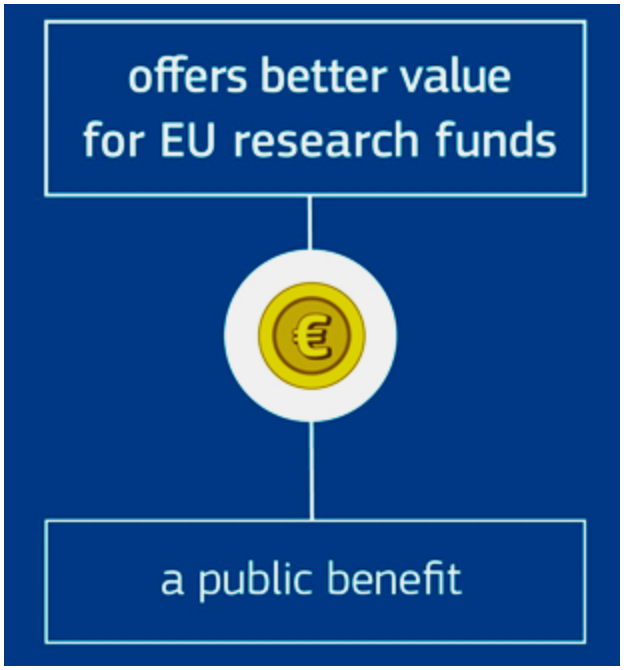
**Data from the OpenAIRE project based on of 171,258 FP7 publications*

Horizon 2020

- Obligation to provide OA, either through the **Green** or **Gold** way in all areas (deposition mandatory either way)
- Allowed embargoes: 6/12m **Gold** OA costs eligible for reimbursement as part of the project budget while the project runs & post-grant support being piloted through OpenAIRE
- Authors encouraged to retain copyright and grant licences instead

Open Research Data in H2020: The Challenge

Wider access to scientific facts & knowledge helps researchers, innovators & the public find & re-use data, and check research results:



Open Research Data in H2020: The Solution

European Commission

Horizon2020 already mandates OA to all scientific publications.

From 2017 research data is open by default with possibilities to opt out.

*As open as possible, as closed necessary…**

Open Research Data in H2020

European Commission

* ***As open as possible, as closed as necessary…***

… is an approach tested during an Horizon2020 pilot action

Basis: 3699 Horizon 2020 signed grant agreements

Calls in core-areas: 65,4% stay in, opt out 34,6%

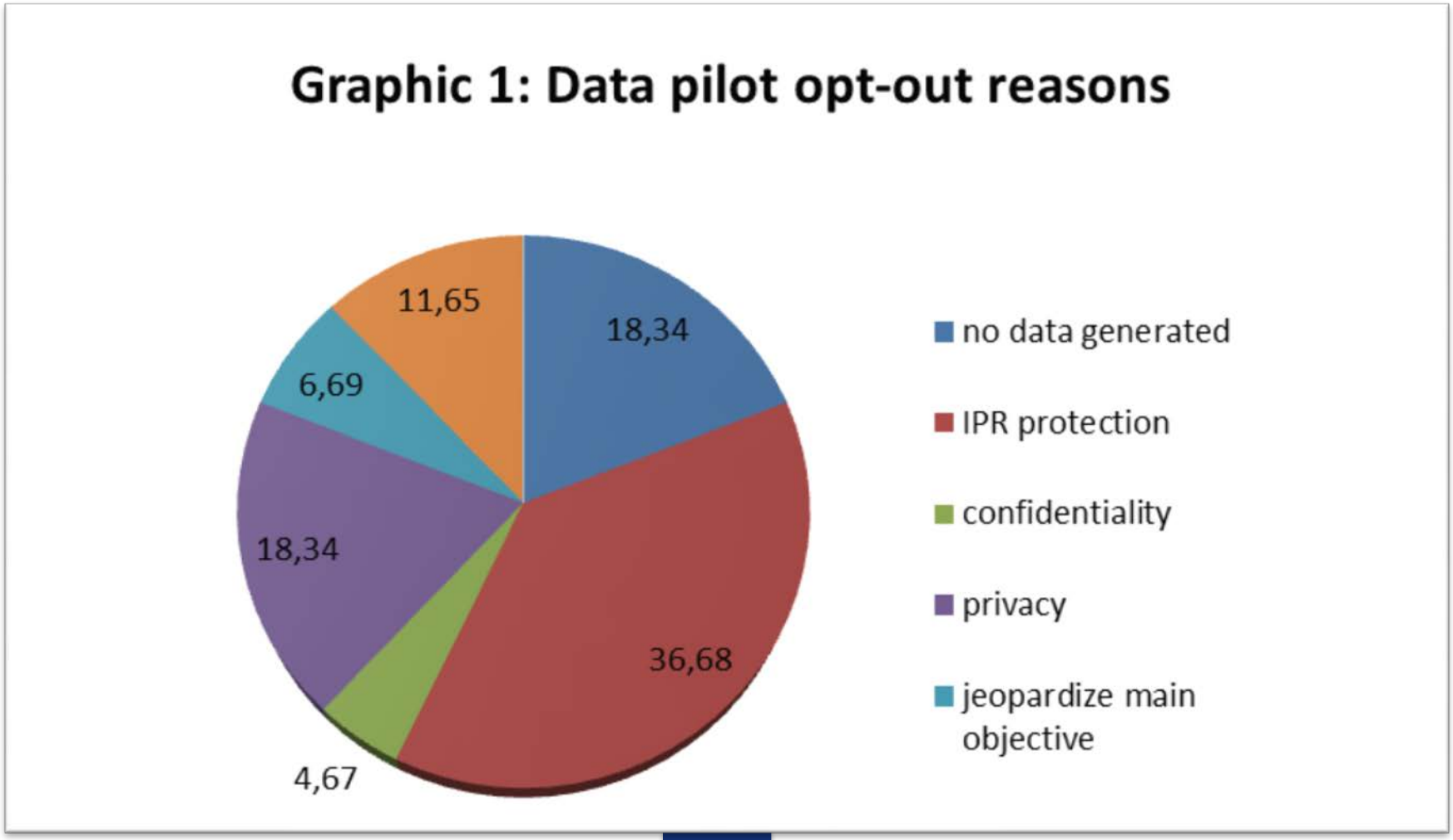
Other areas: voluntary opt in 11,9% (409/3268)

Open Research Data Pilot in H2020

Projects may opt out in a series of cases, at any 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)

ORD Pilot: opt-out reasons among proposals



Research Data Open by Default

Data management costs are eligible for funding

No repository imposed: deposit data where you want

H2020 Grantees are required to:

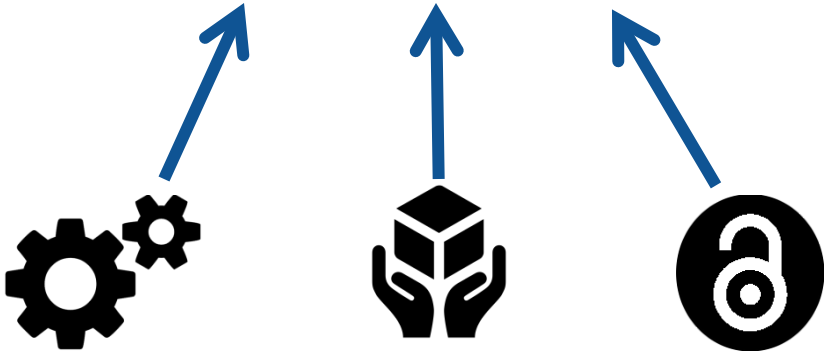
- Take measures to ensure OA to the data underlying publications
- Provide open access to any other research data of their choice
- Are encouraged to also share datasets beyond publication

OA DATA Requirements

Projects must have:

DATA MANAGEMENT PLAN (DMP)

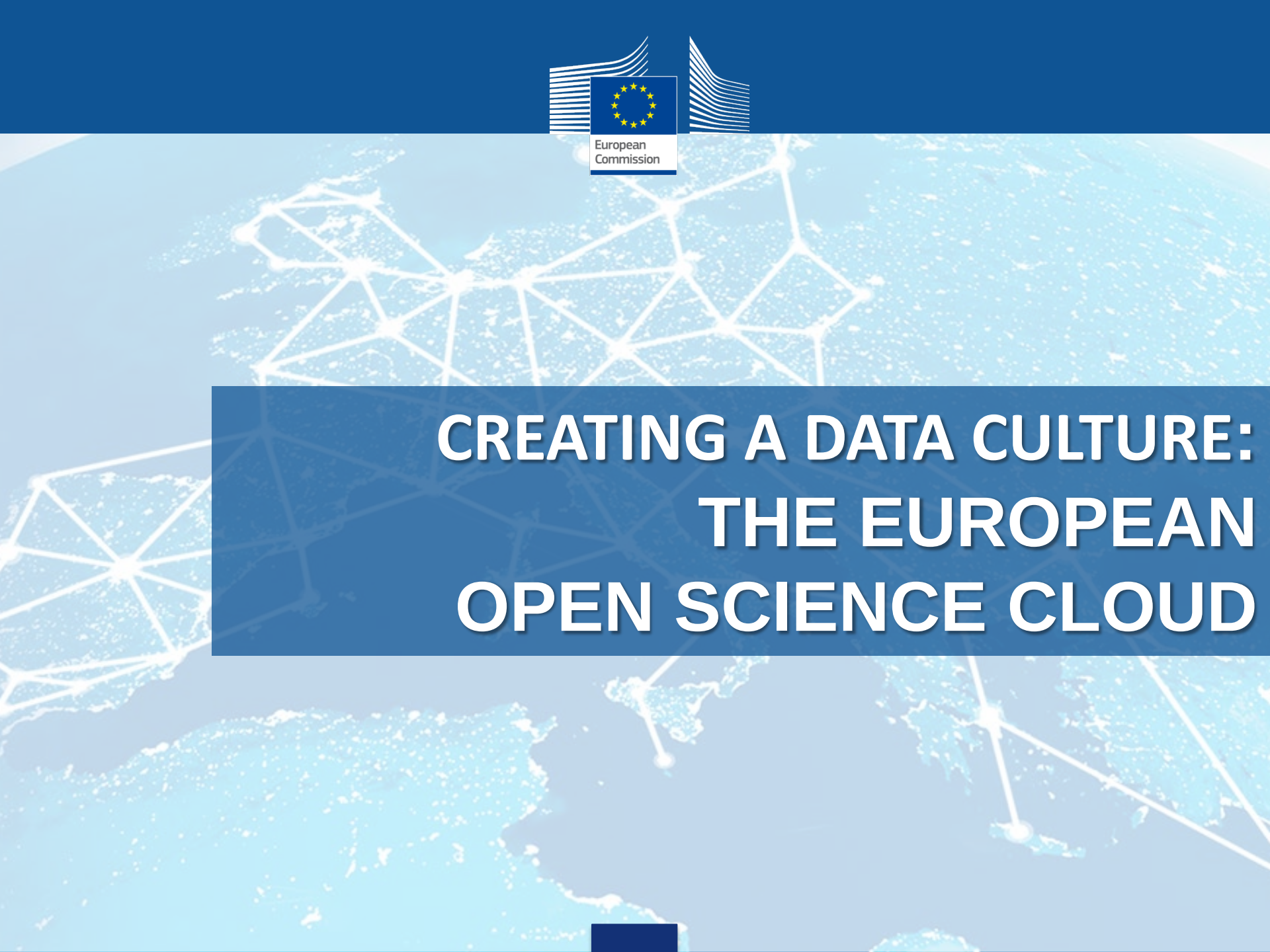
Providing information on:



Data that will be generated

How to ensure curation, preservation, sustainability

What parts of the data will be open and how



CREATING A DATA CULTURE: THE EUROPEAN OPEN SCIENCE CLOUD

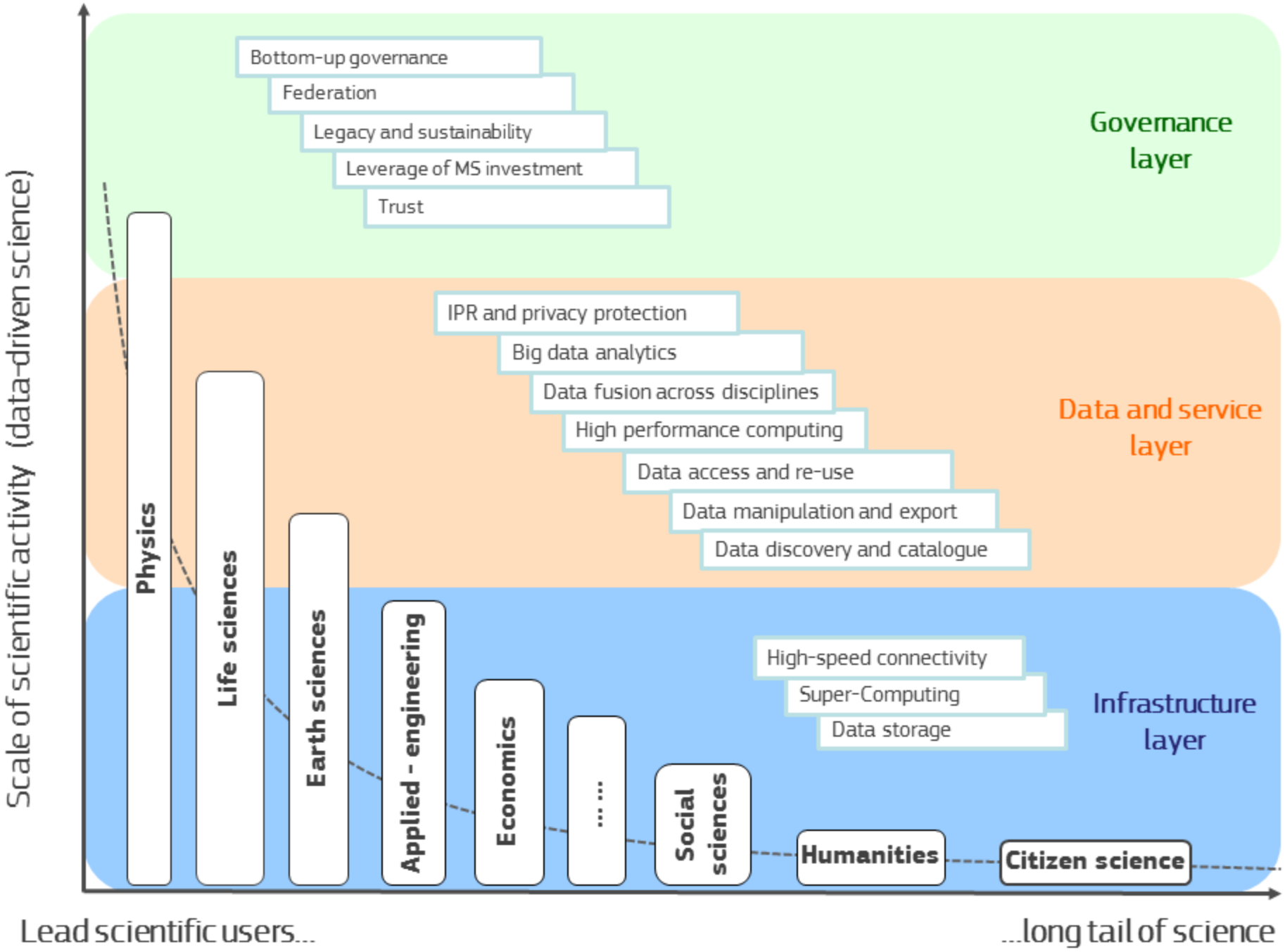
European Open Science Cloud is part of Europe´s ambition to **support the transition to Open Science** and make the most of **data-driven science**.

- Virtual environment **for all European researchers** to store, manage, analysis and re-use data

- **Strongly stated need:** cost-effective, privacy and IPR-conscious.

- **Federation of existing and emerging** data infrastructures

- **Added value:** scale, data-driven science, interdisciplinarity, data to knowledge to innovation


Scale of scientific activity (data-driven science)
Governance layer
Bottom-up governance
Federation
Legacy and sustainability
Leverage of MS investment
Trust
Data and service layer
IPR and privacy protection
Big data analytics
Data fusion across disciplines
High performance computing
Data access and re-use
Data manipulation and export
Data discovery and catalogue
Infrastructure layer
High-speed connectivity
Super-Computing
Data storage
Physics
Life sciences
Earth sciences
Applied - engineering
Economics
... ...
Social sciences
Humanities
Citizen science
Lead scientific users...
...long tail of science

Publication of the HLEG EOSC report (11 October 2016)

- Publication of First report by the Commission High Level Expert Group on the European Open Science Cloud
- Including recommendations on Policy, Governance and Implementation

http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

EOSC Governance: Eco-system functioning

Innovative Funding & Business Models

RECOMMENDATIONS INCLUDED IN THE GOVERNANCE MODEL

OBJECTIVES OF GOVERNANCE

ACTORS & ROLES:

- EOSC MEMBERS
- EOSC FUNDERS
- GOVERNORS
- ASSOCIATES
- USERS
- SECRETARIAT

LOGISTICS & OPERATIONS (PROCUREMENT)

DECISION STRUCTURE & FLOWS

- Funding Models
- Business Models

RULES OF PARTICIPATION

SECURITY ASPECTS & BUSINESS CONTINUITY

INTERNATIONAL AGREEMENTS

Open access and EOSC part of a broader co-developed European Open Science policy

Extensive stakeholder consultation

- ✓ Public consultation (July-September 2014)
- ✓ Validation workshops (October-December 2014)
- ✓ Final report (February 2015): http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf

8 key policy priorities

Strong support by MS and Council

- ✓ Policy debate & Council conclusions 'data-driven economy' May 2015
- ✓ Presidency conference Open Science & Council conclusions 'open science') May 2016

Reflected in the Commission top priorities and actions

- ✓ Included in the Digital Single Market strategy May 2015
- ✓ European Open Science Agenda May 2015
- ✓ High Level Expert Groups on 8 Action Lines
- ✓ Open Science Policy Platform

Still a long way to go for a Data culture

1. Reward systems
2. Measuring quality & impact: altmetrics
3. Changing business models for publishing
4. FAIR open data
5. Open Science Clou
6. Research integrity
7. Citizen Science
8. Open education & skills

Contribute!

Thank you

Jean-Claude.BURGELMAN@ec.europa.eu