

Policy Steps Towards an Open-Data Research Environment

Juan Bicarregui

STFC, Head of Data Division

RCUK, Chair of Data Policy subgroup

RDS, Co-chair of Organisational Advisory Board

GSO, UK representative on Data Working Group

(RDMF 15, April 2016)

Contents

- Global Things
- European Things
- UK Things

Communique of G7 Science Ministers

October 2015

The GSO:

- **Group of Senior Officials on Global Research Infrastructures**
- established in 2008 at the G8 Ministerial meeting, Okinawa

GSO was established to:

- provide a non-binding and open forum for policy exchanges on GRIs and to inform and improve international cooperation;
- share information about existing and planned new RIs;
- establish principles for the development of new partnerships and collaborations.

2013 - G8 Ministerial Statement

London, June 2013

1. Global Challenges

- Antimicrobial resistance

2. Global Research Infrastructures

- Renewed mandate for GSO

3. Open Scientific Research Data

- Next slide

4. Open access to publications

- Another talk...

G8: Open Scientific Research Data

2013

- i. To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.*
- ii. Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.*
- iii. To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.*
- iv. To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.*

2015 G7 Ministerial Communiqué

Berlin Oct 2015

1. Neglected tropical diseases
2. Future of the Seas and Oceans

3. Global Research Infrastructures (GRIs)

“...[4 items about Global (physical) Research Infrastructures]...

- *Further progress on sharing and managing scientific data and information should be achieved, especially by continuing engagement with community based activities such as the **Research Data Alliance RDA**.*
- *We encourage the GSO to continue their work on convergence and alignment of inter-operable data management that could accomplish an **effective open-data science environment** at the G7 level and beyond.”*

4. Clean Energy

The Research Data Alliance



Created 2.5 years ago by merging initiatives in the EU and US

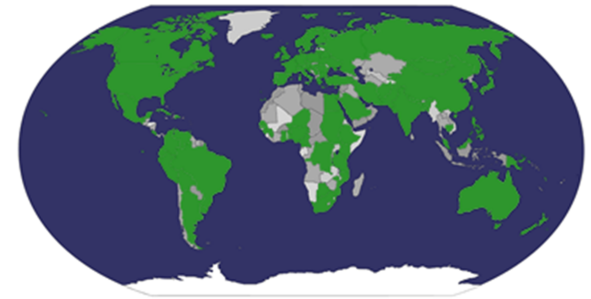
Vision

Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the social and technical bridges that enable open sharing of data.

- Now ~4000 members from 110 countries
 - UK membership is ~400 (second only to the US)
- ~60 Working Groups
 - Domains focused
 - Structural biology, Marine, Agricultural, Materials, Urban,
 - Technical focused
 - Reproducibility, Provenance, Terminology, Policies, Ids
 - Community focused
 - Legal, Data citation, Education and training,
- CODATA and Science Europe are also relevant



700 Members attended the RDA Plenary 6

An “Open-data Science Environment”

In Europe this is embodied in *The European Open Science Cloud*:

- *Open* in two senses: (Open-Science) Cloud & an Open (Science-Cloud)
- *Science* includes research in all domains, public and private
- *Cloud* in the sense of virtualised, transparent, service oriented,...

2014 EC Communication: *A Digital Single Market Strategy for Europe*:

- European Free flow of data initiative
- Seamless, interoperable digital services
- Digital *ecosystems* of hardware, software, applications and data
- Removing technical and legislative barriers to data driven science
- Enabling new services for data-driven science through open systems and services and cross-border flow of data
- Data as a catalyst for economic growth and innovation
- **A transition towards more efficient Open Science**

Now expanded in Communication on European Cloud Initiative, April 19 2016

Contents

- Global Things
- European Things
- UK Things

EC Communication on European Cloud Initiative [April 19, 2016]

- Part of Communication package on *Digital Single Market Technologies and Public Service Modernisation Package*:
- Builds on previous Communications eg *Communication and Recommendation* on access to and preservation of scientific information (2012):
 - Ensure that: “*research data that result from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures*”
- Covers 3 topics:
 1. European Open Science Cloud
 2. European Data Infrastructure
 3. Widening Access and Building Trust

European Cloud Initiative

A trusted, open environment for storing, sharing and re-using scientific data and results and supporting Open Science practices.

- A virtual environment for all European researchers to store, manage, analysis and re-use data.
- Strongly stated needs: cost-effective, user-driven, privacy and IPR-conscious.
- Bringing together existing and emerging data infrastructures.
- Added value: scale, data-driven science, inter-disciplinarity, data to knowledge to innovation.

It will provide 1.7m EU researchers an environment with free, open services for data storage, management, analysis and re-use across disciplines.

5 reasons why Europe is not yet fully tapping into the potential of data

- data from publically funded research is not always open and lack of incentives and rewards for data sharing
- lack of interoperability required for data sharing ... noting deep-rooted walls between disciplines.
- Fragmentation between data infrastructures ...split by scientific and economic domains, countries and governance models and have different access policies
- A surging demand for High Performance Computing at scale ... in a competitive time frame.
- reuse data employing advance analysis techniques in a dependable environment that ensures adequate protection of personal data considering forthcoming revision of Copyright legislation.

European Open Science Cloud (EOSC)

To develop EOSC it will be necessary to:

- **Make all scientific data produced by the Horizon 2020 programme open by default.**
- Raise awareness and change incentive structures for academics industry and public services to share their data.
- Develop specification for interoperability and data sharing across disciplines and infrastructures
- Create a fit-for-purpose pan-European governance structure to federate scientific data infrastructures and overcome fragmentation.
- Develop cloud based services for Open science supported by the necessary data infrastructure
- Enlarge the scientific user base to researchers and innovators from all disciplines.

European Data Infrastructure (EDI)

will underpin the European Open Science Cloud with:

- data infrastructures which store and manage data;
- high-bandwidth networks which transport data;
- ever more powerful computers which can be used to process the data.

Aside: High Performance Computing

*“The European Data Infrastructure will also support the EU to rank among the world's top supercomputing powers by **realising exascale supercomputers around 2022, based on EU technology, which would rank in the first 3 places of the world**”.*

High Level Expert Group Report

- Has recommendations on:
 - Policy,
 - Governance
 - Implementation
- For example:
 - *“Make adequate data stewardship mandatory for all research proposals ”*
- Publication delayed until June
- Recommendations available in various talks: eg
 - http://ec.europa.eu/research/participants/data/ref/h2020/other/events/2016-04-15/h2020-infoday-eosc-overview_en.pdf

EOSC Info Day

15 April 2016

‘Governance’ layer of the EOSC

1. Renew/strengthen OA/ORD political commitment at EU+MS level.
2. Develop a business model for the sustainable funding of the development & operation of the science cloud (incl. a market-place to involve private sector).
3. Develop a mechanism to ensure policy steering, incl. the interaction amongst/with research funders and stakeholders.

EOSC Info Day

15 April 2016

‘Software’ or ‘service’ layer of the EOSC

4. Develop common implement principles for ORD cost eligibility and research data stewardship, including DMPs and certification/quality schemes.
5. Define rules of engagement for scientific users of the science cloud and the business providers of solutions (incl. a market-place or platform).
6. Ensure ‘semantic interoperability’ of research data services/clouds/infrastructures before developing a common research data ‘meta-language’ including research data identifiers.

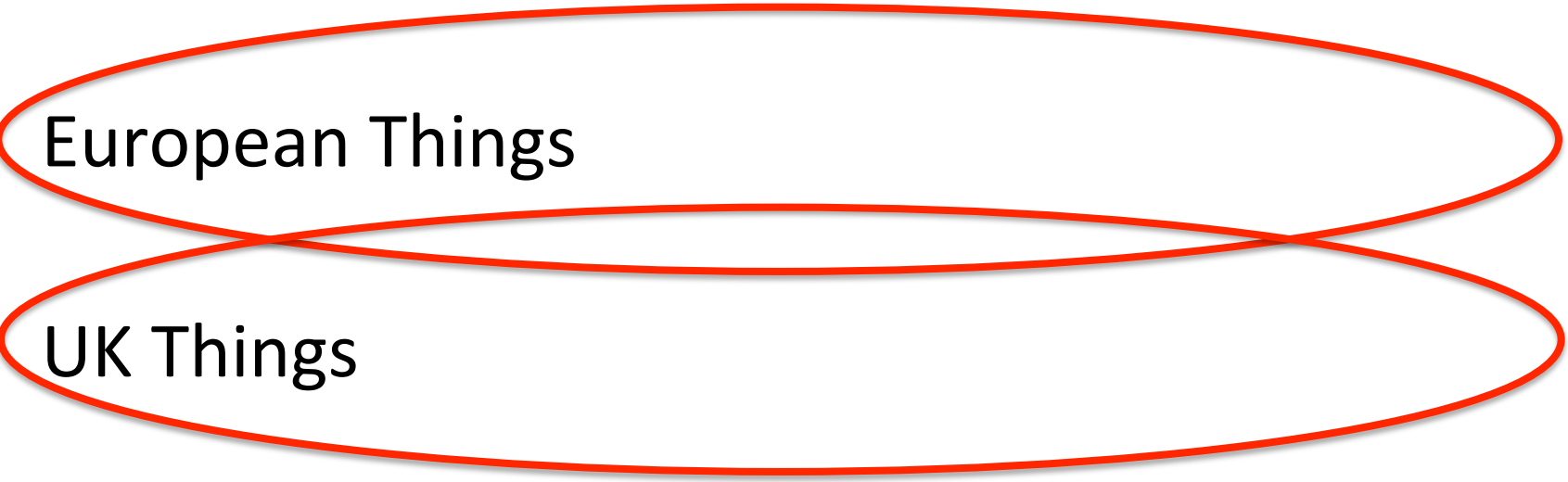
EOSC Info Day

15 April 2016'

'Infrastructure' layer of the EOSC

7. Leverage support to ESFRIs to make them lead customers and solution providers.
8. Continue developing the data infrastructures.
9. Support the clustering of domain-specific science clouds and their interoperability.

Contents

- Global Things
 - European Things
 - UK Things
- 
- Two red hand-drawn ovals are present on the slide. The first oval encircles the text 'European Things' and the second oval encircles the text 'UK Things'. The two ovals overlap in the middle of the list.

RCUK Guidelines (2015)

- RCUK published:
 - **Guidance on best practice in the management of research data** (July 2015)
- Explanatory text for each of the 7 Common Principles from 2011
 - Available from:
RCUK Common Principles on Data Policy webpage
- Input to *Concordat on Open Research Data* which is currently under discussion

RCUK Principles on Data Policy

1) Data are a Public Good

Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner

2) Data should be managed...

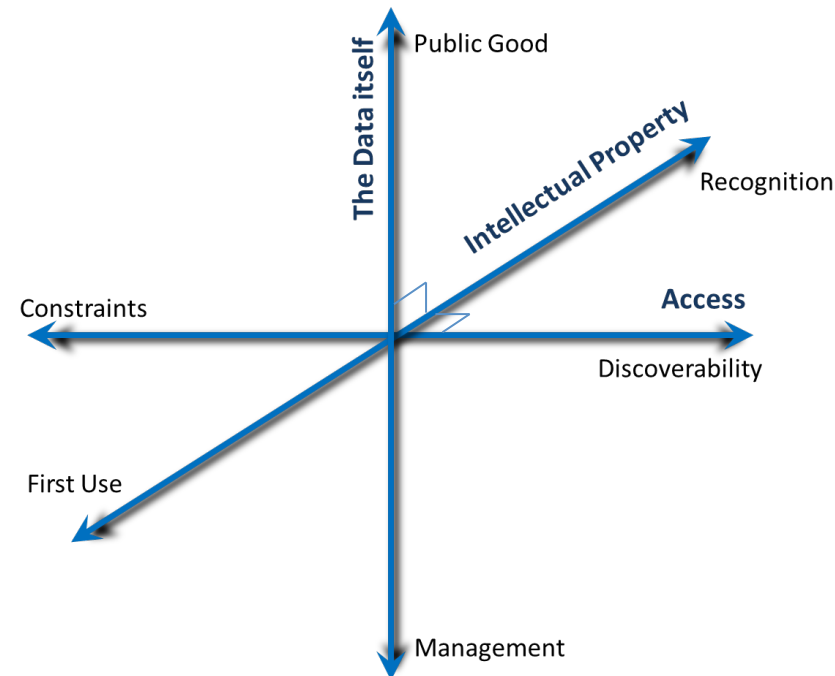
3) Data should be discoverable...

4) There may be constraints...

5) Originators may have first use...

6) Reusers have responsibilities...

7) Data sharing is not free...



Concordat on Open Research Data (2016)

- Developed by UK Open Research Data Forum
 - includes:
 - RCUK, HEFCE, JISC,
 - Wellcome Trust,
 - Universities UK,
 - ...
- *Still under discussion*

Scope of Concordat

Definition of Research Data:

- *the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence.*

Purpose of open research data:

- *to provide the information necessary to support or validate a research project's observations, findings or outputs.*

Applies to all fields of research :

- *Emphasises responsibilities and accountabilities*
- *Recognises the autonomy of researchers*
- *Complements existing frameworks*

Principle #1

Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice.

- **Researchers** will make their research data open and usable within a short and well-defined period, which may vary by subject and disciplinary area
- **Employers of Researchers** will foster a research environment which recognises the value of open data and will seek to provide appropriate access to infrastructure systems and services to enable their researchers to make research data open and usable
- **Funders of Research** will support open research data by appropriately acknowledging and supporting its costs, and by supporting the wider agenda with appropriate policy and investment activities

Principles 2-9

Principle #2

- *There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable.*

Principle #3

- *Open access to research data carries a significant cost, which should be respected by all parties.*

Principle #4

- *The right of the creators of research data to reasonable first use is recognised.*

Principles 2-9

Principle #5

- *Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.*

Principle #6

- *Good data management is fundamental to all stages of the research process and should be established at the outset.*

Principle #7

- *Data curation is vital to make data useful for others and for long-term preservation of data*

Principles 2-9

Principle #8

- *Data supporting publications should be accessible by the publication date and should be in a citeable form.*

Principle #9

- *Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders*

Principle #10

- *Regular reviews of progress towards open research data should be undertaken.*

Contents

- Global Things
 - European Things
 - UK Things
- 

*"By academic freedom,
I understand the right to
search for truth and to
publish and teach what
one holds to be true.
This right implies also a
duty; one must not
conceal any part of what
one has recognized to be
true."*

Albert Einstein
Letter on his seventy-fifth birthday 1954

