



'RDM for All: What does good practice look like?

Juan Bicarregui

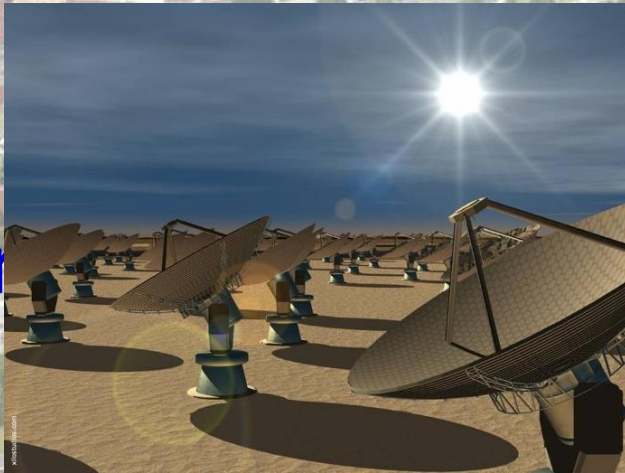
Head of Data Division
Scientific Computing Department
STFC – Rutherford Appleton Laboratory

Research Data Management for All
January 2018



- Who thinks Data management is important?
 - Why is Data Management Important?
- What do Governments say about Data Management?
 - What do Research Councils say?
- Why all the fuss about Open Science?
 - What is the European Open Science Cloud all about?

STFC – Science and Technology Facilities Council



Square Kilometre Array

- Lasers
- Space Science
- Particle Physics
- Computing and Data



des.

source
n Source



Large Hadron Collider



Daresbury Laboratory



ESRF & ILL, Grenoble



Science & Technology
Facilities Council

Scientific Computing Department

- Data handling and Analysis from scientific experiments
- Applications development, optimization and support
- High-performance compute and data facilities and services

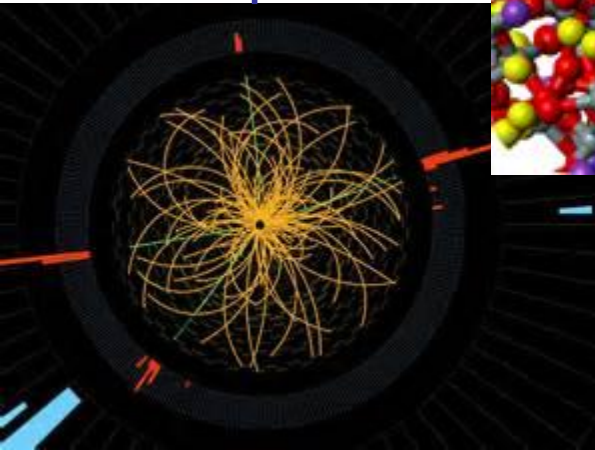
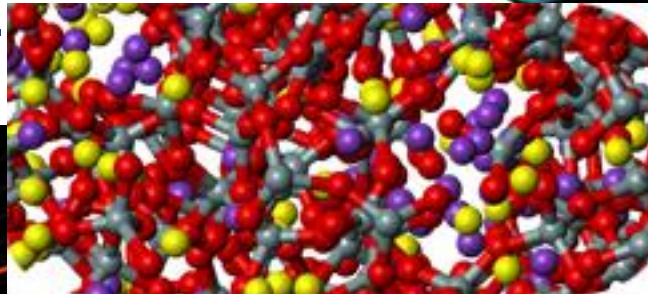
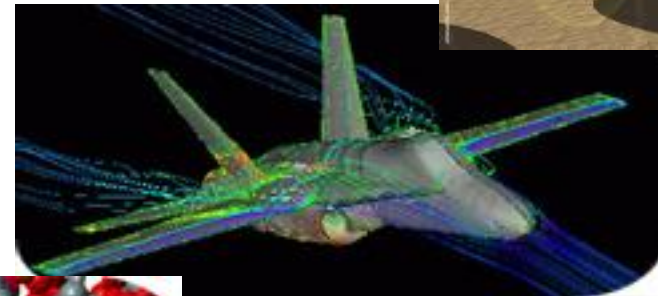
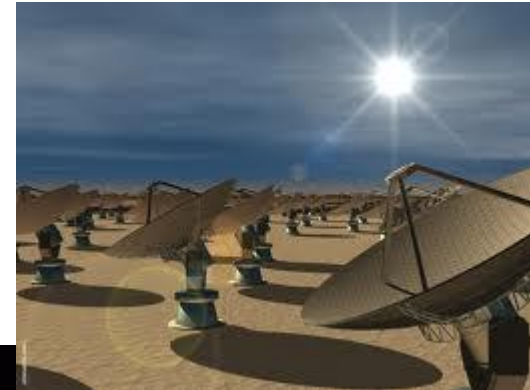
~160 staff

~10,000 users

~ 3500 training days per annum

~100 pub'ns p.a.

~ £12m pa.



*Maximising
the impact of
Scientific Computing
Big Data for Big Science*



- Who thinks Data management is important?
 - Why is Data Management Important?
- What do Governments say about Data Management?
 - What do Research Councils say?
- Why all the fuss about Open Science?
 - What is the European Open Science Cloud all about?



2006, OECD

Recommendation on Access to Research *Data* from *Public* Funding.

2007, European Commission

Recommendation on access to and preservation of scientific information

2007, European Commission

Communication and Conclusions on scientific information in the digital age

2010, HLEG on Scientific Data

Riding the wave: How Europe can gain from the rising tide of scientific data

2011, G8+5

Global Research Infrastructure Group on Data

2012, European Commission

Recommendation on access to and preservation of scientific information

2013, G8 Ministerial Communiqué

“... [publically funded] scientific research data should be open...”

2015, G7 Ministerial Communiqué, October

“...accomplish an effective open-data science environment...”

2016, European Commission

Communication on European Cloud Initiatives

2017, G7 Ministerial Communiqué, September 2017



- Who thinks Data management is important?
 - Why is Data Management Important?
- What do Governments say about Data Management?
 - What do Research Councils say?
- Why all the fuss about Open Science?
 - What is the European Open Science Cloud all about?



Royal Society, *Science as an open enterprise*. June 2012

“Open inquiry is at the heart of the scientific enterprise.

Publication of scientific theories

- and of the experimental and observational data on which they are based –

permits others

- to identify errors,
- to support, reject or refine theories and
- to reuse data for further understanding and knowledge.

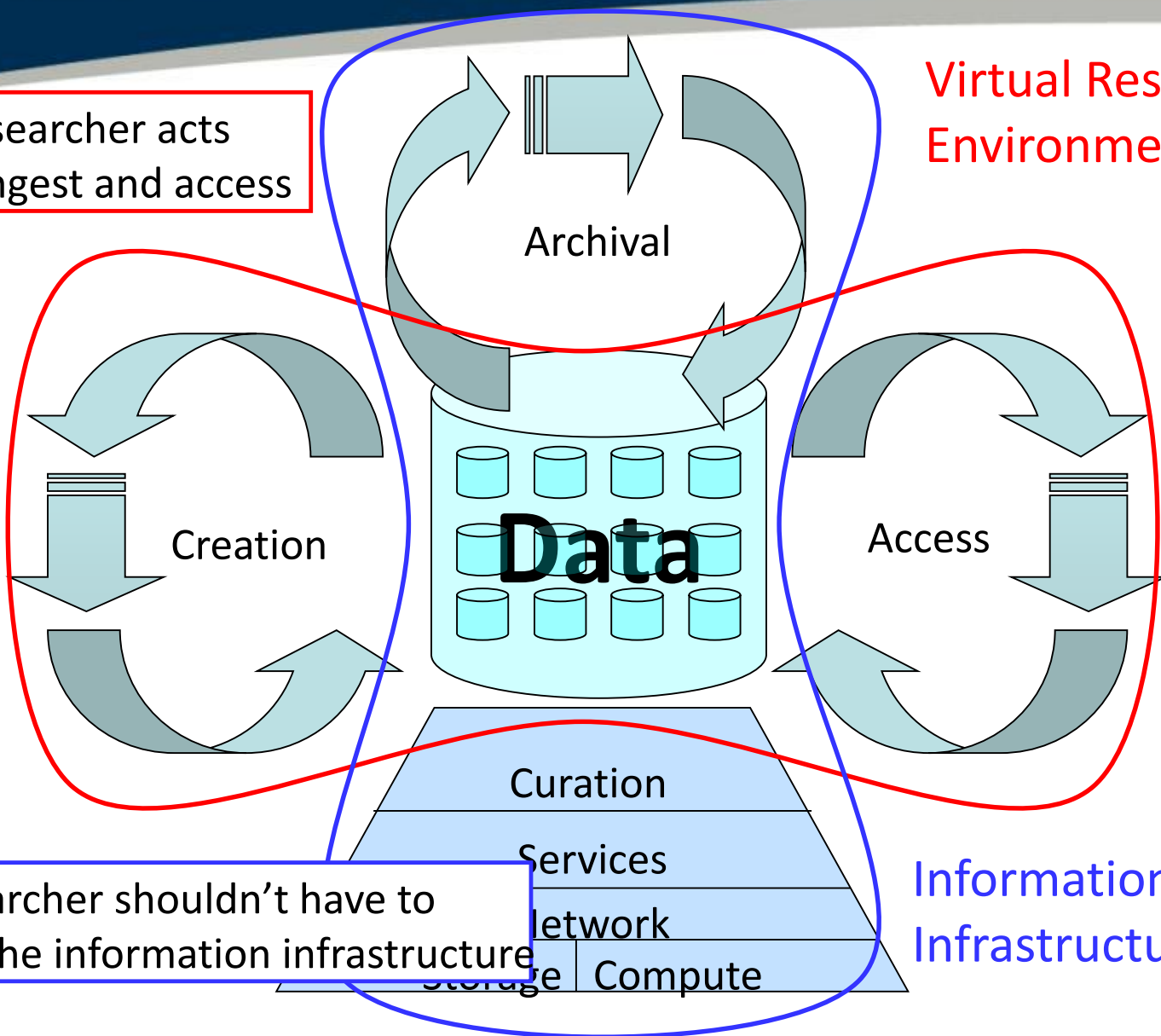
Science’s powerful capacity for self-correction comes from this openness to scrutiny and challenge.”



What is Data Management

the researcher acts
through ingest and access

Virtual Research
Environment



the researcher shouldn't have to
worry about the information infrastructure

Information
Infrastructure



- Who thinks Data management is important?
 - Why is Data Management Important?
- What do Governments say about Data Management?
 - What do Research Councils say?
- Why all the fuss about Open Science?
 - What is the European Open Science Cloud all about?



London, June 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**.*



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

www.rd-alliance.org

*building the social and technical
bridges that enable open sharing of
data*

18 FLAGSHIP OUTPUTS

of which 4 ICT
Technical
Specifications

75 ADOPTION CASES

across multiple
disciplines,
organisations &
countries

89 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

of which 30 WORKING GROUPS
& 59 INTEREST GROUPS

6,283 INDIVIDUAL MEMBERS FROM 132 COUNTRIES

67% Academia & Research
15% Public Administration
11% Enterprise & Industry

43 ORGANISATIONAL MEMBERS & 8 AFFILIATE MEMBERS

RDA



RDA Vision

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

RDA Mission

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



Turin, September 2017

...

Open Science

- ...the evaluation of research careers should *better recognize and reward Open Science activities*.
- ...all researchers should be able to *deposit, access and analyse* scientific data *across disciplines* and *at the global scale*,
- ...research data should adhere to the *FAIR* principles of being *findable, accessible, interoperable, and reusable*.



- Who thinks Data management is important?
 - Why is Data Management Important?
- What do Governments say about Data Management?
 - What do Research Councils say?
- Why all the fuss about Open Science?
 - What is the European Open Science Cloud all about?



- 2011 - RCUK **Principles on Data Policy**
- 2015 - RCUK **Guidance on best practice in the management of research data**
- 2016 – Joint **Concordat on Open Research Data**
 - RCUK, HEFCE, JISC, Wellcome Trust, Universities UK
- 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). ...*
- 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



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...

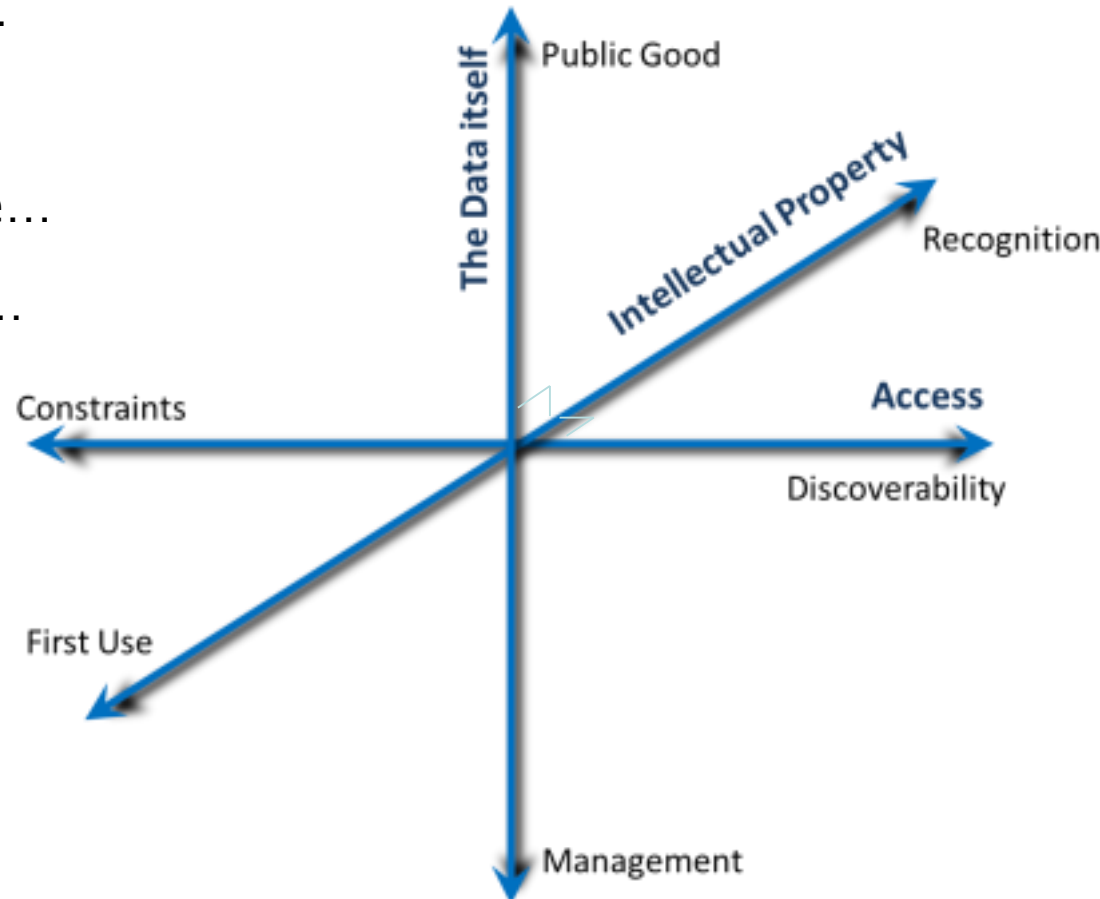
6) Reusers have responsibilities...

7) Data sharing is not free...

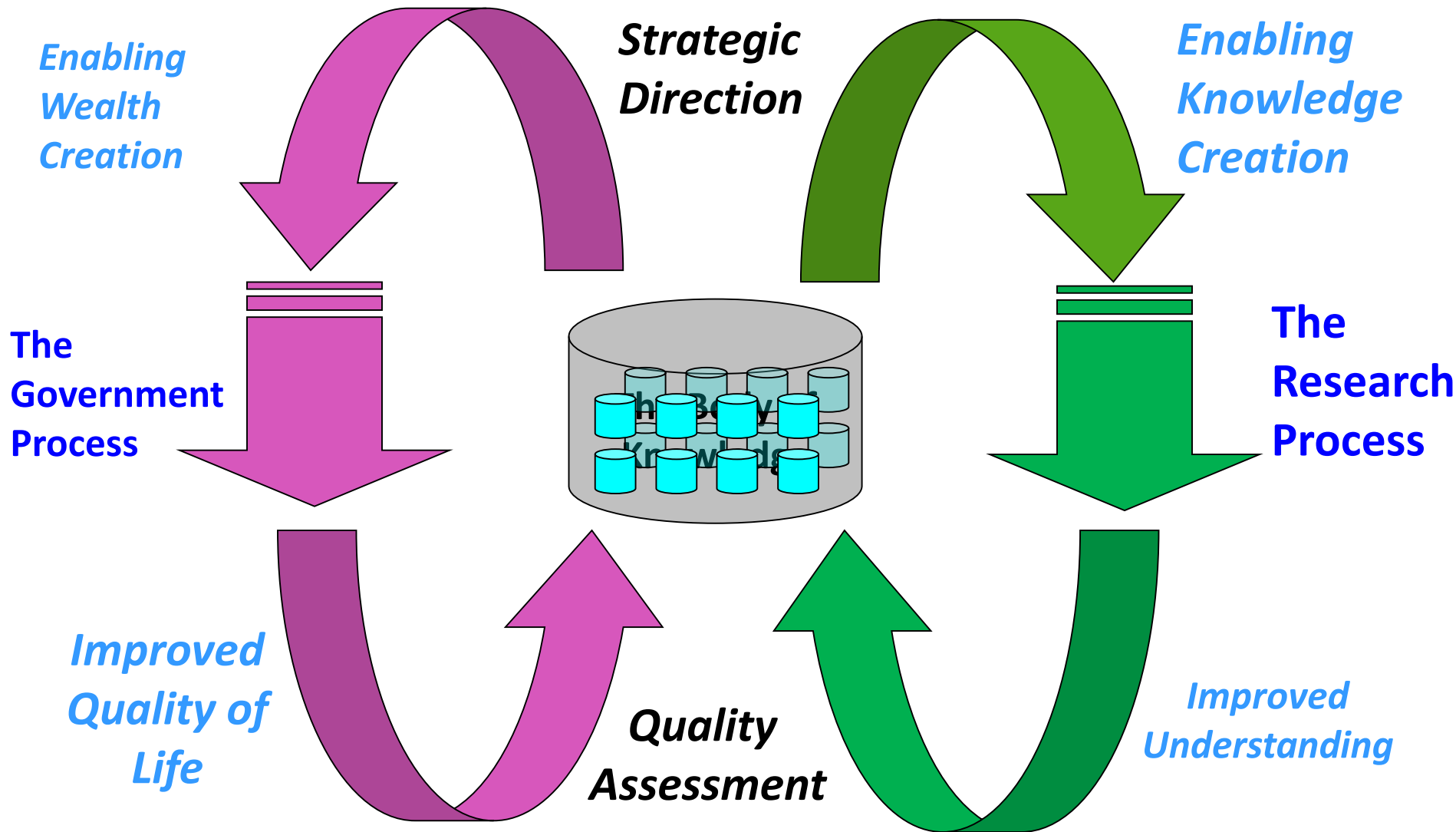


Tensions

- 1) Data are a Public Good
- 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...



The Innovation Lifecycle



Aggregation of Knowledge lies at the heart of the innovation lifecycle



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

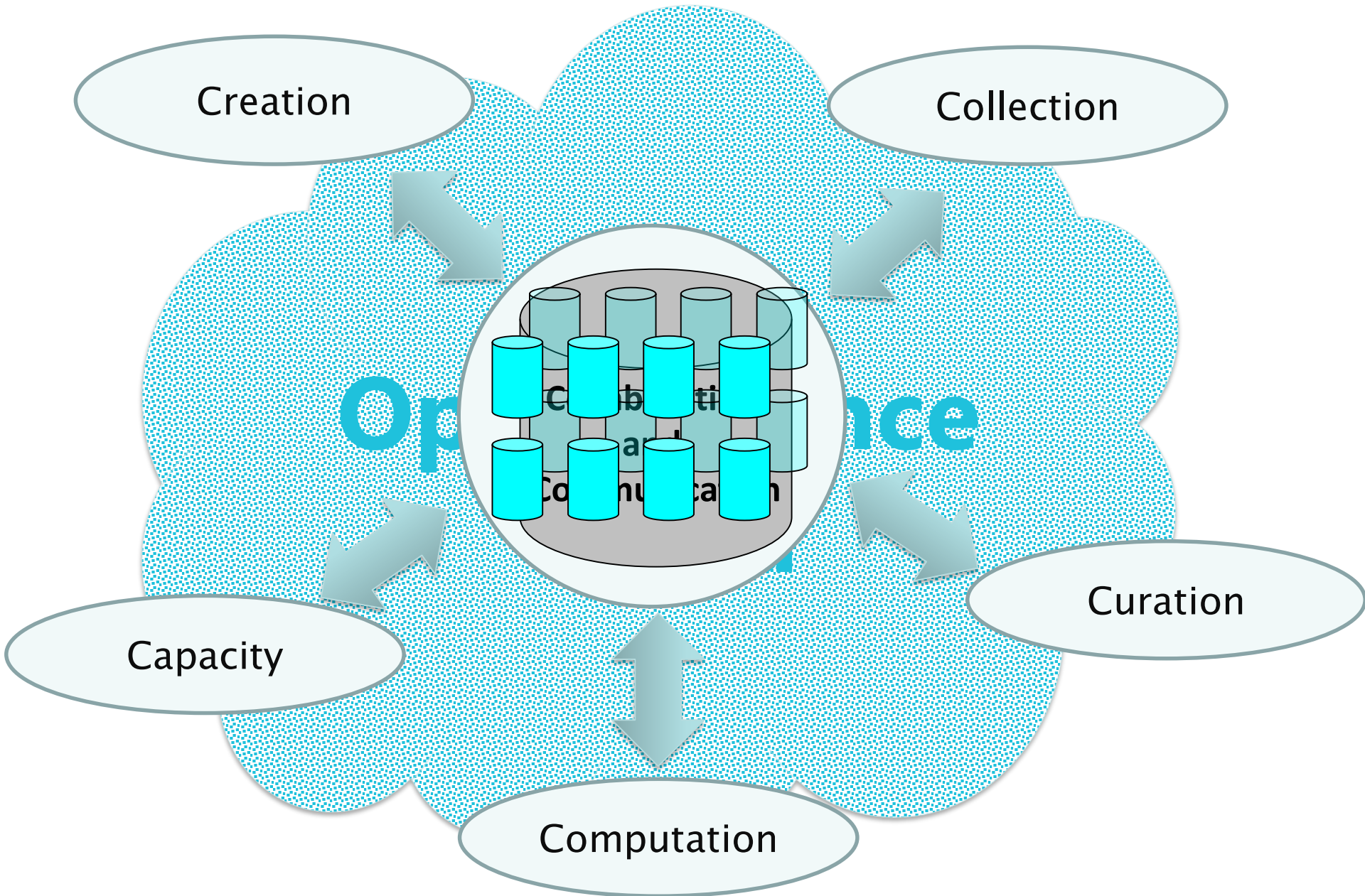
- **Data not always open** and **lack of incentives and rewards** for data sharing
- **Lack of interoperability** required for data sharing ... noting deep-rooted walls between disciplines.
- **Fragmentation between data infrastructures** that are split by scientific and economic domains, countries and governance models
- Surging demand for **High Performance Computing** at a scale above single member state resources
- **Data reuse employing advance analysis techniques** adequate protection of personal data considering forthcoming revision of Copyright legislation.



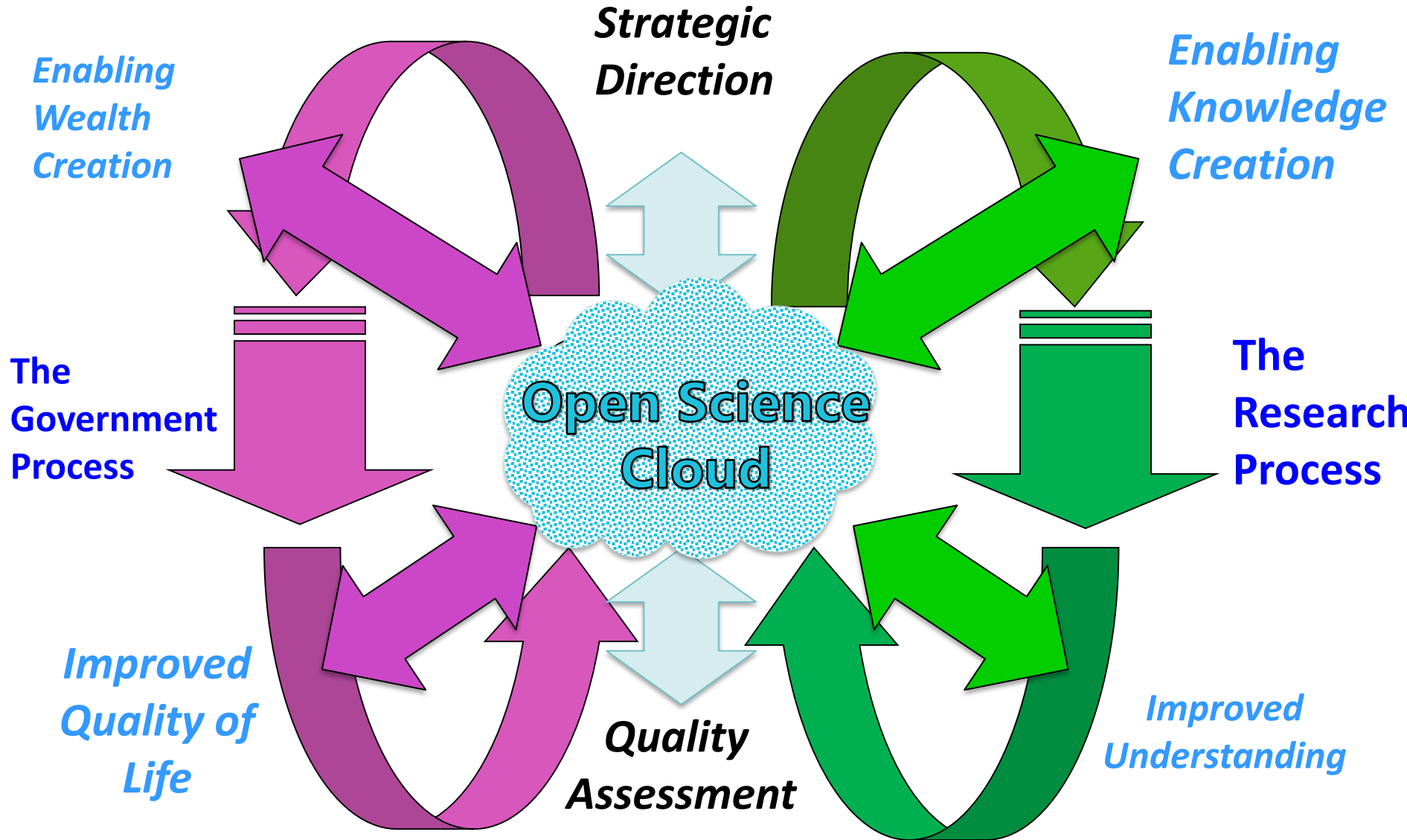
To develop the European Open Science Cloud (EOSC) it will be necessary to:

- **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 specifications 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.

The Innovation Lifecycle



The Innovation Lifecycle



Aggregation of Knowledge lies at the heart of the innovation lifecycle



The *EOSCpilot* project will support the first phase of development of EOSC:

Scientific Challenges are really *Opportunities*

🔗 **Scientific Challenges:** deploying the EOSC to deliver Open Science

Technical Challenges are *Barriers to overcome*

🔗 **Technical Challenges:** developing technical solutions that meet the scientific needs

Cultural Challenges are also *Barriers*

🔗 **Cultural Challenges:** adopting new, more open ways of working



Horizontal and Vertical

The EOSC merges Domain Specific (vertical) and Horizontal e-Infrastructures.....





EOSC*pilot* is a *pilot*

EOSCpilot is just a pilot – it will not build the EOSC

A pilot not a design study

A set of experiments and design proposals

A requirements study?

		2017	2018	2019	2020	2021	2022
EOSCpilot							
EOSC-Hub							
Openaire							
RDA/FREYA/etc							
INFRAEOSC-01							
INFRAEOSC-04							
INFRAEOSC-05 (Gov and Fair)							
INFRAEOSC-02							
INFRAEOSC-05 (RIA)							
INFRAEOSC-06							
Also Thematic Clouds							



- Who thinks Data management is important?
 - Why is Data Management Important?
- What do Governments say about Data Management?
 - What do Research Councils say?
- Why all the fuss about Open Science?
 - What is the European Open Science Cloud all about?



Science & Technology
Facilities Council

"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



An aerial photograph of Berlin, Germany, featuring the Spire of Silence (Fernsehturm) and a river with a boat. The image is overlaid with a semi-transparent dark blue rectangle containing white and yellow text. On the right side, there are several white, wavy, curved lines that sweep across the image.

RDA

11 PLENARY

MEETING

21-23 MARCH 2018

Berlin, Germany



RESEARCH DATA ALLIANCE
RESEARCH DATA SHARING WITHOUT BARRIERS



'RDM for All: What does good practice look like?

Juan Bicarregui

Head of Data Division
Scientific Computing Department
STFC – Rutherford Appleton Laboratory

Research Data Management for All
January 2018