Science & Technology Facilities Council

The European Open Science Cloud for Research

Juan Bicarregui
Head of Data Division, STFC
Co-Chair of RDA Organisaional Asssembly
Coordinator of EOSCpilot project

Cloud Forward
18 October 2016

Science & Technology Facilities Council

Outline

STFC

- Science and Technology Facilities Council
- Scientific Computing Department

Open Science Policy

- UK, EU, G7
- 2004, 2007, 2011, 2013, 2015, 2016…

RDA

- Research Data Alliance

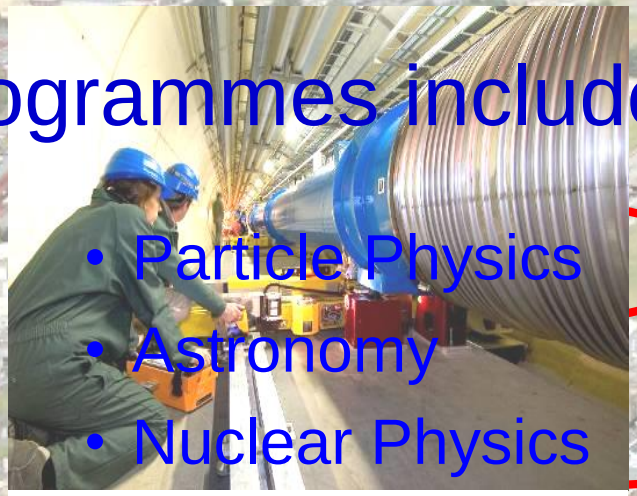
European Open Science Cloud

- EOSC pilot project

The Science and Technolgy Facilities Council

Programmes include:

- Particle Physics
- Astronomy
- Nuclear Physics



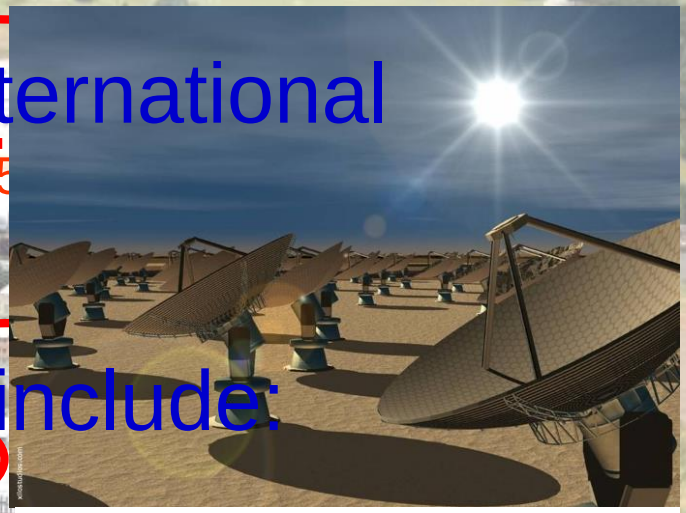
Large Hadron Collider

International

25

include:

- CERN
- ESO
- SKA
- ESS
- ESRF
- ILL



Square Kilometre Array

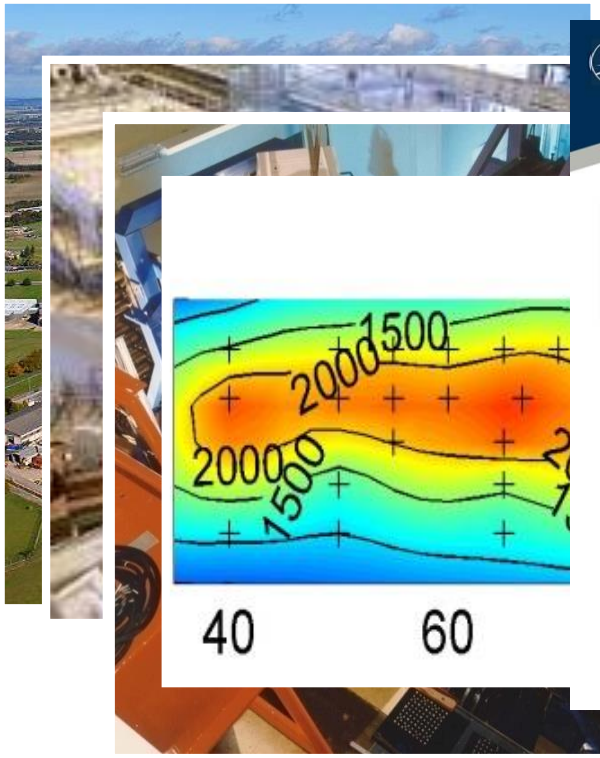
National Laboratories include:

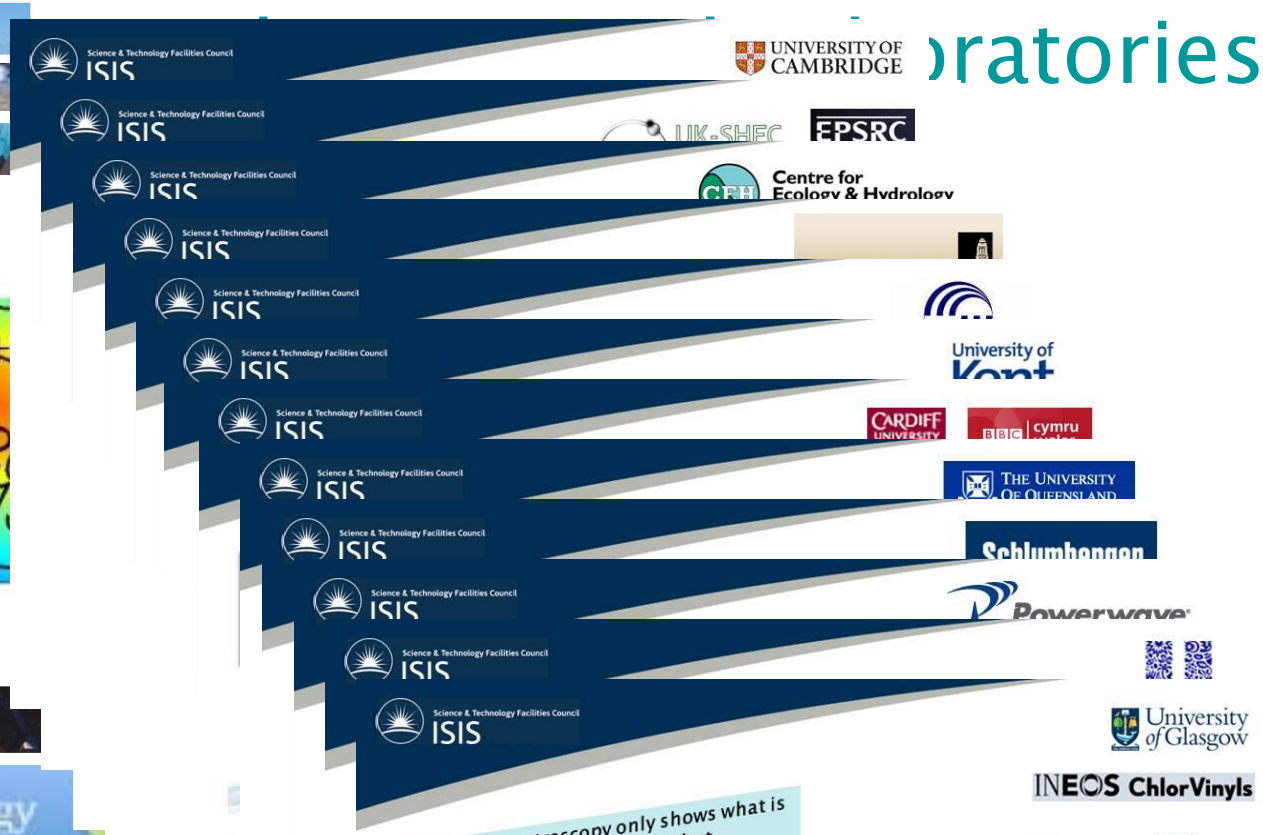
- Neutron and Muon Source
- Synchrotron Radiation Source
- High Power Lasers
- Scientific Computing/Hartree Centre
- Space Science
- Electronics, Cryogenics

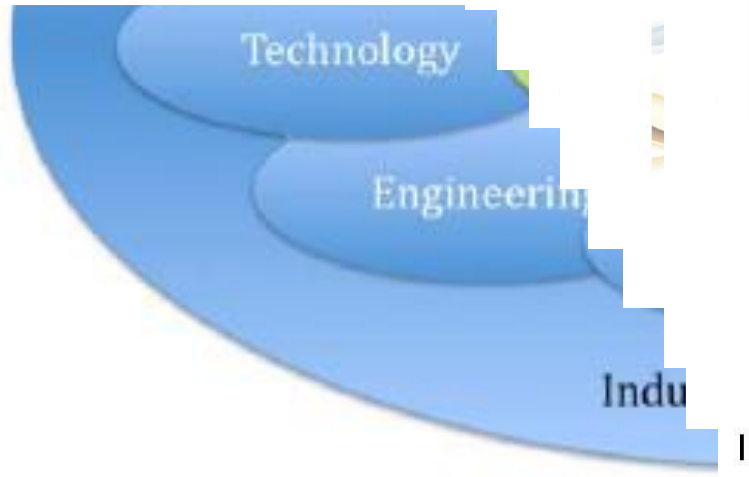
Daresbury Laboratory

ESRF & ILL, Grenoble

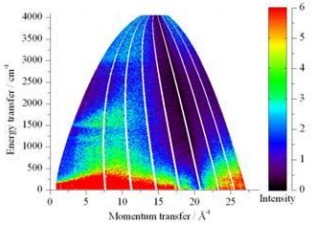
oratories

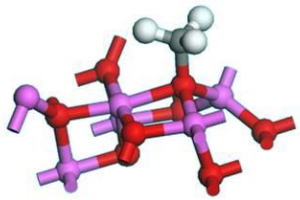






"Infrared spectroscopy only shows what is on the surface of the catalyst, but neutrons give the whole picture. We were able to modify the catalyst to give 50% less unwanted derivative."
–Dr Stewart Parker, ISIS and Professor David Lennon, University of Glasgow





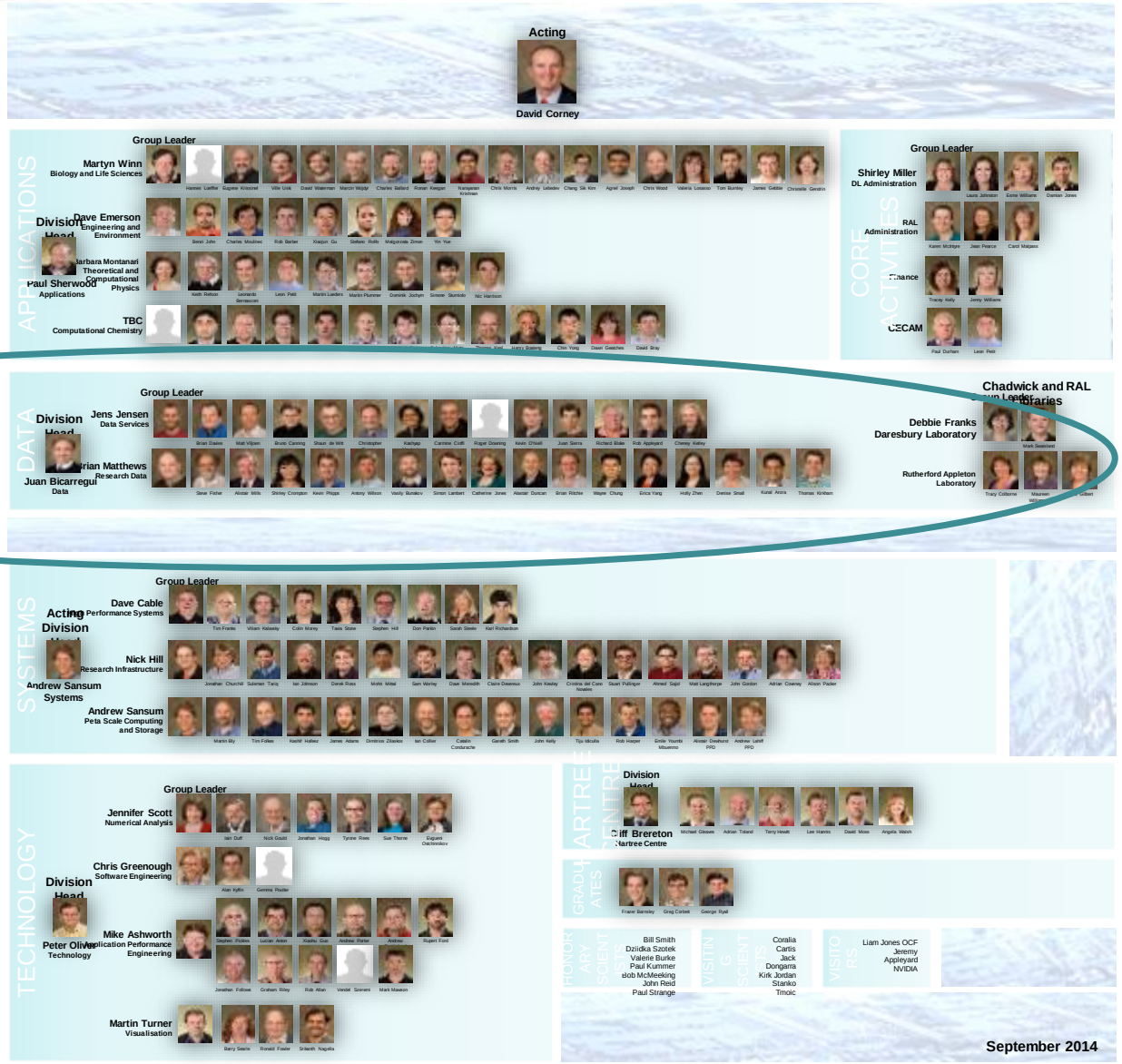
Methyl chloride synthesis: Neutrons help industry

Scientific Computing Department

Science & Technology Facilities Council

4 Divisions:

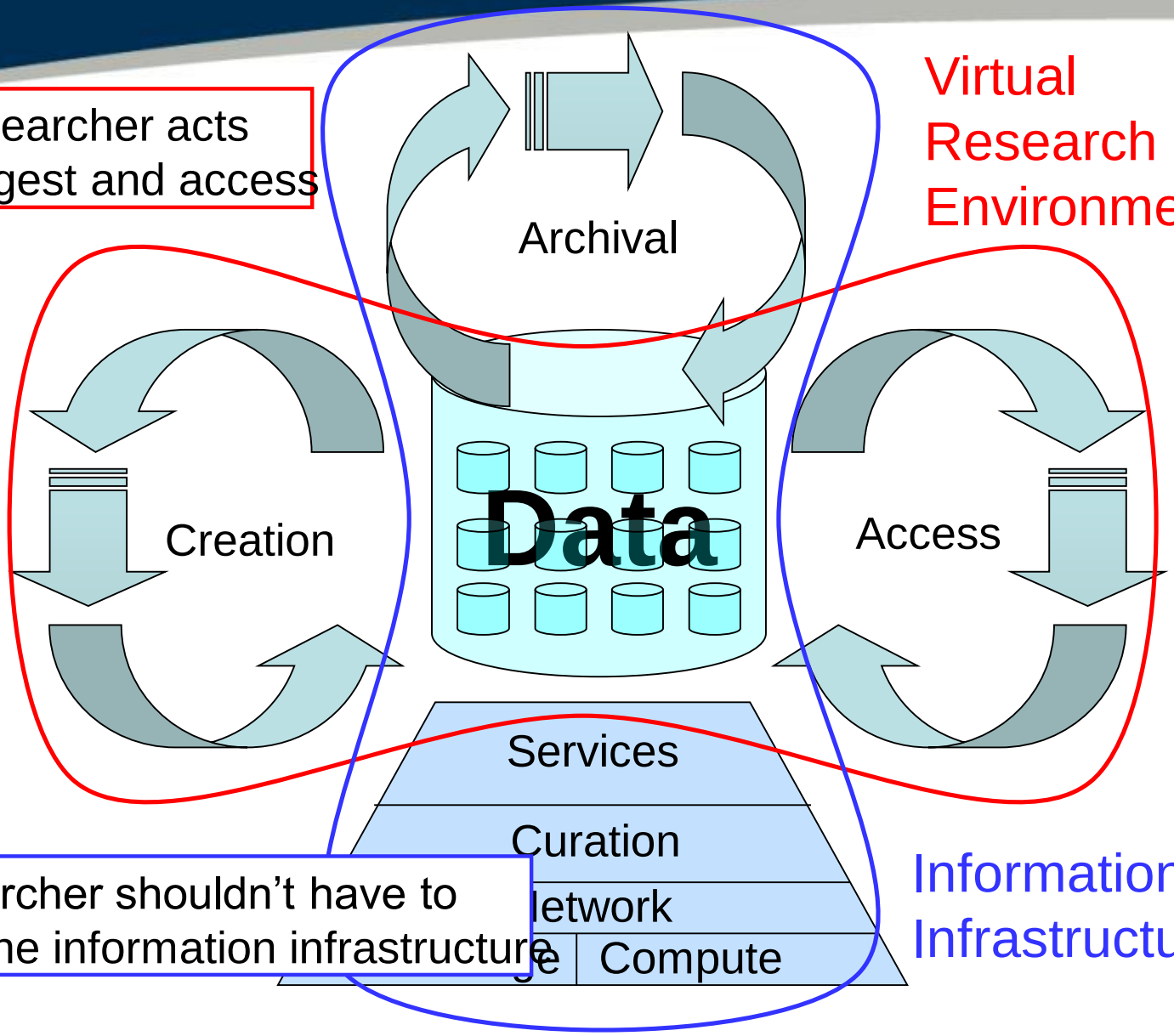
- **Applications**
- **Data**
- **Systems**
- **Technology**



Data centric view of research

Science & Technology Facilities Council

the researcher acts through ingest and access



Virtual Research Environment

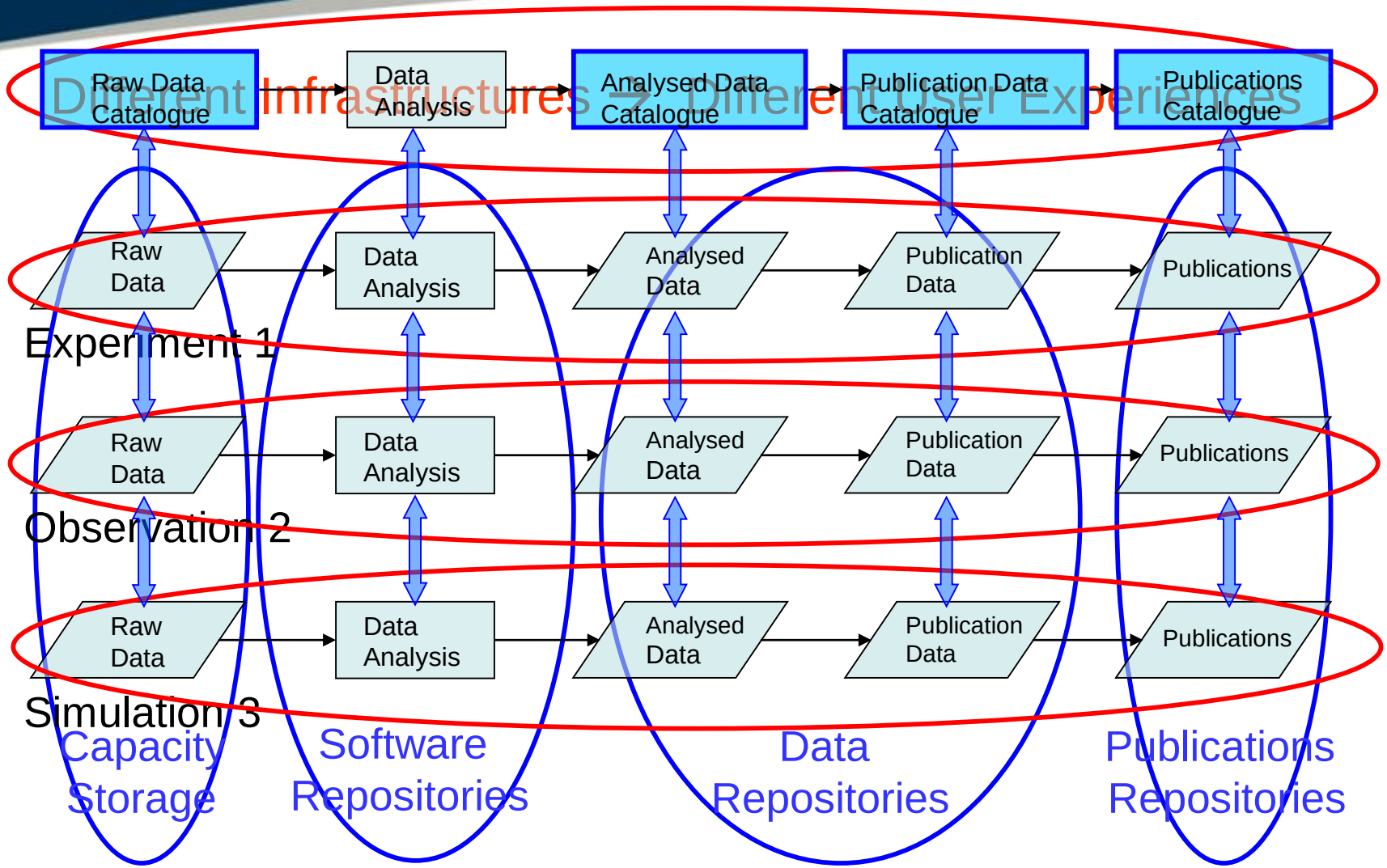
the researcher shouldn't have to worry about the information infrastructure

Information Infrastructure

Data Sharing Vision

Science & Technology Facilities Council

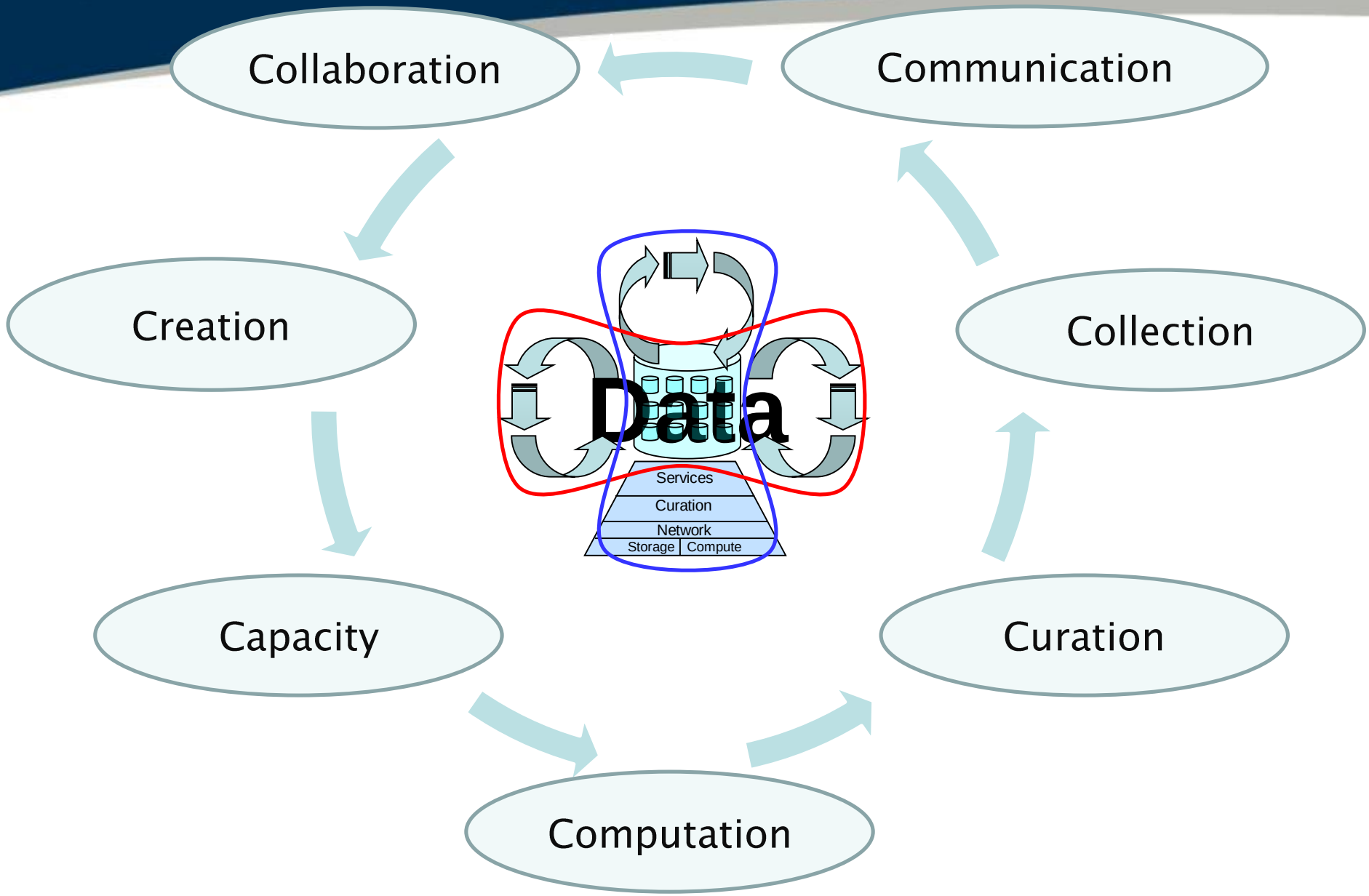
Single Infrastructure → Single User Experience



PaN-Data Infrastructure for Photon and Neutron Sources (2007)

Science & Technology Facilities Council

The 7 C's



Science & Technology Facilities Council

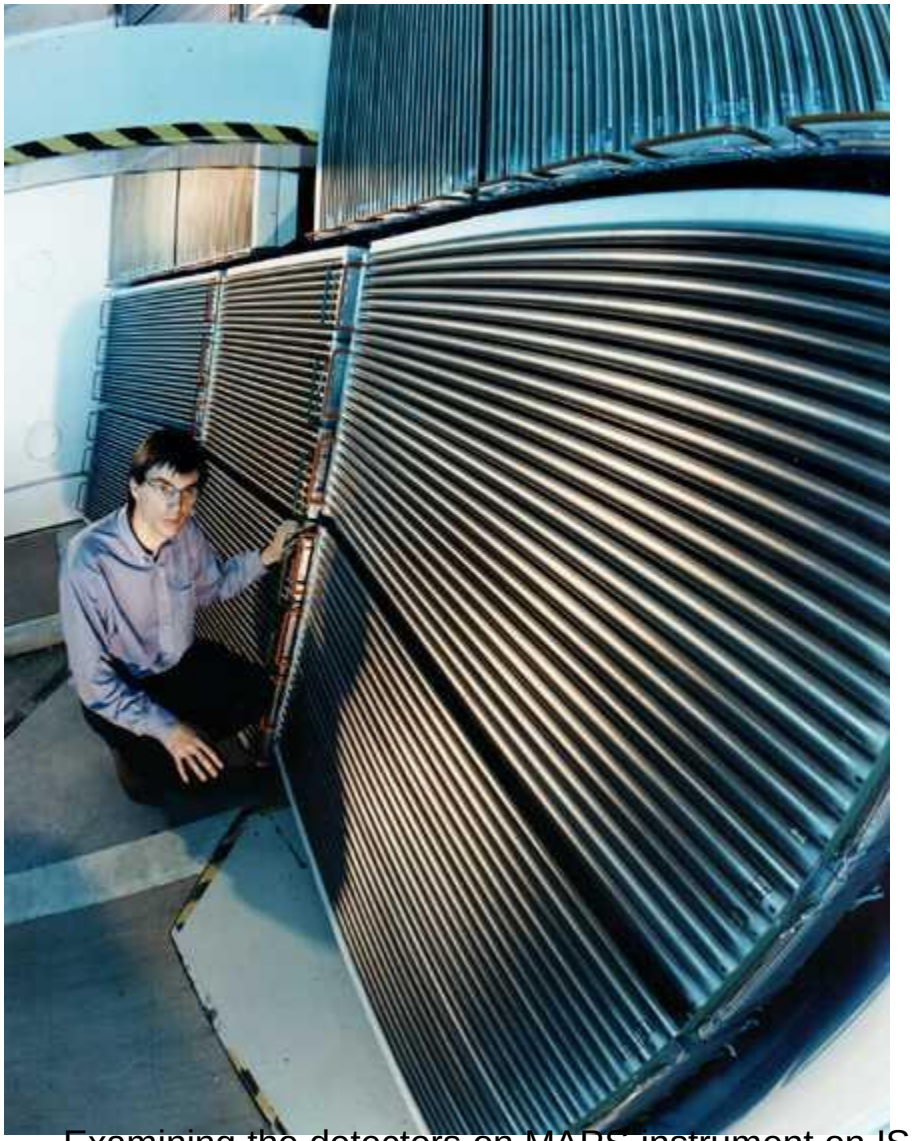
Creation

Linked systems for:

- Proposal submission
- User management
- Data acquisition

Metadata carried from each system to the next

Detectors moving from Hz to KHz



Examining the detectors on MAPS instrument on ISIS

Science & Technology Facilities Council

Capacity

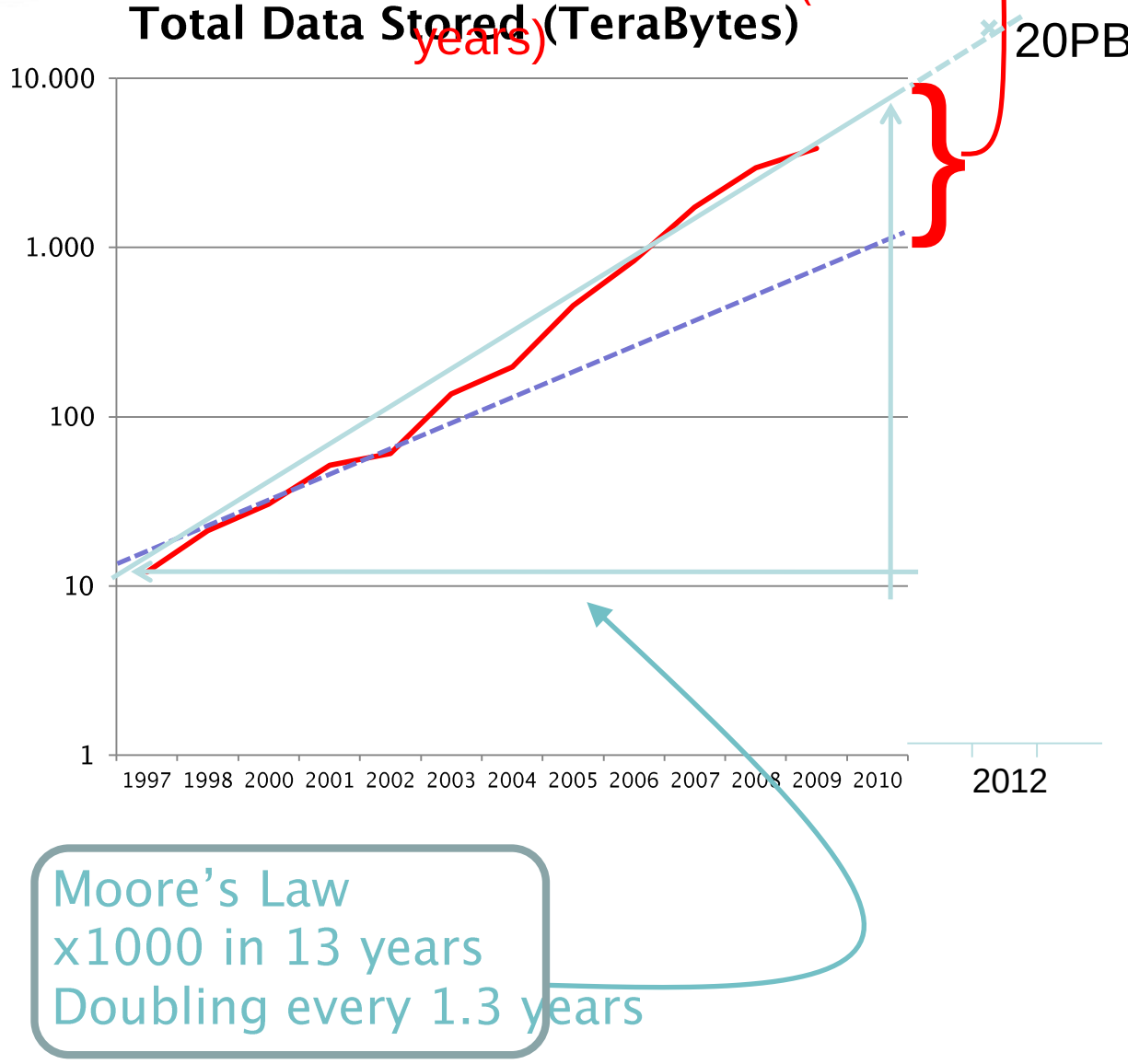
1997-2010

Moore's law for us

was about 15 months

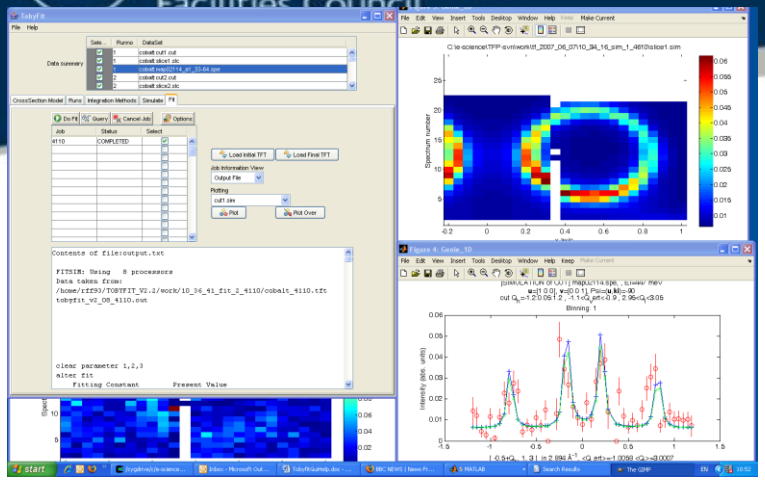
2012: ~20 PetaBytes

2015: ~50PetaBytes



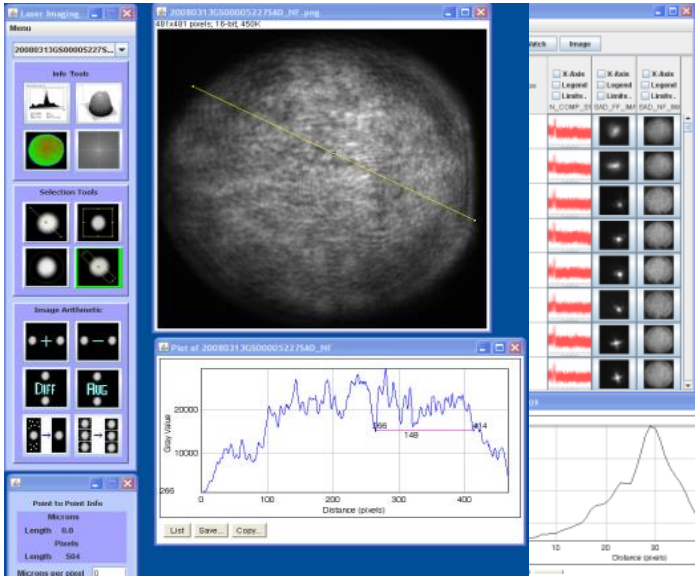
Computation

Science & Technology Facilities Council

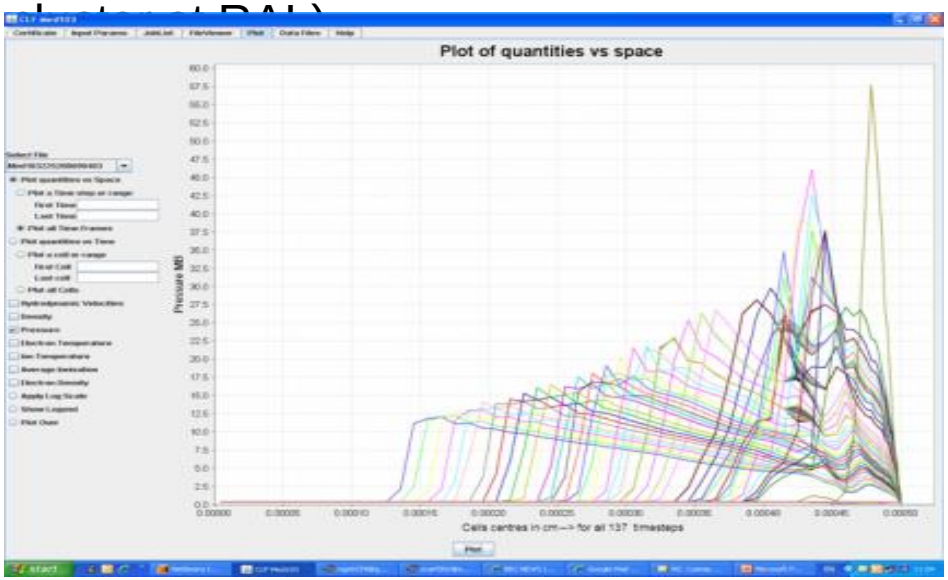


Fitting of experimental data to model

Compute intensive components on HPC (BlueGene/Q,1.2 Pflop,#13; Emerald GPGPU



Real-time diagnostics of instrument performance and



Computational derivation of properties from

Curation

Science & Technology Facilities Council

RAL Facility Archives

- All **ISIS** data (~30 years) > 3,000,000 files
- All **Diamond** Data (~10 years) > 1,000,000,000 files

LHC Tier 1

- UK hub for **LHC** data (30PB)

NERC

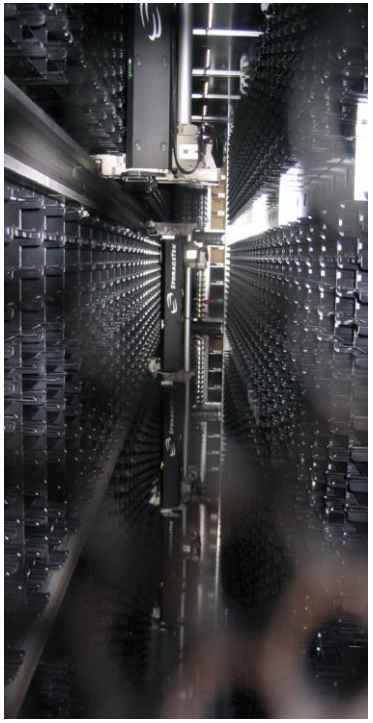
- JASMIN+CEMS super data cluster

Universities

- **Imperial College** – National Service for Computational Chemistry Software
- **Oxford**, **UCL**, **Southampton** **Bristol**, Emerald GPGPU cluster

Publications:

- The STFC Publications Archive



The StorageTek tape robots 160PB Capacity

JASMIN

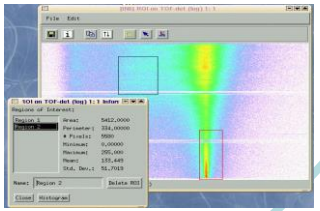
Science & Technology Facilities Council

Collection

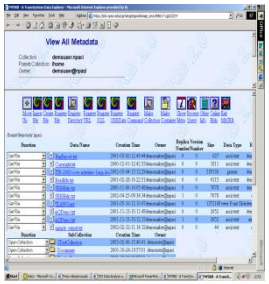
ICAT

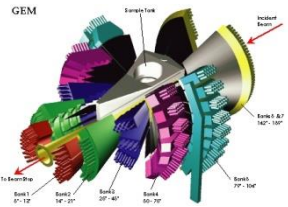
Proposal

Record Publication



Approval





Subsequent publication registered with facility

Data analysis

Scientist submits application for beamtime

Scheduling

Experiment

Data cleansing

Tools for processing made available

Facility committee approves application

Facility registers, trains, and schedules scientist's visit

Scientists visits, facility run's experiment

Raw data filtered and cleansed

Science & Technology Facilities Council

Communication

Immense Expectations !

- Web technology enables
  - Anything and Everything on-demand
- Openness enables
  - Validation of results
  - Repetition of experiment
- Discoverability enables:
  - Repurposing of knowledge
    - new from old

STFC's "e-pubs" Institutional Repository has records of 40,000 publications spanning 30 years

Science & Technology Facilities Council

Outline

STFC

- Science and Technology Facilities Council
- Scientific Computing Department

Open Science Policy

- UK, EU, G7
- 2004, 2007, 2011, 2013, 2015, 2016…

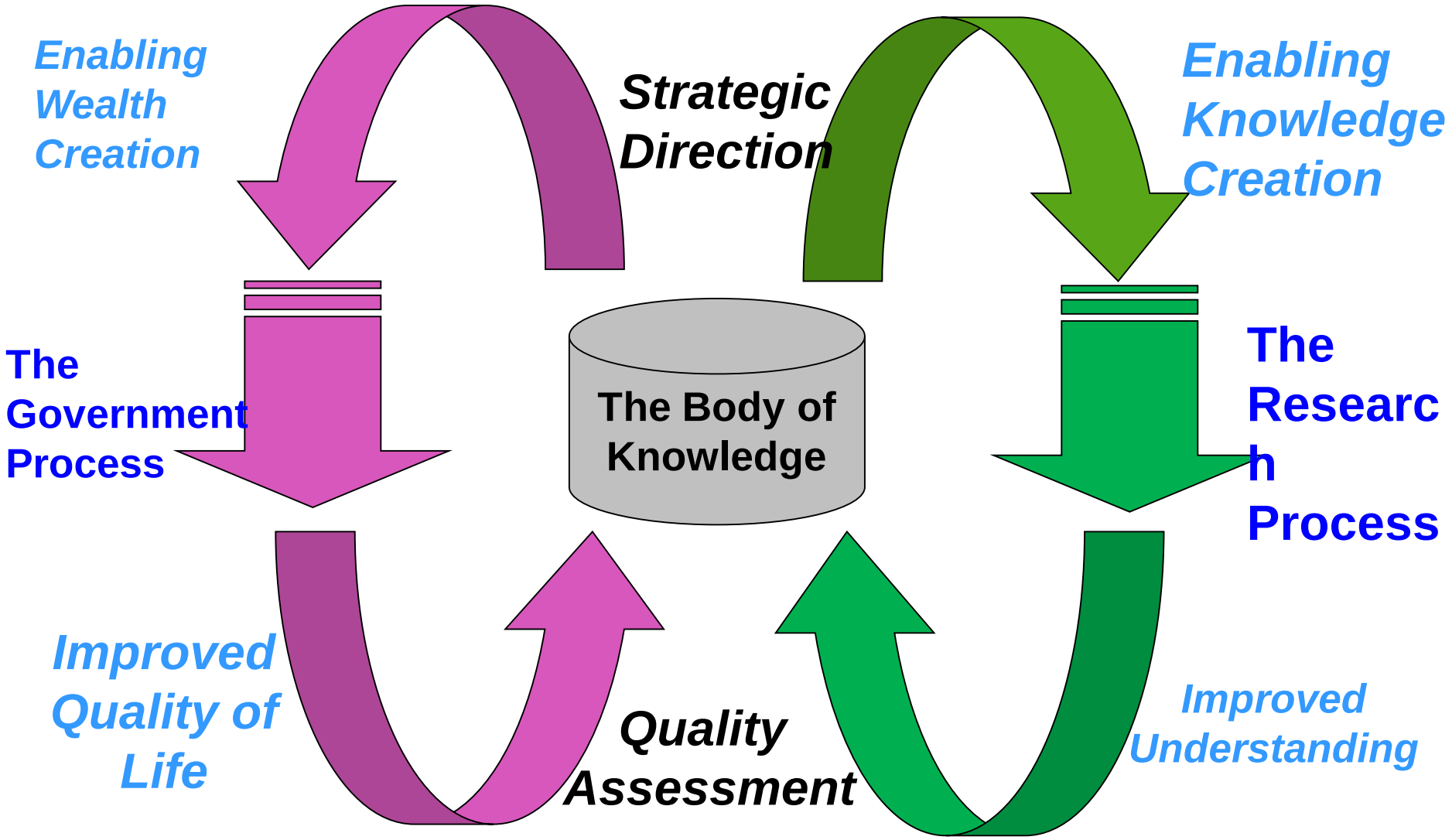
RDA

- Research Data Alliance

European Open Science Cloud

- EOSC pilot project

The Innovation Lifecycle



Aggregation of Knowledge lies at the heart of the innovation lifecycle

The Policy Context 2004–2012

- OECD, 2004–2006
  - Principles and *Guidelines* for Access to Research *Data* from *Public* Funding
- EC, 2007–2012
  - Recommendation on access to and preservation of scientific information
- G8+5, 2011–2012
  - Global Research Infrastructure Sub Group on Data

- Research Councils UK, Joint Principles on Data 2011
  - *Publicly funded research data are a public good….*

- [Royal Society, 2011–2012]
  - Science as an Open Exercise

The Policy Context 2013 –

- G8 Ministerial Communiqué, 2013
  - *"… [publically funded] scientific research data should be open…"*
- [EC Communication and OSTP Memo 2013]
- G7 Ministerial Communiqué, October 2015

  "…accomplish an effective open-data science environment at the G7 level and beyond."
- EC Communication on European Cloud Initiatives, April 2016
  - LATER

OECD 2004 – 2006
ON ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

2003 – Science and Technology Ministers called on the OECD to develop a set of guidelines based on commonly agreed principles to **facilitate cost-effective access to digital research data from public funding.**

- *Declaration* adopted on 30 January 2004

2006 – *Recommendation* of the Council concerning Access to Research Data from Public Funding

- *Principles and Guidelines endorsed by the OECD Council on 14 December 2006.* *[C(2006)184]*

OECD Principles and Guidelines for Access to Research Data from Public Funding

13 principles

A – Openness

- **Openness means access on equal terms for the international research community at the lowest possible cost, ...**

B – Flexibility, C – Transparency, D – Legal conformity, E – Protection of intellectual property, F – Formal responsibility, G – Professionalism

H – Interoperability

- **Technological and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of research data. ...**

I – Quality, J – Security, K – Efficiency, L – Accountability

M – Sustainability

- **... taking administrative responsibility for the measures to guarantee permanent access to data that have been determined to require long-term retention.**

[http://www.oecd.org/dataoecd/9/61/38500813.pdf]

EUROPEAN COMMISSION 2007-2012

- **2007**, Commission adopted a **Communication** and **Conclusions** on scientific information in the digital age
  - access, dissemination and preservation

- **2010** Riding the wave: How Europe can gain from the rising tide of scientific data
  - Final report of the High Level Expert Group on Scientific Data

- **2012** – COMMISSION **RECOMMENDATION** to member states on access to and preservation of scientific information
  - Covers Publications, Data, and Infrastructure

17 July 2012 – COMMISSION **RECOMMENDATION** on Access to and Preservation of Scientific Information

"publicly funded research should be widely disseminated through open access publication of scientific data and papers."

Regarding ***Open access to research data***, member states should:

- Define clear policies for the dissemination of and open access to research data resulting from publicly funded research.
- Ensure that, as a result of these policies:
  - research data that result from publicly funded research become **publicly accessible, usable and re-usable** through digital e-infrastructures. ...
  - datasets are made **easily identifiable** and can be linked to other datasets and publications ...

G8+5 Global Research Infrastructure Subgroup on Data report 2011

In 2020/2030…

- Researchers and practitioners from any discipline are able to **find, access and process the data they need** in a timely manner.

- They are confident in their **ability to use and understand** data, and they can evaluate the degree to which that data can be trusted. ...

- Data are managed, shared, and preserved in a way that **optimizes scientific discovery, innovation, and societal benefit**. Where appropriate, producers of data benefit from opening it to broad access and routinely deposit their data in reliable repositories. A framework of repositories work to international standards, to ensure they are trustworthy...

G8+5 are: Canada, France, Germany, Italy, Japan, Russia, United Kingdom, United States, + Brazil, China, India, Mexico, South Africa.

G8+5 2012

Framework for Global Research Infrastructures.

- Global scientific data infrastructure providers and users **should establish an international forum for data interoperability**. It should facilitate the exchange and interoperability of data across disciplines and national boundaries by producing high quality, relevant technical documents and procedures that influence the way researchers store, use, and manage data.

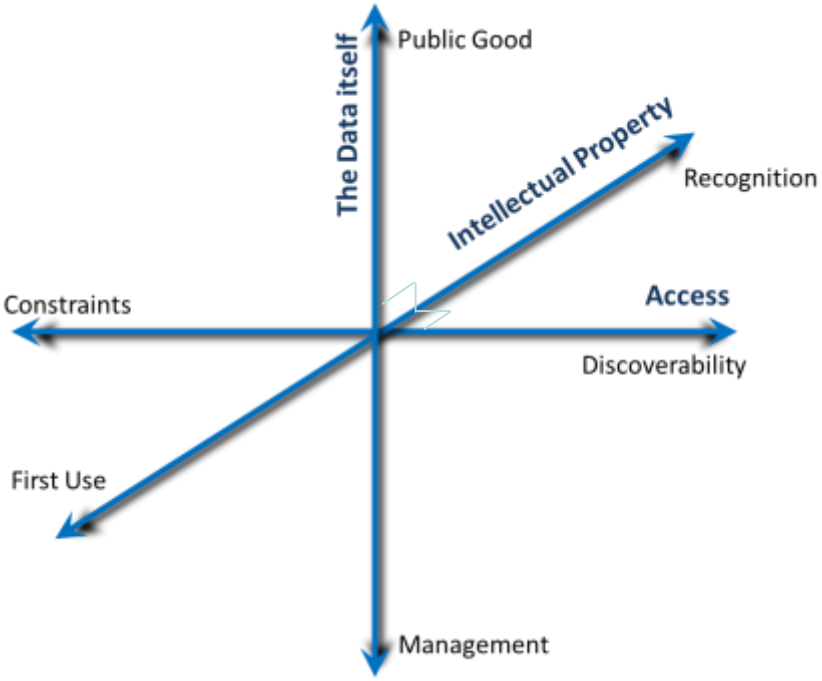
→ RDA

RCUK Principles on Data Policy (2011)

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…



G8: Open Scientific Research Data. London, 2013

- *Antimicrobial Resistance*

- *Open Access to Publications*

- *Open Data:*

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 environmen**t at the G7 level and beyond."*

4. Clean Energy

Science & Technology Facilities Council

Outline

STFC

- Science and Technology Facilities Council
- Scientific Computing Department

Open Science Policy

- UK, EU, G7
- 2004, 2007, 2011, 2013, 2015, 2016…

RDA

- Research Data Alliance

European Open Science Cloud

- EOSC pilot project

RDA

The Research Data Alliance

research data sharing without barriers

rd-alliance.org

RDA – Vision and Purpose (2013)

Vision

Researchers around the world

sharing and using research data without barriers.

Purpose

… to accelerate international

data-driven innovation and discovery

by facilitating research data

***sharing** and **exchange**,*

***use** and **re-use**,*

***standards harmonization**, and **discoverability**.*

*…through the development and adoption of **infrastructure, policy, practice, standards**, and other deliverables.*

RDA Principles

Openness

- Membership is open to all interested organizations
- Meetings are public, Processes are transparent
- Products are freely available to the public

Consensus

- RDA moves forward by achieving consensus
- resolves disagreements through appropriate voting mechanisms;

Balance

- Organized on the principle of balanced representation for individual organizations and stakeholder communities;

Harmonization

- Works to achieve harmonization across standards, policies, technologies, tools, and other data infrastructure elements;

Voluntary

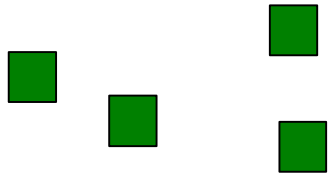
- The RDA is not a government organization or regulatory body
- is a public body responsive to its members

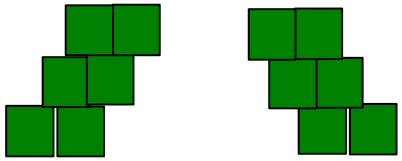
Non-profit

- Not a commercial organization and will not design, promote, endorse, or sell commercial products, technologies, or services.

“Building Bridges”

- Bridges to the future
  - data preservation
- Bridges to research partners
- Bridges across disciplines
- Bridges across regions
- Bridges to integration

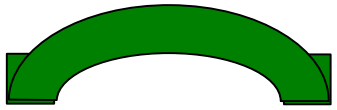




Two types of bridges we can build:

- Connecting Data
- Connecting People

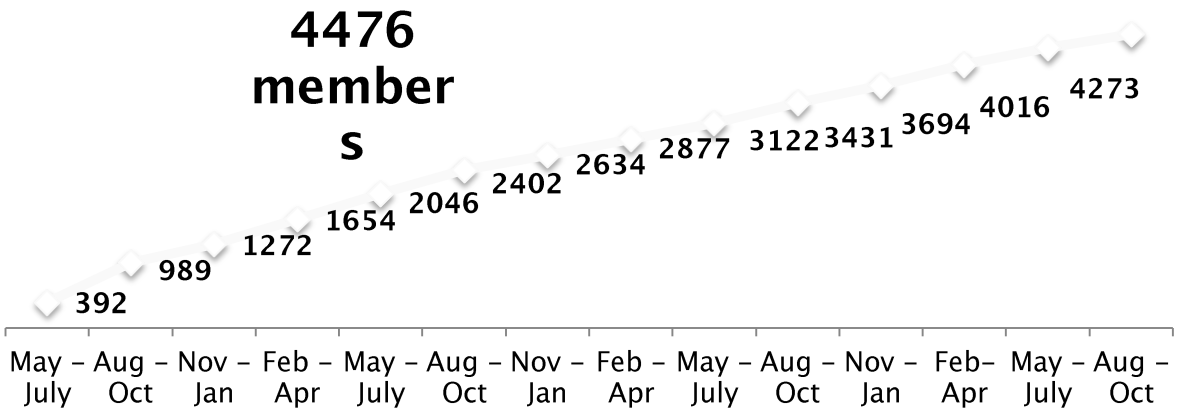
What kind of organisation do we need to do this?

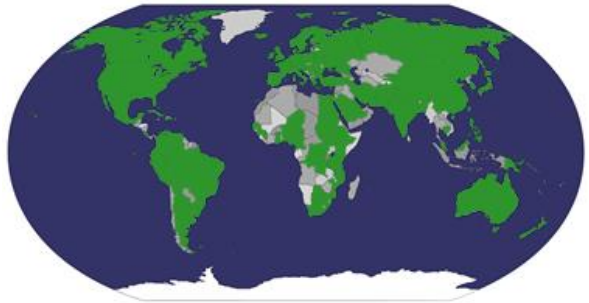


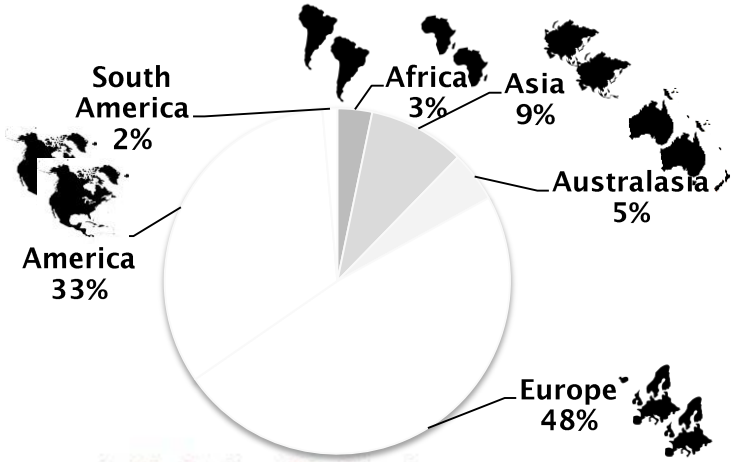
Who is RDA?

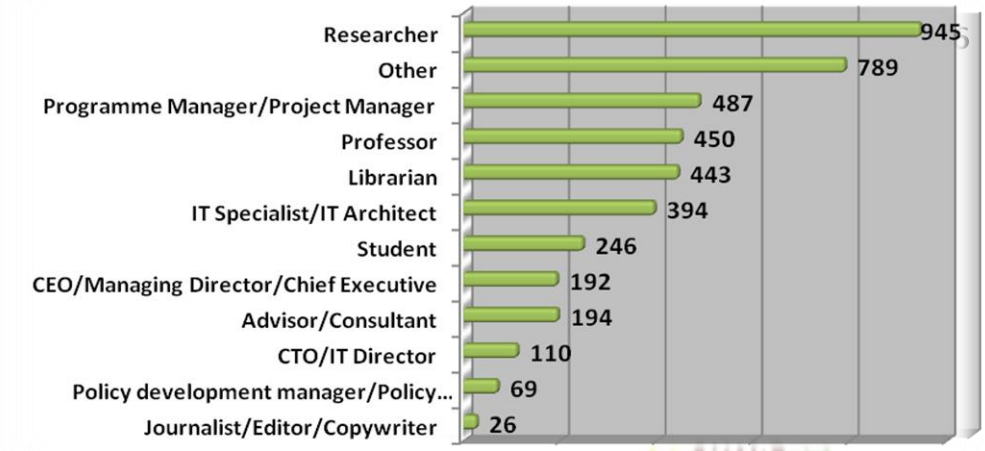
Type	Actual members (Oct. 2016)
Press and Media	27
Policy/Funding Agency	66
Large Enterprise	99
IT Consultancy/Development	146
Small and Medium Enterprise	257
Other (please specify) – obligatory	241
Government/Public Services	690
Academia/Research	2950

from 112 countries









research data sharing without barriers

Organisational & Affiliate Members

44 RDA Organisational Members

6 RDA Affiliate Members

RDA Working Groups
Create Enabling Infrastructure

Presented at Plenary 1 March 2013

RDA Working Group

Data Use, Sharing and Exchange

Community deployment of adopted / implemented infrastructure, tools, policy, practice, standards facilitating data sharing and exchange

Foundations

RDA Working Group Deliverables

RDA Interest (IG) & Working Groups (WG) by Focus (1 of 3)

Total 73 groups: 27 Working Groups & 46 Interest Groups

Domain Science - focused

Agrisemantics WG

BioSharing Registry: connecting data policies, standards & databases in life sciences WG

Rice Data Interoperability WG

Wheat Data Interoperability WG

Agriculture Data IG (IGAD)

Biodiversity Data Integration IG

Chemistry Research Data IG

Digital Practices in History and Ethnography IG

Geospatial IG

Global Water Information IG

Health Data IG

Marine Data Harmonization IG

Metabolomics Data Interoperability IG

Quality of Urban Life IG

RDA/CODATA Materials Data, Infrastructure & Interoperability

Research data needs of the Photon and Neutron Science community IG

Structural Biology IG

Community Needs - focused

RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG

Archives & Records Professionals for Research Data IG

Data for Development IG

Development of Cloud Computing Capacity and Education in Developing World Research IG

Education and Training on handling of research data IG

Engagement IG

Ethics and Social Aspects of Data IG

RDA Interest (IG) & Working Groups (WG) by Focus (2 of 3)

Referencing and Sharing - focused

Data Citation WG

Data Description Registry Interoperability WG

Data Security and Trust WG

Empirical Humanities Metadata WG

RDA / WDS Publishing Data Bibliometrics WG

Research Data Collections WG

QoS-DataLC Definitions WG

International Materials Resource Registries WG

National Data Services IG

RDA/CODATA Legal Interoperability IG

Reproducibility IG

New Paradigms for Data Discovery

Partnership Groups

RDA / TDWG Metadata Standards for attribution of physical and digital collections stewardship

RDA/NISO Privacy Implications of Research Data Sets WG

Repository Audit and Certification DSA–WDS Partnership WG

RDA/WDS Publishing Data IG

ELIXIR Bridging Force IG

RDA Interest (IG) & Working Groups (WG) by Focus (3 of 3)

Data Stewardship and Services – focused

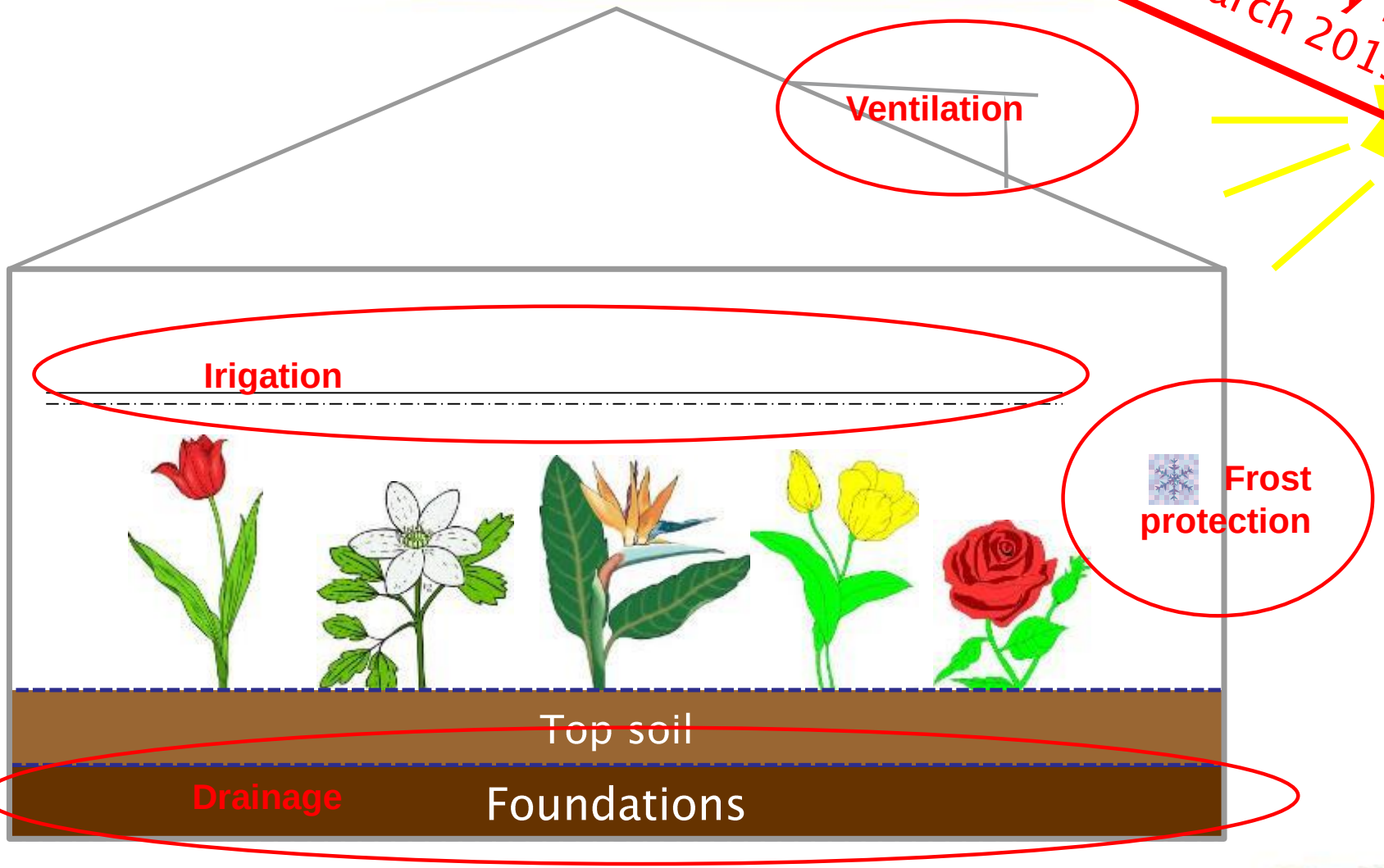
Brokering Framework WG
Brokering Governance WG
RDA / WDS Publishing Data Services WG
RDA / WDS Publishing Data Workflows WG
Active Data Management Plans IG
Data in Context IG
Data Rescue IG
Domain Repositories IG
Libraries for Research Data IG
Long tail of research data IG
Preservation e-Infrastructure IG
RDA/WDS Certification of Digital Repositories IG
RDA/WDS Publishing Data Cost Recovery for Data Centres IG
Repository Platforms for Research Data IG
Research Data Provenance IG
Virtual Research Environments IG

Base Infrastructure – focused

Array Database WG
Data Foundation and Terminology WG
Data Type Registries WG
Metadata Standards Catalog WG
Metadata Standards Directory WG
PID Information Types WG
Practical Policy WG
Data Fabric IG
Data Foundations and Terminology IG
Data in Context IG
Big Data IG
Brokering IG
Federated Identity Management IG
Metadata IG
PID IG
Service Management IG
Vocabulary Services IG

Let the flowers grow....

Presented at Plenary 1 March 2013



RDA Organizational Structure provides coordinating framework

RDA Membership

Working Groups
Responsible for impactful, outcome-oriented efforts

Interest Groups
Responsible for defining and refining common issues

Technical Advisory Board
Responsible for Technical roadmap and interactions

Secretary-General and Secretariat
Responsible for administration and operations

Organizational Advisory Board and Organizational Assembly
Responsible for organizational and strategic advice

RDA Council
Responsible for overarching mission, vision, impact of RDA

RDA Funders Forum (Informal)
Operational and community sponsorship

RDA Recommendations & Outputs

- **Data Foundation & Terminology**: a model for data in the registered domain.
- **PID Information Types**: a common protocol for providers and users of persistent ID services worldwide.
- **Data Type Registries**: allowing humans and machines to act on unknown, but registered, data types.
- **Practical Policy**: defining best practices of how to deal with data automatically and in a documented way with computer e policy.

RDA Recommendations & Outputs

- **Metadata standards directory**: Community curated standards catalogue for metadata interoperability
- **Data Citation**: defining mechanisms to reliably cite dynamic data
- **Data Description Registry Interoperability** solutions enabling cross platform discovery based on existing open protocols and standards
- **Wheat Data Interoperability** impacting the discoverability, reusability and interoperability of wheat data by building a common framework for describing, representing linking and publishing wheat data

RDA Recommendations & Outputs

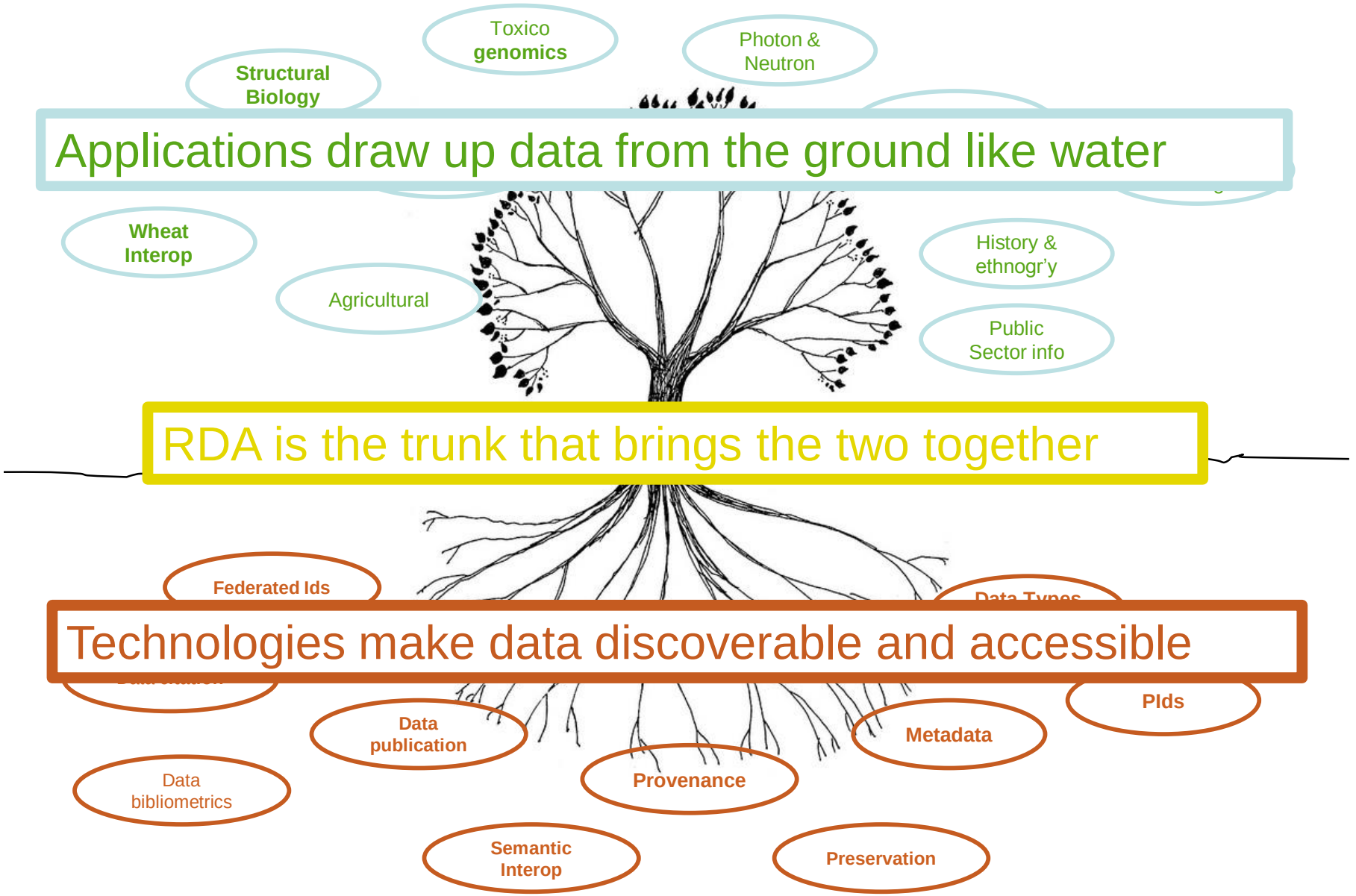
- **Repository Audit and Certification DSA-WDS:** A convergent DSA-WDS certification standard to help eliminate duplication of effort, increase certification procedure coherence and compatibility thus benefitting researchers, data managers, librarians and scientific communities.
- **RDA/WDS Publishing Data Bibliometrics:** improved research data metrics and corresponding services, with the final goal of increasing the overall availability and quality of citations and research data itself.
- **RDA/WDS Publishing Data Services:** A universal interlinking service between data and the scientific literature.
- **RDA/WDS Publishing Data Workflows:** enhance the possibilities for greater discoverability and a more efficient and reliable reuse of research data benefitting other stakeholders like publishers, libraries and data centres.

RDA Recommendations & Outputs

- **Brokering Governance WG:** Sustainable Business Models for Brokering Middleware to support Research Interoperability

- **RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG:** A framework to run a series of Summer Schools in Data Science and data sharing in low and middle income countries (LMICs

RDA – Data is the New ~~Oil~~ Water



RDA Plenaries

Plenary 8, Denver, September 2016

69 Breakout session meetings:

- 31 Interest group meetings
- 10 Working group meetings
- 12 BoF Meetings
- 16 Joint group meetings

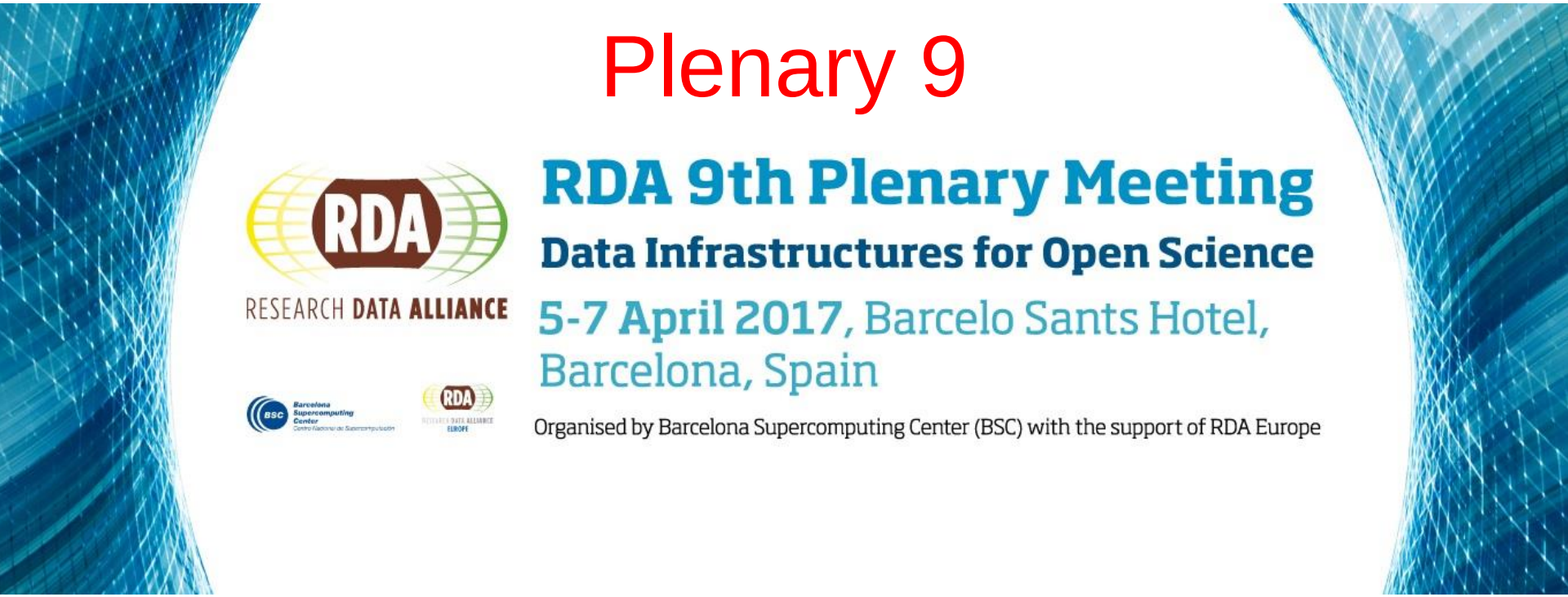
549 participants
from 33 countries

Also

- Newcomers session
- Organisational Members meetings
- Technical Advisory Board
- Group Chairs session

RDA Plenaries

Plenary 9



The 9th RDA Plenary meeting is organised by
Barcelona Supercomputing Center
Centro Nacional de Supercomputación (BSC-CNS)
with support from RDA Europe.

https://rd-alliance.org/plenary-meetings/rda-ninth-plenary-meeting.html

Looking forward to seeing you ALL there

Science & Technology Facilities Council

Outline

STFC

- Science and Technology Facilities Council
- Scientific Computing Department

Open Science Policy

- UK, EU, G7
- 2004, 2007, 2011, 2013, 2015, 2016…

RDA

- Research Data Alliance

European Open Science Cloud

- EOSC pilot project

An "Open-data Science Environment"

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,…

Relates to:
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

EC Communication on European Science Cloud Initiatives 19th April 2016

- European Open Science Cloud (EOSC)
- European Data Infrastructure (EDI)
- [Exascale and Quantum Computing]

- High Level Expert Group on EOSC
- EOSC pilot project

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 **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 se**rvices for Open science **supported by** the necessary **data infrastructure**
- **Enlarge the scientific user base** to researchers and innovators from all disciplines.

Evolution of infrastructure

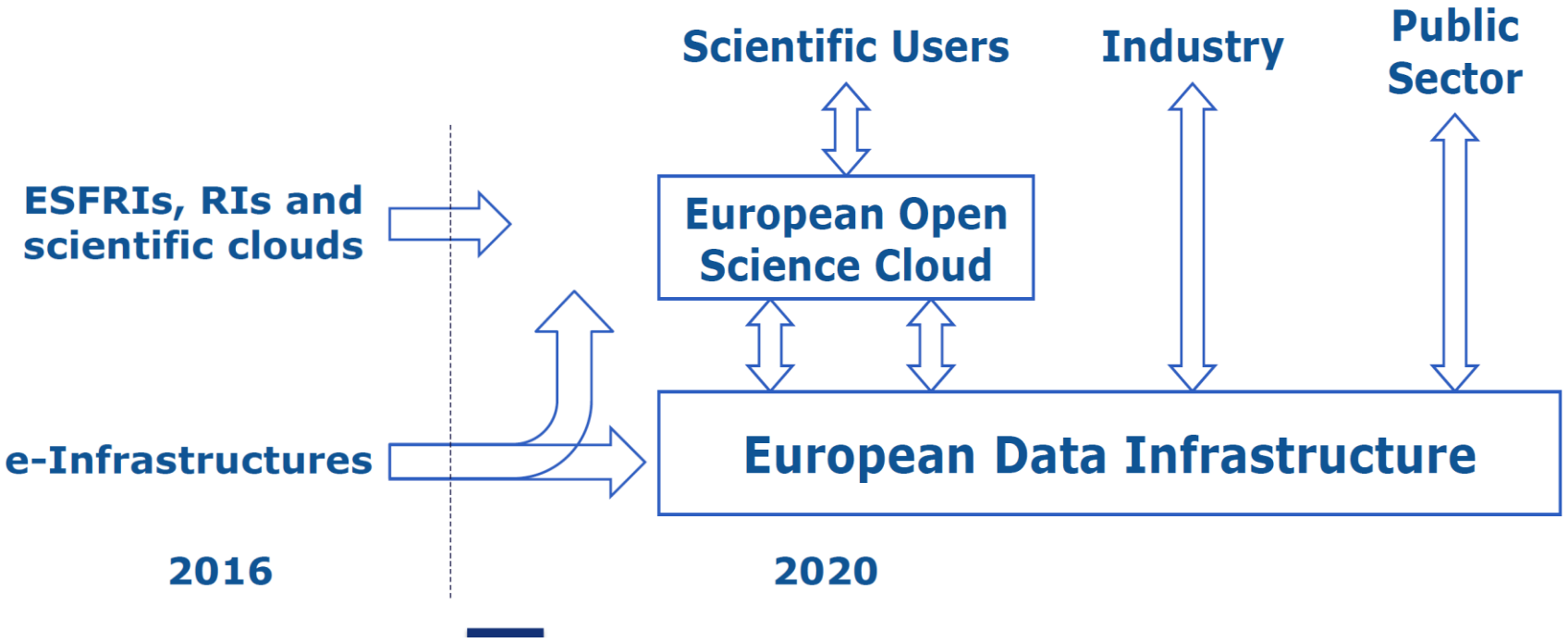
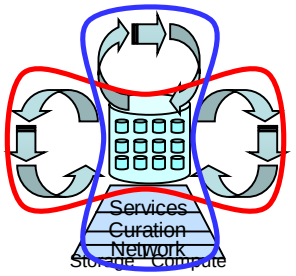


Diagram: Augusto BURGUEÑO ARJONA



Published: 11 October 2016

High Level Expert Group for the "European Open Science Cloud"

Final draft

Monday 20 June 2016

A Cloud on the 2020 Horizon

Commission High Level Expert Group on the European Open Science Cloud

Realising the European Open Science Cloud: first report and recommendations

Preface by Barend Mons, Chair

This report aims to lay out a high level, living roadmap for the realisation of the European Open Science Cloud (EOSC). The High Level Expert Group, with ten members from European countries, Japan and Australia, discussed extensively in several meetings, conferences, policy events and met with key stakeholders (30 November 2015) and research funders (15 March 2016). Based on these consultations, on many 'white papers' and on a range of presentations and feed-back at international meetings, we are confident that our recommendations count on a high-level of consensus amongst all stakeholders. This was a solid basis to embark on this challenging journey with the Commission, the Member States and

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

HLEG General Observations

- The majority of the **challenges are social** rather than technical.
- The major technical challenge is **complexity of data** rather than size
- There is an alarming **shortage of data experts**
- The **article culture** is preventing effective data publishing and re-use.
- **A chasm** between e-infrastructure providers and domains specialists.
- **Fragmentation across domains** produces repetitive & isolated sol'ns.
- Short and dispersed **funding cycles** not fit making effective use of global scientific data.
- Ever larger **distributed data sets** …centralised HPC is insufficient
- **Components** needed to create EOSC are largely there but they **are lost in fragmentation**
- There is no **dedicated and mandated effort** or instrument to coordinate EOSC-type activities across Member States.

Policy recommendations

- P1: Take immediate, affirmative action on the EOSC in close concert with Member States.
- P2: Close discussions about the 'perceived need'.
- P3: Build on existing capacity and expertise where possible.
- P4: Frame the EOSC as the EU contribution to an Internet of FAIR Data and Services underpinned with open protocols.

Governance recommendations

- G1: Aim at the **lightest possible**, internationally effective **governance**.
- G2: **Guidance only where guidance is due**
  - (this relates to technical issues, best practices and social change).
- G3: Define **Rules of Engagement** for service provision in the EOSC.
- G4: **Federate the gems** and amplify good practice.

Implementation recommendations

- I1: Turn the **HLEG report** into a high-level guide to scope and guide the EOSC initiative.
- I2: Develop, endorse and implement the **Rules of Engagement** for the EOSC.
- I2.1: Set initial guiding principles to **kick-start** the initiative as quickly as possible.
- I3: Fund a concerted effort to develop **core data expertise** in Europe.
- I4: Develop a concrete plan for the **architecture of data interoperability** of the EOSC.
- I5: Install an innovative guided funding scheme for the **preparatory phase**.
- I6: Make adequate **data stewardship mandatory** for all research proposals.
- I7: Provide a clear **operational timeline** to deal with the early preparatory phase of the EOSC.
-

European Open Science Cloud for Research Pilot Project

EOSCpilot

- Response to INFRADEV-4 call
  - 33 Partners (Leaves and Roots)
  - 2 years
  - €10M
  - Starts 1 January 2017

- A pilot
  - Running some demonstrators
    - requirements and benefits
    - Application pull and Technology push
  - Governance for the EOSC
  - Policy
  - Skills, Training etc

EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSCpilot Partners

- **Domain specific research infrastructures providers, projects and clusters**
  - STFC, EMBL, MPG, INFN, INGV, DESY, DANS, ICOS, INAF, BBMRI, ESS, BGS, XFEL, and ERCIN.
- **Horizontal e-Infrastructure providers**
  - CSC, SURF, CNRS, JISC, PRACE, BSC, GEANT, CEA, CINECA, EGI and LIBER.
  - **(Projects** EUDat, Indigo-DataCloud, EGI-Engage, AARC, and OPENAire+)
- **Research performing and support organisations**
  - EGI, UEDIN-DCC, LIBER, TRUST-IT, ARC, CNR, DANS, KIT, UEDIN, UGOE, UMAN, PIN and BGS
- **Research Funding Organisations**
  - STFC, SURF, CNRS and CNR

EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSCpilot Objectives

- **Establish the governance framework** for the EOSC and contribute to the development of European open science policy and best practice;

- **Develop a number of demonstrators** functioning as high-profile pilots that integrate services and infrastructures to show interoperability and its benefits in a number of scientific domains;

- **Engage with a broad range of stakeholders**, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research.

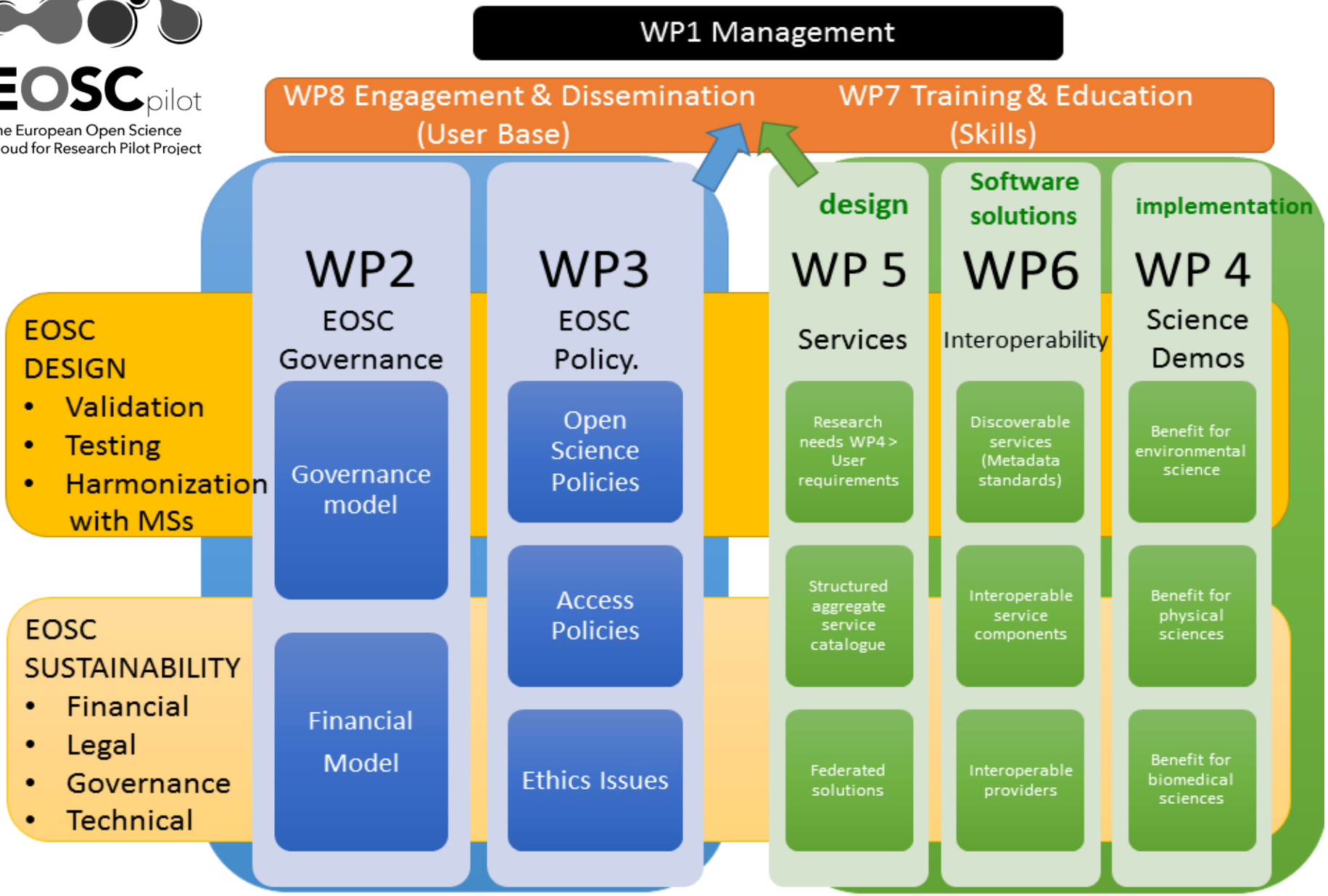
EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSCpilot Workpackages

1. Management
2. Governance
3. Policy
4. Science Demonstrators
5. Services
6. Interoperability
7. Skills
8. Engagement

Close match to objectives of EC Comm on ECI

EOSC pilot
The European Open Science Cloud for Research Pilot Project



EOSCpilot
The European Open Science Cloud for Research Pilot Project

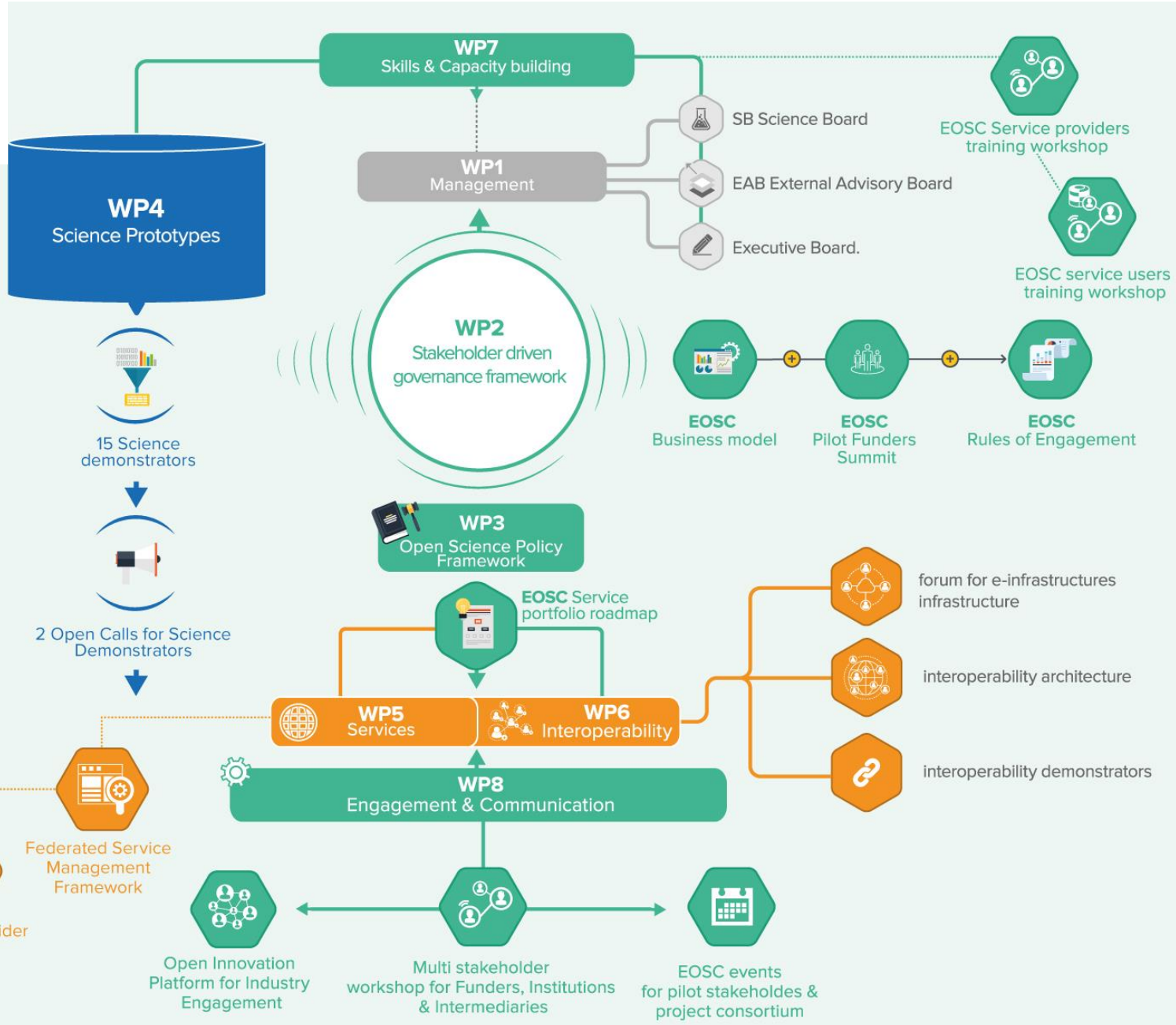
EOSCpilot
Science Demonstrators

- Requirements to EOSC
- Evaluate whether Services from EOSC meet science needs

- 28 EoIs
- 15 science demonstrators planned
- 5 selected for immediate start
  - **PanCancer Analysis of Whole Genomes:** Led by EMBL,.
  - **Research with Photons and Neutrons:** Led by DESY
  - **ENVRI Radiative Forcing Integration:** Led by ICOS ERIC,.
  - **TextCrowd:** Led by the University of Florence,
  - **WLCG:** Led by CERN,
- Two more sets of 5 will be selected at M6 and M12

EOSCpilot Outputs

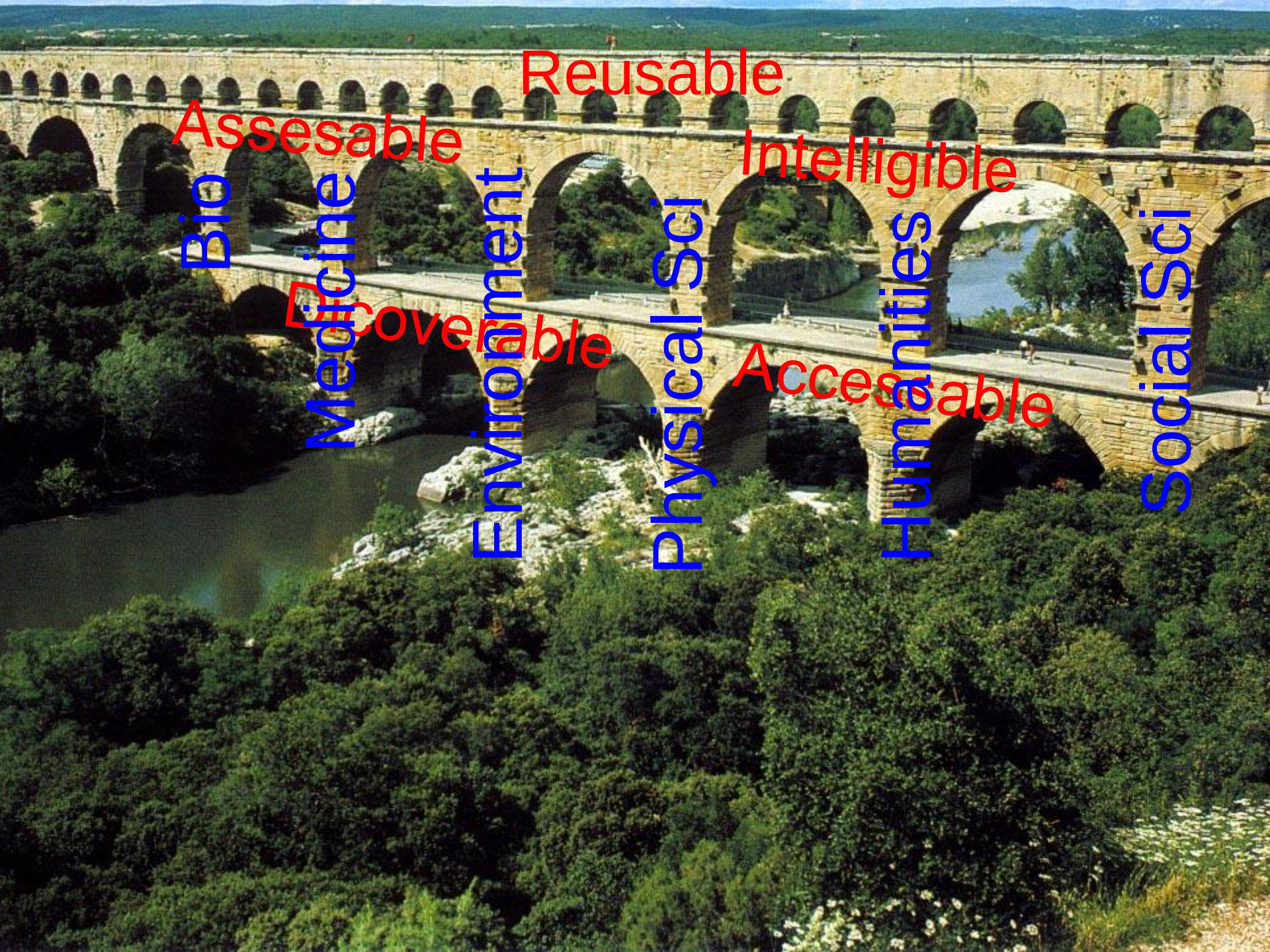
EOSCpilot
The European Open Science Cloud for Research Pilot Project



EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSCpilot is only the first step

- Further calls to come aimed at development of the EOSC:
- *EINFRA-12-2017 – Data and distributed computing e-infrastructures for Open Science (€40 M€, closes 29/3/2017)*
  - **Integration and consolidation of existing e-infrastructure services**
- *EINFRA-21-2017 – Platform-driven e-infrastructure innovation (€20 M€, closes 29/3/2017)*
  - **Service development**
- **Expect more in WP2018-20**


Reusable
Assesable
Intelligible
Bio
Medicine
Discoverable
Environment
Physical Sci
Accessable
Humanities
Social Sci

Science & Technology Facilities Council

Outline

STFC

- Science and Technology Facilities Council
- Scientific Computing Department

Open Science Policy

- UK, EU, G7
- 2004, 2007, 2011, 2013, 2015, 2016…

RDA

- Research Data Alliance

European Open Science Cloud

- EOSC pilot project

"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

Science & Technology Facilities Council

The European Open Science Cloud for Research

Juan Bicarregui
Head of Data Division, STFC
Co-Chair of RDA Organisaional Asssembly
Coordinator of EOSCpilot project

Cloud Forward
18 October 2016