



Facilitate Open Science Training for European Research

SEMINAR:
Preparing research data for open access

December 10th 2014, Social Science Data Archives, Faculty of Social Sciences,
University of Ljubljana



INTRODUCTION TO RDM FROM THE INTERNATIONAL PERSPECTIVE

Angus Whyte, Digital Curation Centre



Overview

1. What is Digital Curation Centre?
 - Quick introduction
2. Why open science?
3. Where is the infrastructure?
4. What can you do?

Overview

1. What is Digital Curation Centre?
2. What is Research Data Management?
 - What is the problem, how do we define its scope?
 - Where is policy coming from?
 - Must all the data be open?
3. Where is the infrastructure?
4. What can you do?

Overview

1. What is Digital Curation Centre?
2. Why open science?
3. Where is the infrastructure?
 - International level
 - National level e.g. UK
 - Institutional level in the UK
4. What can you do?



Overview

1. What is Digital Curation Centre?
2. Why open science?
3. Where is the infrastructure?
4. What can you do?
 - Plan data management throughout the research lifecycle
 - Deal with personal data properly
 - **Select what to keep & where to deposit**



Become a Sponsor | Call for

Latest news

Next events



IDCC15 - Registration now open

28 October, 2014 | in DCC News



Understanding and comparing digital curation costs to support smarter investments

7 November, 2014 | in DCC News



New DCC Checklist on Selecting Data to Keep

31 October, 2014 | in DCC News

Established 2004
 UK wide exchange good practice
 Share good practice
 Original focus on digital preservation



10th International Digital Curation Conference
London, 9 - 12 February 2015

Become a Sponsor | Call for

Latest news

Next events



IDCC15 - Registration now open
28 October, 2014 | in DCC News

**and comparing digital curation
smarter investments**
DCC News

st on Selecting Data to Keep
News

Since 2009 increasing focus on Research Data Management

“Helping to build capacity, capability and skills in data management and curation across the UK’s higher education research community”

How to Develop Research Data Management Services - a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte

Please cite as: Jones, S., Pryor, G. & Whyte, A. (2013). 'How to Develop Research Data Management Services - a guide for HEIs'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>



Digital Curation Centre, March 2013
 This work is licensed under Creative Commons
 Attribution BY 2.5 Scotland

Home Digital curation About us News Events Resources Training

Home > Tailored support > Guidance and support

In this section

Institutional Engagements

Advocacy

Business case

DMP

Data discovery & reuse

Guidance

Policy & strategy

Requirements gathering

Training

Guidance and support

Guidance webpages

Researchers are often unaware of the support available within their institution, so raising awareness is useful. Many institutions have [RDM guidance webpages](#). We can help you to create some of your own.

Guidance documents

The DCC provides a number of guidance documents. See in particular our short, practical [How to guides](#). We can provide you with hard copies and licence [our content as CC-BY](#) to encourage reuse.

We can also develop guidance documents to meet your particular needs. For example, University of Edinburgh asked us to produce a [FAQ on Data Management](#). Queen Mary, University of London asked for a [Quick Guide to RDM](#).

RDM support

We can help you to develop support more generally, such as through our [RDM support](#) touch with other institutions undertaking similar work to help you.

Supported by Jisc

Shared service provider to UK higher education

Catalogue of services

- Digital content
- Network and IT services
- Advice
 - legal aspects of ICT
 - disability and accessibility
 - research data curation and digital preservation
 - innovative use of digital media
- Research & Development

So what is the problem?

1. Researchers do what is required to manage data to pursue the immediate need
2. Ad-hoc solutions, unsupported, un-rewarded for managing research data
3. Digital research data disappears from the research record unless actively managed
4. Research cannot be scrutinised or reproduced
5. Funders' investment is lost along with the data

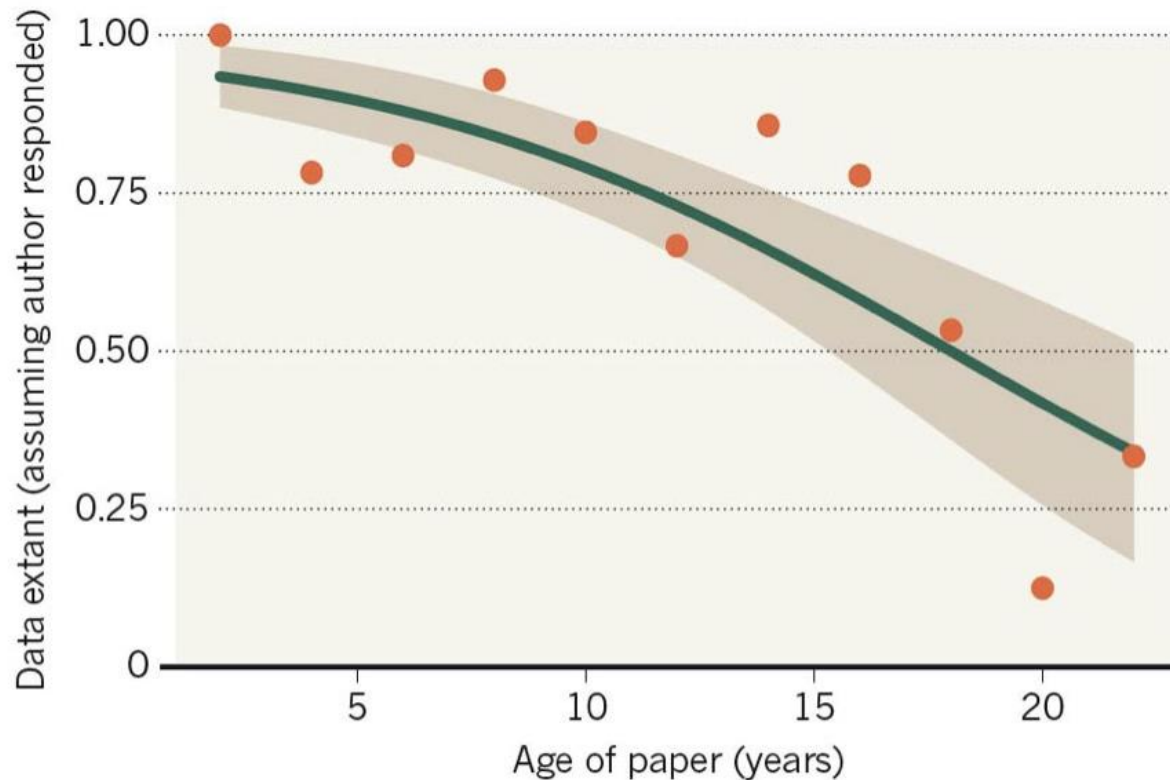
So what is the problem?

1. Researchers do what is required to manage data to pursue the immediate need
2. Ad-hoc solutions, unsupported, un-rewarded for managing research data
3. **Digital research data disappears from the research record unless actively managed**
4. Difficulty scrutinising or reproducing research
5. Funders' investment is lost along with the data!

Disappearing research record

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Concern about reproducibility

from within the research community ...



Statistics' Crisis of Reproducibility

By **Organizing Committee of the Future of the Statistical Sciences Workshop** | Published:
July 21, 2014

- Studies cannot be reproduced
- Data analysis poorly carried out

<http://www.socialsciencespace.com/2014/07/statistics-crisis-of-reproducibility/>

Concern about scrutiny

...and from the public, to manage and share data better to make fraud easier to detect

Retraction Watch

New Dutch psychology scandal? Inquiry cites data manipulation, calls for retraction

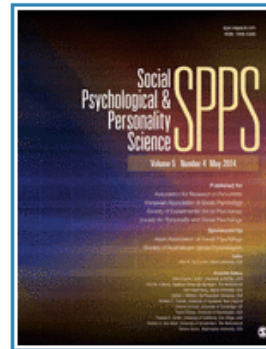
with 28 comments

The University of Amsterdam has called for the retraction of a 2011 paper by two psychology researchers after a school investigation concluded that the article contained bogus data, the Dutch press are [reporting](#).

The paper, "Sense Creative! The Impact of Global and Local Vision, Hearing, Touching, Tasting and Smelling on Creative and Analytic Thought," was written by [Jens Förster](#) and [Markus Denzler](#) and published in *Social Psychological & Personality Science*. It [purported to find that](#):

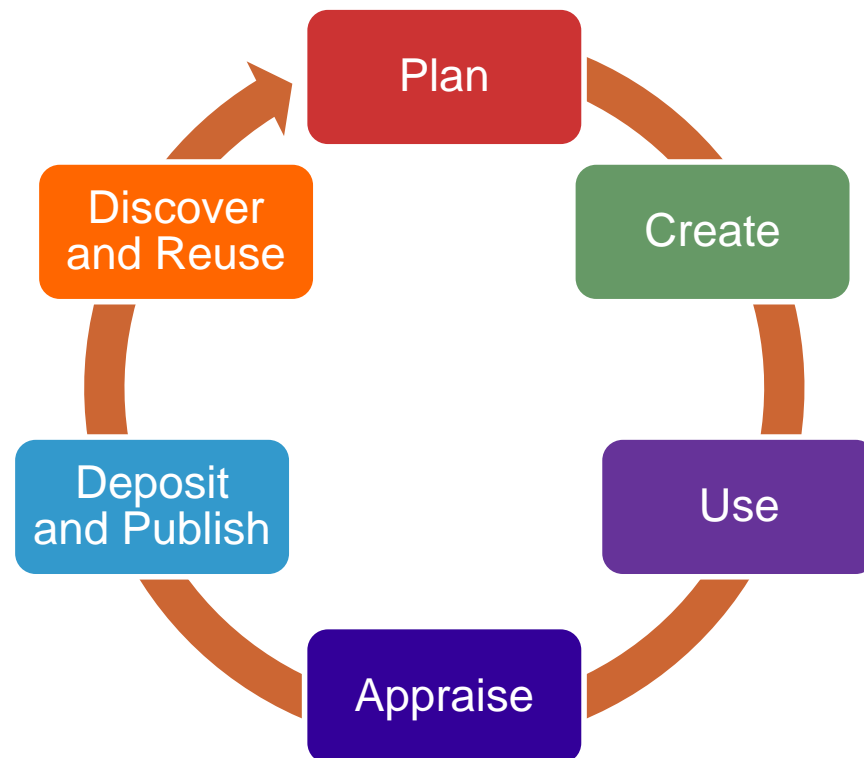


Holistic (global) versus elemental (local) perception reflects a prominent distinction in psychology; however, so far it has almost entirely been examined in the domain of vision. Current work suggests that global/local



Define ‘Research Data Management’?

“An explicit process, covering the creation and stewardship of research materials to enable their use for as long as they retain value” (DCC)



So what research data is ‘published’ ?

Any combination of..

1. **Source data** collected, created, or held elsewhere that the research has used
2. **Assembled datasets** extracted or derived from (1)
3. **Referenced data** ‘supplementary material’ from which conclusions drawn (**=most common***)

* see Wiley Researcher Data Insights Survey 2014 Available at:
<https://scholarlykitchen.files.wordpress.com/2014/11/researcher-data-insights-infographic-final.pdf>

Adapted from: Peter Burnhill, Muriel Mewissen & Adam Rusbridge (2014) ‘ Where data and journal content collide: what does it mean to ‘publish your data’ Presented at ‘Dealing with Data Conference’, 26 August 2014, University of Edinburgh Library.

Available at: <https://www.era.lib.ed.ac.uk/handle/1842/9394>

Must all ‘managed’ data be ‘open’?

No. RDM also for data that needs kept but not shared.

One definition of **open science**

“science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process”

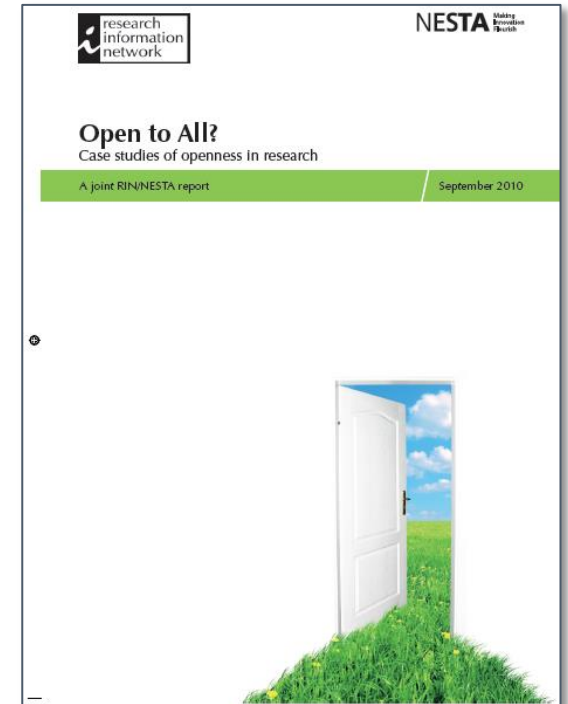
RDM/ Open Science overlap = data sharing

Open to All? Case studies of openness in research. Retrieved from <http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies>

Case study examples

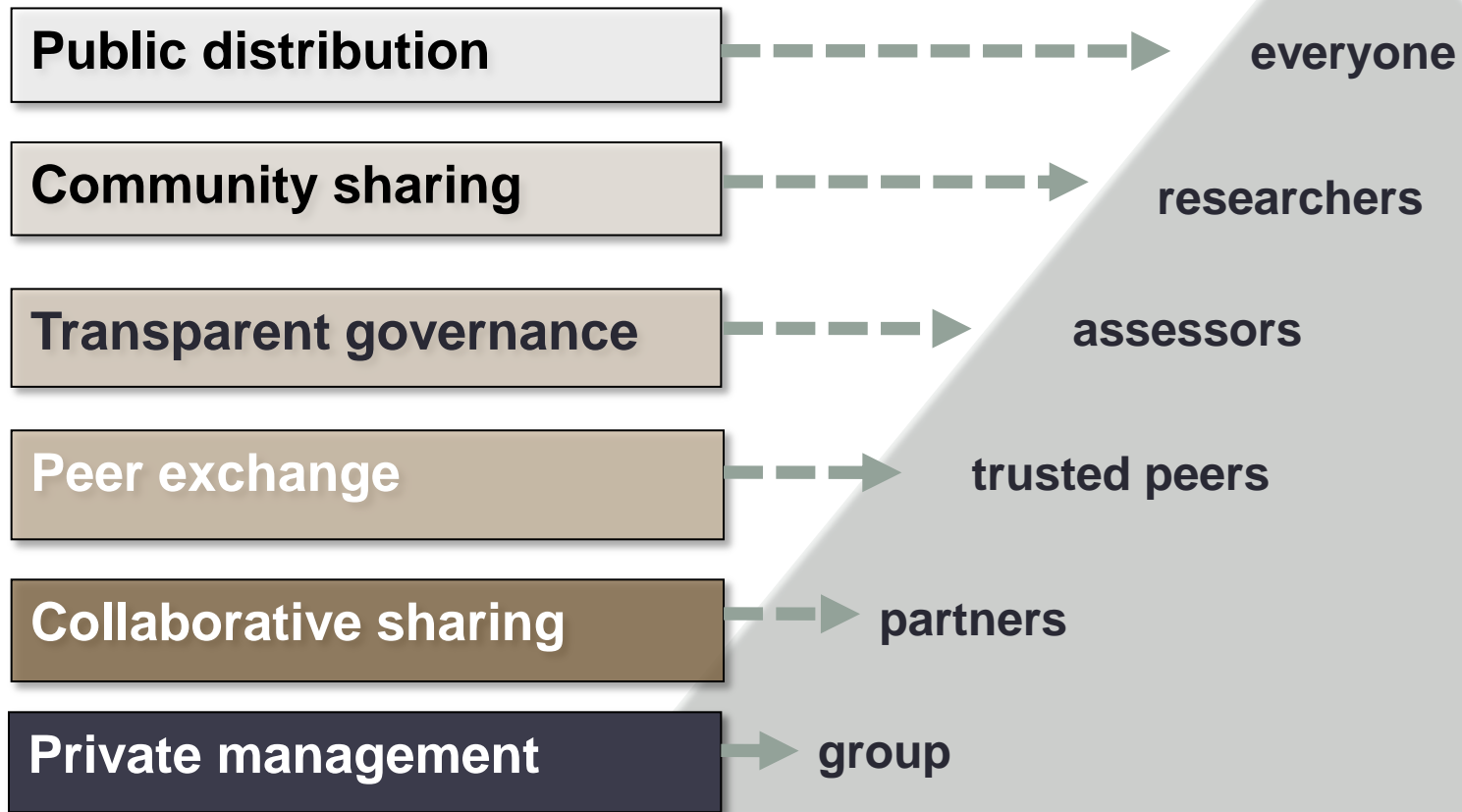
What is open, why? Who open to?

- Interviewed 18 researchers, 6 domains
- All claimed to be working openly to some degree
- All saw benefits in working that way
- What was made accessible and usable by others, and when?



Whyte, A., & Pryor, G. (2011). Open Science in Practice: Researcher Perspectives and Participation. *International Journal of Digital Curation*, 6(1), 199–213. doi:10.2218/ijdc.v6i1.182

Degrees of openness



*“Degrees of openness - extremely important - different people work in different ways and have different constraints imposed upon them”
(Chemistry, Senior Researcher)*



You can view some of the administrative data sources that might be right for your research project.

We've grouped them into themed areas, and given some information about the sources.

This is not a comprehensive list of all possible data sources. We welcome ideas and requests for other data sources which may be valuable for research. Please [get in touch](#) if you have any suggestions.

- All
- Business & Third Sector**
- Crime & Justice
- Economy & Employment
- Education & Learning
- Health & Wellbeing
- Population
- Housing & Environment

[888049 - Inter-Departmental Business Register, 1994/95-](#)

Summary

Applying for access

[888050 - Register of Charities, 1993-](#)

Summary

Applying for access

[888051 - Citizens Advice, 2008-](#)

Summary

Applying for access

Open data policy from the ‘top down’



“**Member states** are invited to:

- Harmonise access and usage policies for research and education-related public **e-infrastructures** ...

Research stakeholder organisations are invited to:

- Adopt and implement **open access** measures for publications and **data** resulting from publicly funded research”

Reinforced European Research Area Partnership for Excellence
and Growth COM(2012) 392 final

http://ec.europa.eu/euraxess/pdf/research_policies/era-communication_en.pdf

Data sharing policy from the ‘top down’



“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 that does not harm intellectual property.”

RCUK Common Principles on Data Policy 2011
<http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>



“To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be **open**...respecting concerns in relation to privacy, safety, security and commercial interests [and] legitimate concerns of private partners.”

G8 Science Ministers Statement- June 2013

But what makes data a ‘public good’?

| | | Subtractability / Rivalry | |
|-----------|-----------|---|--|
| | | Low | High |
| Exclusion | Difficult | Public Goods General knowledge Public domain data | Common-pool resources Libraries Data archives |
| | Easy | Toll or Club Goods Subscription journals Subscription data | Private Goods Printed books Raw or competitive data |

Christine Borgman, RDA Plenary 4 Keynote "*Data, Data, Everywhere, Nor Any Drop to Drink.*" 23.9.2014
<https://rd-alliance.org/plenary-2-session-chaired-cees-de-laet-university-amsterdam.html>

But what makes data a ‘public good’?

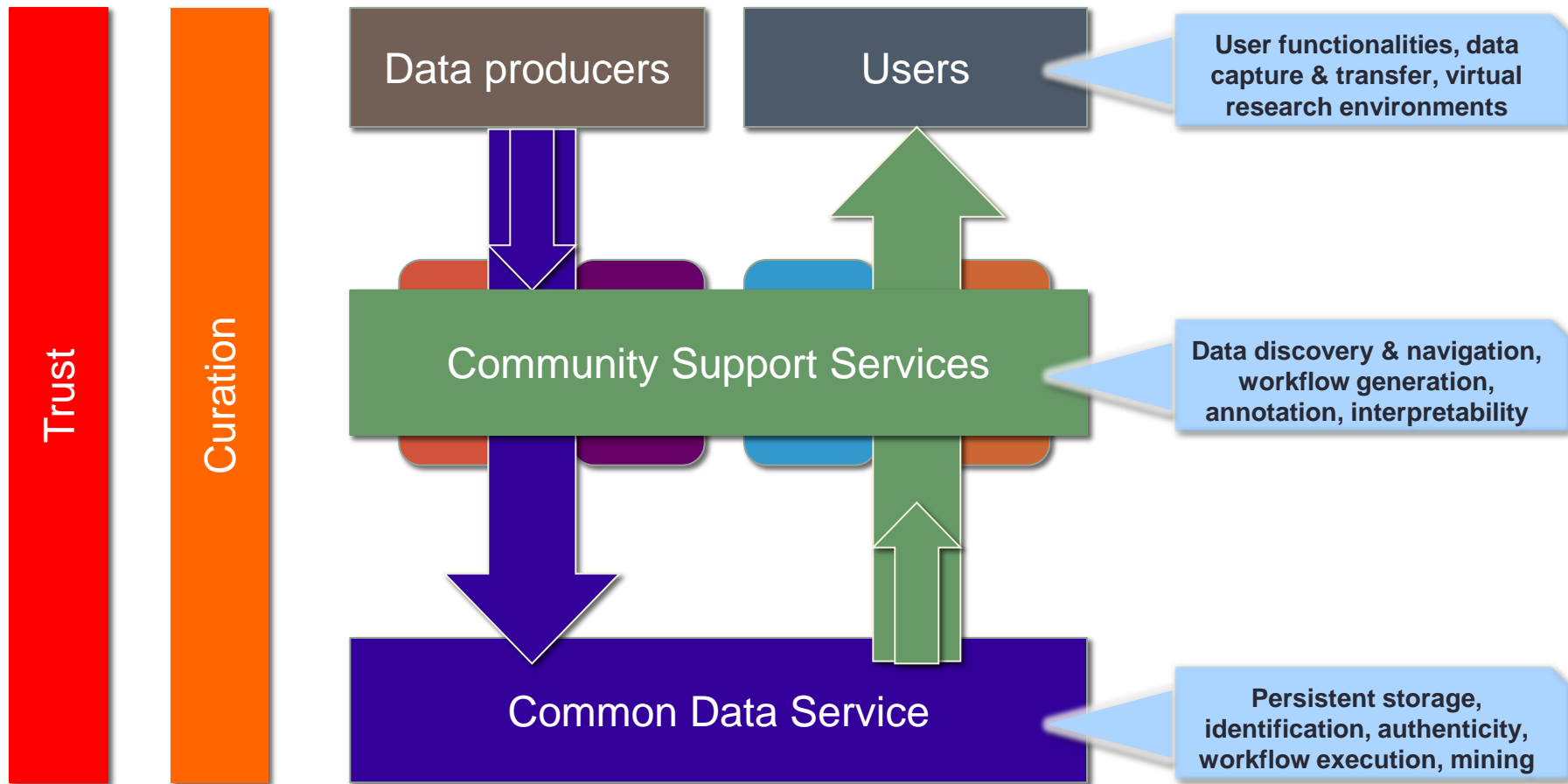
| | | Subtractability / Rivalry | |
|------------------|-----------|---|--|
| | | NO FIREWALL, PAYWALL OR LICENSE RESTRICTION | High |
| Exclusion | Difficult | Public Goods General knowledge Public domain data | Common-pool resources Libraries Data archives |
| | Easy | Toll or Club Goods Subscription journals Subscription data | Private Goods Printed books Raw or competitive data |

Christine Borgman, RDA Plenary 4 Keynote "*Data, Data, Everywhere, Nor Any Drop to Drink.*" 23.9.2014
<https://rd-alliance.org/plenary-2-session-chaired-cees-de-laet-university-amsterdam.html>

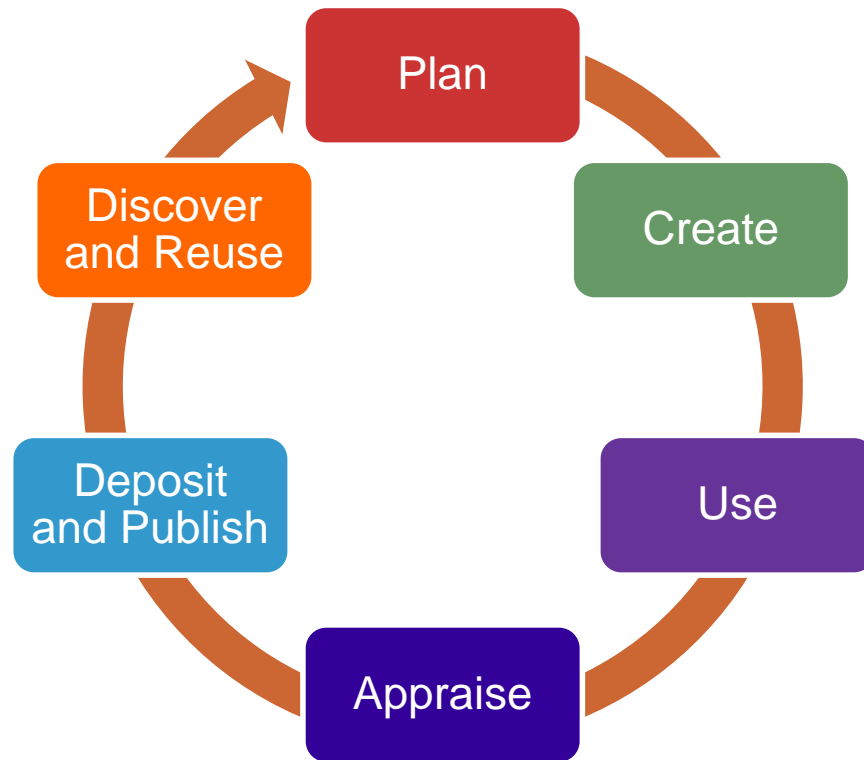
But what makes data a ‘public good’?

| | | Subtractability / Rivalry | |
|------------------|-----------|---|---|
| | | Low | INFRASTRUCTURE Common-pool resources Libraries Data archives |
| Exclusion | Difficult | Public Goods General knowledge Public domain data | |
| | Easy | Toll or Club Goods Subscription journals Subscription data | Private Goods Printed books Raw or competitive data |

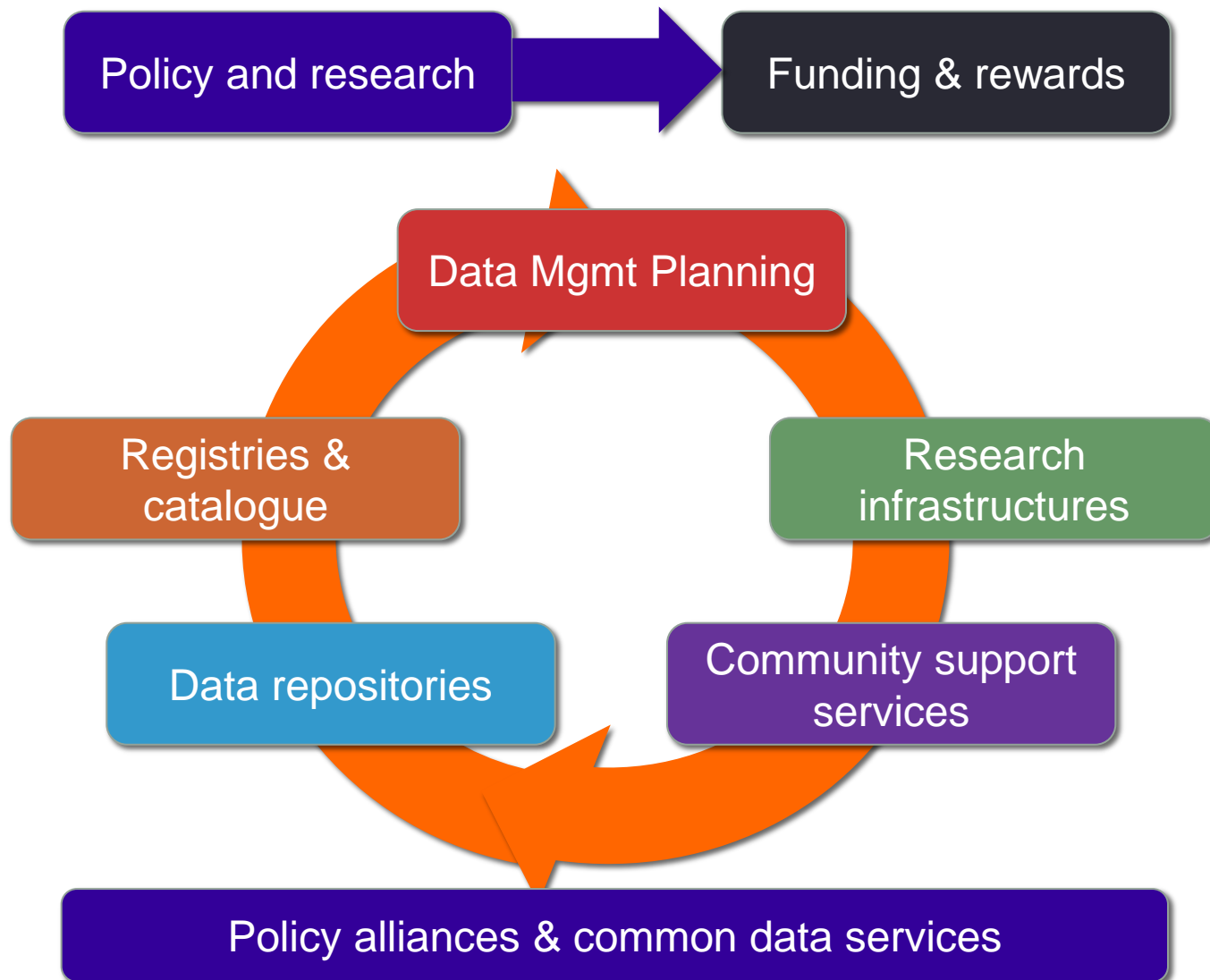
Collaborative Data Infrastructure



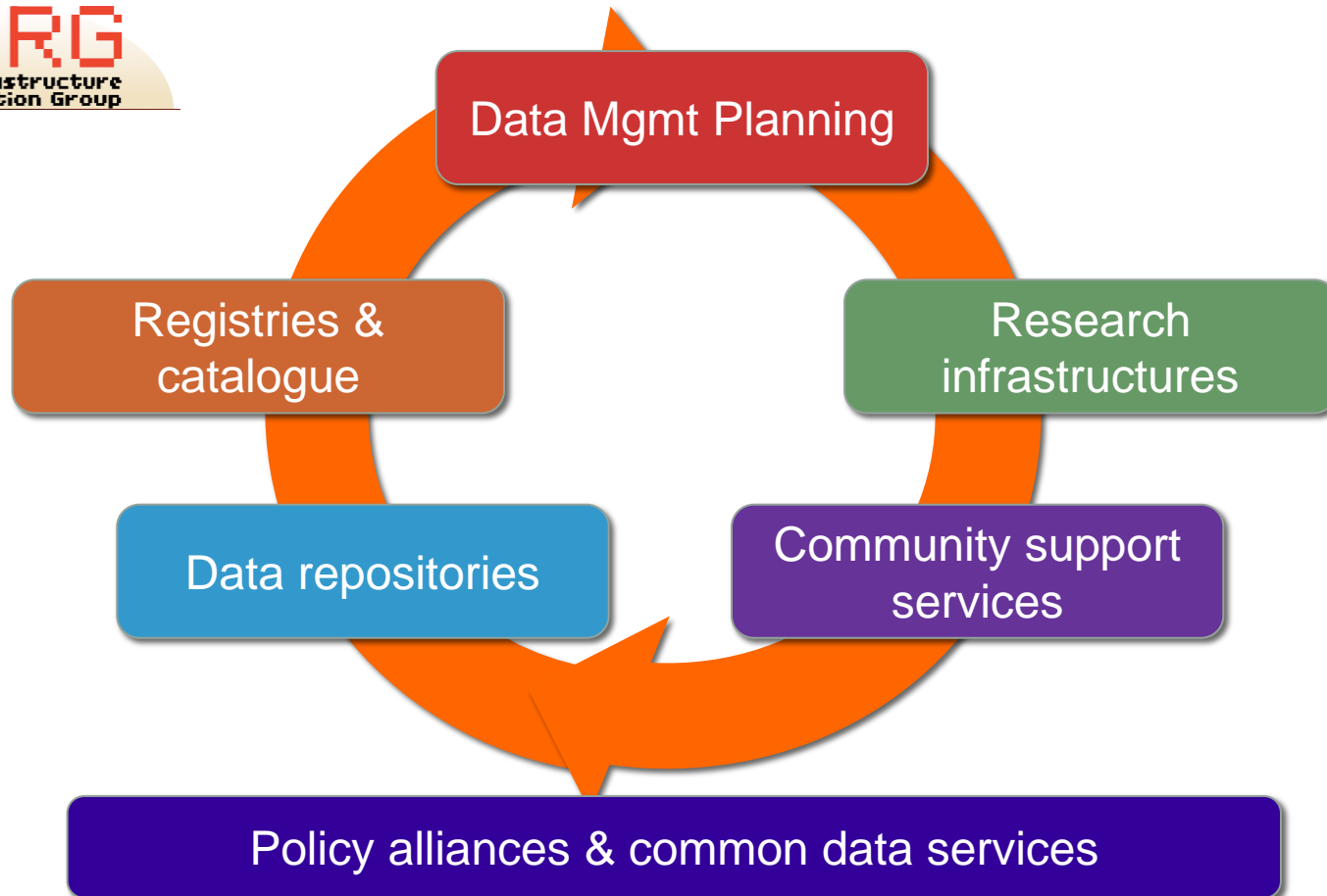
How does that support RDM lifecycle?



International infrastructure



International infrastructure





e-Infrastructure Reflection Group

Policy and research

Common policy on e-infrastructure development

Start News Events Publications Downloads Knowledge Base About Delegates Space

Start



e-Infrastructure Reflection Group

Upcoming e-IRG events

More information will follow soon...

- e-IRG meeting: 16 February (Riga)
- e-IRG workshop:
- e-IRG meeting:

Meet us at

- Research Data Alliance and Global Data and Computing e-Infrastructure Challenges, December 11-12 2014

Calendar on more e-infrastructure events, [read more](#)

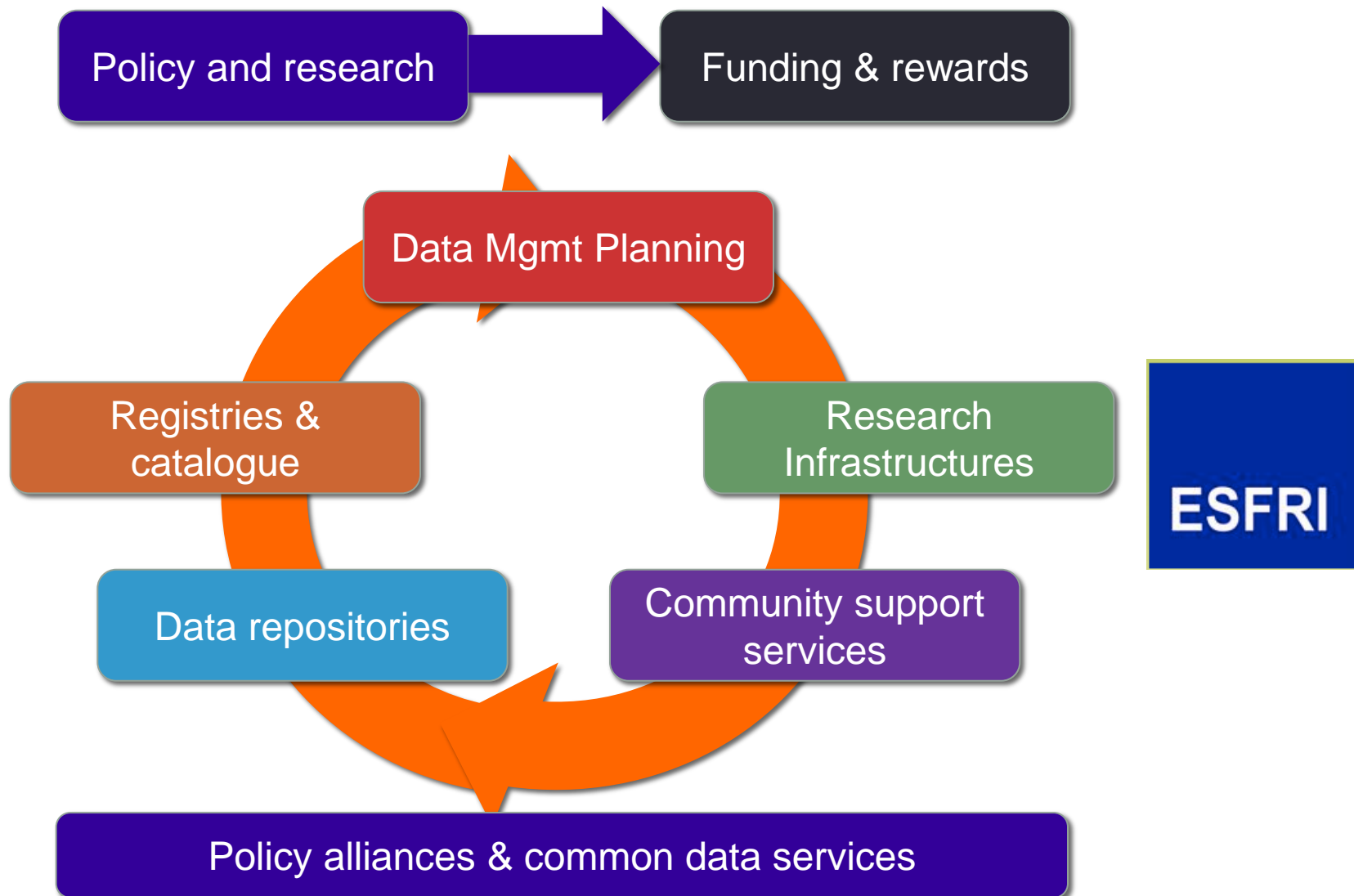
News

New e-IRG web presence

By Ad Emmen, On 04/12/14 09:24

As you may have noticed, the e-IRG has a new web presence. Over the years the web site has grown considerably, but the design stayed the same. With the new design we tried to organise the information in a more logical

International infrastructure





RESEARCH & INNOVATION Infrastructures

Research
Infrastructures

Coordinates EU
Research
Infrastructures

European Commission > Research & Innovation > Research infrastructures > Interactive map

Map of research infrastructures



Research
Infrastructures

Display: Map

Listing

Search:

Filter

HOME

WHAT ARE RIs ?

MAP of RIs

THE EUROPEAN LANDSCAPE

EU FINANCIAL SUPPORT

ERIC-LEGAL FRAMEWORK

SYNERGIES - EU INITIATIVES

INTERNATIONAL
COOPERATION

ESFRI

Press corner

Slovenia

RI name & associated project

JSI TRIGA REACTOR

AIDA Advanced European Infrastructures for Detectors at Accelerators

SLOANMR

Bio-NMR NMR for Structural Biology

SLOANMR

EAST-NMR Enhancing Access and Services To East European users towards an efficient and coordinated panEuropean pool of NMR capacities to enable global collaborative research & boost technological advancements

P3-IMI-MF

EVA European Virus Archive

NODC-SI

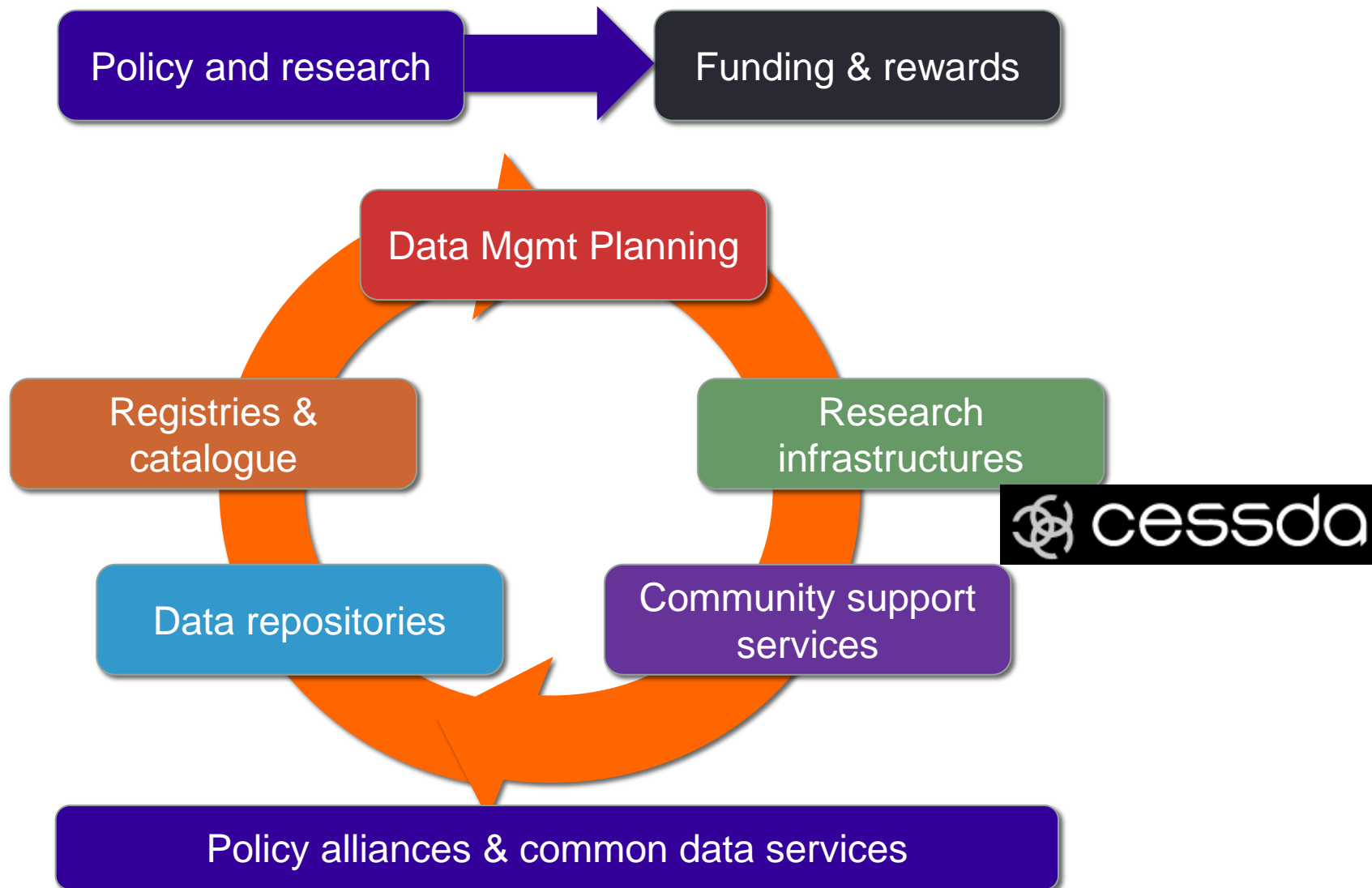
SeaDataNet II SeaDataNet II: Pan-European infrastructure for ocean and marine data management

TANIJS

SPIRIT Support of Public and Industrial Research using Ion Beam Technology

http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=mapri

International infrastructure



Expert Seminars

CESSDA Expert Seminars are held annually. The seminars concentrate on practical information and on various aspects of data archiving.

| Year | Host | Location | Theme |
|------|-----------------|-----------------------|---|
| 2014 | UK Data Archive | Colchester, UK | Research integrity - a digital curation problem? [website] |
| 2013 | WISDOM | Vienna, Austria | Research Data Management: Cost and Incentives |
| 2012 | FSD | Tampere, Finland | Data Acquisition and Licence Agreements [website] |
| 2011 | FORS | Lausanne, Switzerland | Question data banks - towards greater integration? [website] |
| 2010 | SND | Gothenburg, Sweden | Cooperation between CESSDA archives and NSIs [website] |
| 2009 | ADP | Ljubljana, Slovenia | Towards the CESSDA-ERIC metadata model and DDI3 [website] |
| 2008 | DDA | Odense, Denmark | Cessda, DDI, Tools and Methods [website] |
| 2007 | UKDA | Colchester, UK | Developing the CESSDA Portal Further [programme] |
| 2006 | GSDB | Athens, Greece | Open Access to Data: Anonymisation, Data Protection & Confidentiality [website] |

Data without Boundaries Training course on EU Census Microdata - IECM

Data without Boundaries, in cooperation with Eurostat, organises a course on the Integrated European Union Census Microdata (IECM) in Barcelona 21 – 23 January 2015, hosted by Centre d'Estudis Demogràfics, Universitat Autònoma de Barcelona. [\[More..\]](#)

CESSDA - Introducing the new team

In June 2013 CESSDA was established as a permanent legal entity owned and financed by the individual member states' ministry of research or a delegated institution. Norway is hosting CESSDA, and the main office is located in Bergen. But who are the faces behind this new infrastructure? A small team of three staff started work and moved into their office, CESSDA House in Bergen, in September 2014. [\[More..\]](#)

CESSDA Expert Seminar, UK Data Archive, 13-14 November 2014



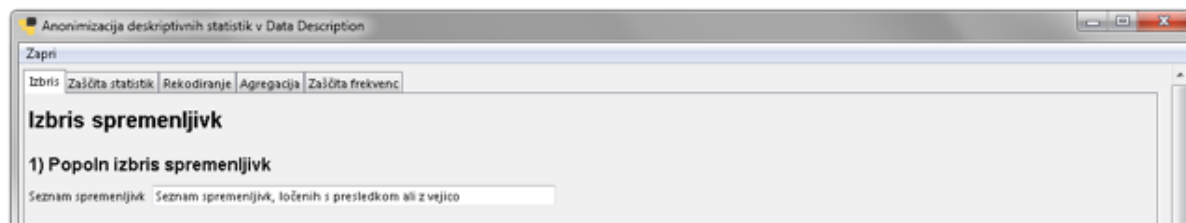
DDI Data Description protection software

DDI Data Description Statistics Protection Software je orodje, ki omogoča enostavno anonimizacijo univariatnih statistik, ki so vključene po DDI (Data Documentation Initiative) standardu. Orodje je bilo razvito, da bi rešilo problem zaščite zaupnosti pri distribuciji agregiranih podatkov posameznih spremenljivk. Ti so javno dostopni, da bi zagotovili koristne informacije znanstvenim raziskovalcem in s tem promovirali rabo mikropodatkov uradne statistike v znanstvene namene.

Funkcije orodja za avtomatsko zaščito statistik, vključenih v opis podatkov, temeljijo na tako metodah in tehnikah anonimizacije mikropodatkov, kot tudi na tistih uporabljenih pri zaščiti tabelarnih podatkov. Sledijo pa tudi posebnostim zaupnosti podatkov uradne statistike. Tehnike zaščite so tematsko razdeljene glede na tip anonimizacijske tehnike, ki se uporablja za zaščito agregiranih statistik.

Orodje omogoča hitro in enostavno avtomatsko zaščito podatkov z uporabo metod za zaščito podatkov kot so rekodiranje, zaščita najnižjih in najvišjih vrednosti, izbris (numeričnih informacij) spremenljivk, zaščita deskriptivnih statistik številskih spremenljivk in zaščita frekvenc vrednosti spremenljivk, pri čemer se upošteva pravilo najnižje dovoljene prikazane frekvence. V nasprotju z obstoječimi orodji za zaščito mikropodatkov in agregiranih podatkov to razvito orodje zaščiti agregirane podatke direktno v XML kodi.

>> Prenesi orodje




Search

Search the ADP website

 **Seminar for researchers:
Preparing research data for
open access**



 **ADP also preserves study...
Tolerance, CATI**

 **Open data project report**

Useful links

ADP is founded by Ministry of Higher Education, Science and Technology in scope of infrastructure program "The Research Infrastructural Centres Network of University of Ljubljana." Social Science Data Archives is a member of CESSDA, the organisation for social science data archives across Europe.

What is DDI?

Last Updated: Thu, 2013-11-21 15:24 — stephanie.roth

About the DDI standards

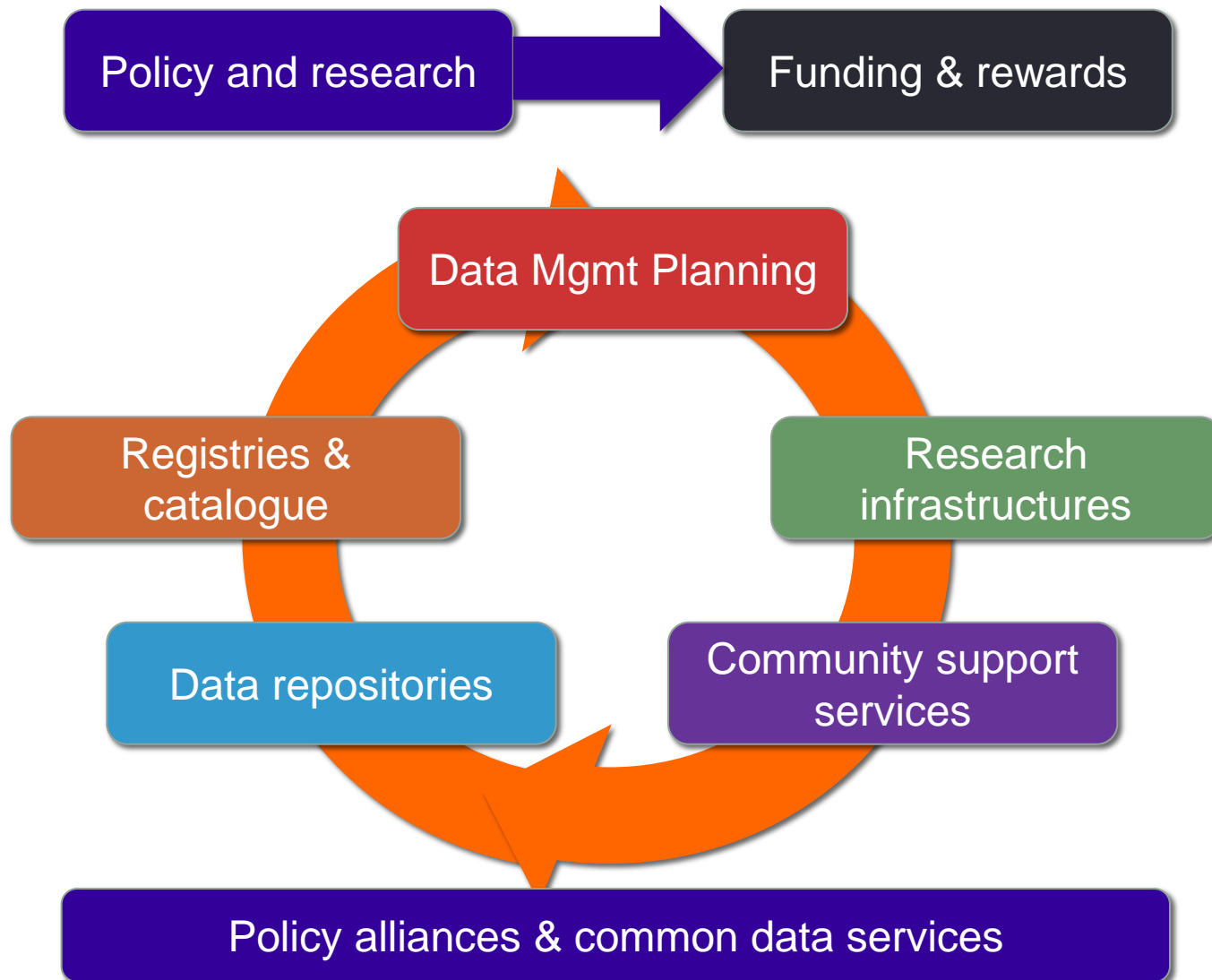
The **Data Documentation Initiative (DDI)** is an effort to create an international standard for describing data from the social, behavioral, and economic sciences. Expressed in [XML](#), the DDI metadata [specification](#) now supports the entire research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving.



CESSDA uses and supports the DDI international metadata standard

Enabling better access and reuse, e.g. find which longitudinal studies have asked which questions

International infrastructure



What do you think?

Should ResourceType be a required field? Your opinion counts!

[Read more](#)



WHAT CAN DATACITE DO FOR YOU?



CITE YOUR DATA



FORMAT YOUR CITATION



FIND A REPOSITORY

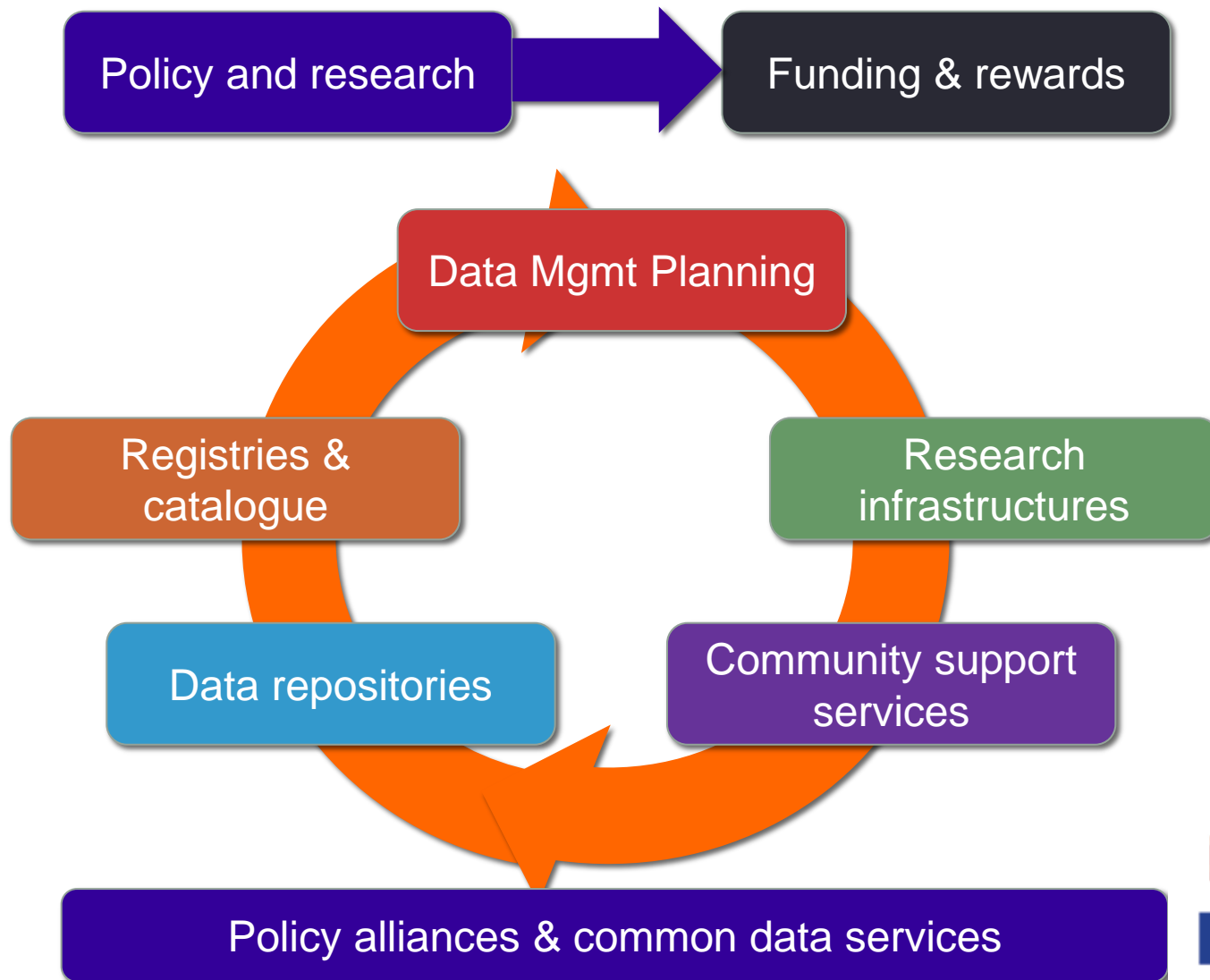


FIND A DATASET



GET YOUR DOI STATISTICS

International infrastructure



Common data services



Home Services & Support EUDAT Communities EUDAT Events Working Groups News & Publications

Search

B2DROP

Sync and Exchange Research Data

- What is B2DROP
- Use B2DROP
- Contact



B2SHARE

Store and Share Research Data

- What is B2SHARE
- Use B2SHARE
- Contact



B2SAFE

Replicate Research Data Safely

- What is B2SAFE
- Contact



B2STAGE

Get Data to Computation

- What is B2STAGE
- Contact



B2FIND

Find Research Data

- What is B2FIND
- Use B2FIND
- Contact

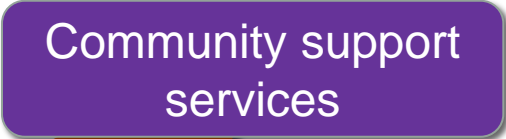
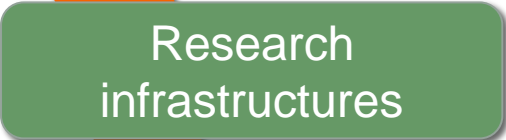
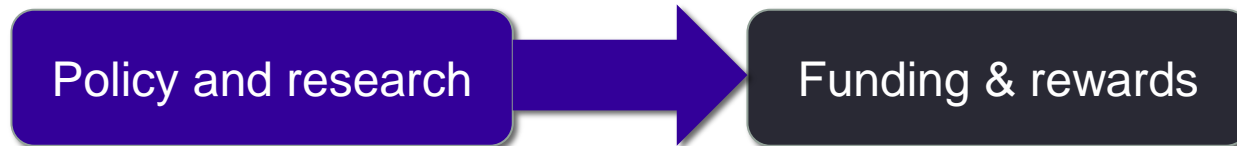


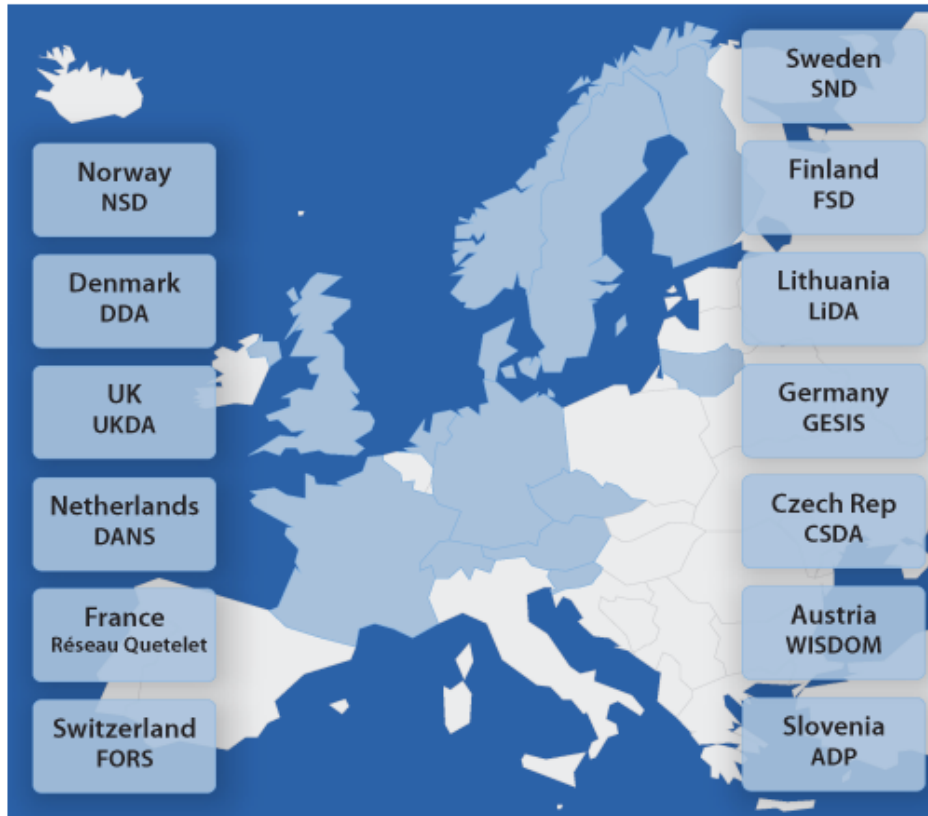
IN FOCUS

Want to know more about EUDAT's B2 services?

Find out more about the B2 services - B2DROP, B2SHARE, B2SAFE, B2STAGE and B2FIND - in this presentation.

International infrastructure





Registries & catalogue

Access across national social science data repositories

Data repositories

CESSDA

New pan-European Research Infrastructure

<http://www.cessda.net/>

Search



English

- [-] CESSDA Catalogue
 - [+] Browse by Topic
 - [+] Browse by Keyword
 - [+] Browse by Data Publisher
 - ADP
 - ADPSS-Sociodata
 - APIS
 - CSDA
 - DANS
 - FORS
 - FSD
 - GESIS ZACAT
 - GSDB
 - ISSDA
 - LiDA
 - NSD
 - NSD Metadata
 - Quetelet-ADISP
 - Quetelet-CDSP
 - SND
 - UKDA

Search Term: **ADP**

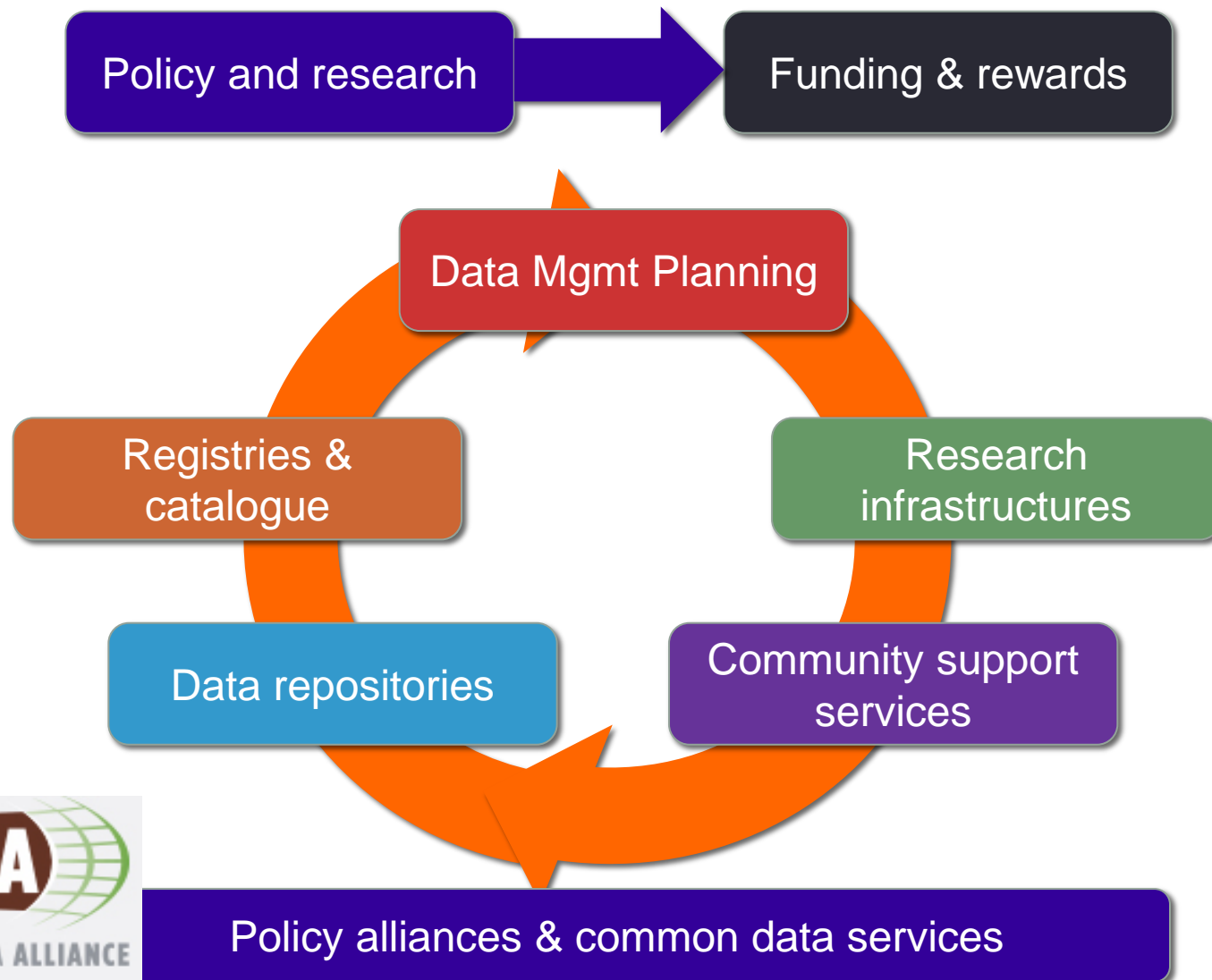
Study Section Variable

| | Study |
|---|--|
| [ABR97] ¹ Mednarodna raziskava o veri in odnosu do cerkve 1997 |  |
| [ADS] ³ Anketa o delovni sili : Opis serije |  |
| [ADS001] ³ Anketa o delovni sili, 2000/1 |  |
| [ADS002] ³ Anketa o delovni sili, 2000/2 |  |
| [ADS003] ³ Anketa o delovni sili, 2000/3 |  |
| [ADS004] ³ Anketa o delovni sili, 2000/4 |  |
| [ADS01] ³ Anketa o delovni sili, 2001 : Letna datoteka - brez omejitev zaupnosti |  |
| [ADS02] ³ Anketa o delovni sili, 2002 : Letna datoteka - brez omejitev zaupnosti |  |
| [ADS03] ³ Anketa o delovni sili, 2003 : Letna datoteka - brez omejitev zaupnosti |  |
| [ADS04] ³ Anketa o delovni sili, 2004 : Letna datoteka - brez omejitev zaupnosti |  |

1-10 of 1121 | Next >

ADP are the central Social Science Data Archives in Slovenia.

International infrastructure





Research Data Sharing
without barriers

Policy alliances

Practical policy-
problem solving

- Home
- Organisation
- Working and Interest Groups
- Plenary Meetings
- News & Events
- Early Career Programmes
- About

**Research Data Alliance and global Data and Computing
e-Infrastructure challenges**
11&12 December 2014, Rome, Italy

**Plenary 5
FIND OUT MORE & REGISTER**




Call for Session Proposals

**CALL FOR EXPRESSIONS OF INTEREST
TO ORGANISE AN RDA PLENARY MEETING**



WORKING GROUPS

- 
The BioSharing Registry: connecting data policies, standards & databases in life sciences
 32
 Susanna-Assunta Sansone, Rebecca Lawrence, & Simon Hodson
- 
Urban Quality of Life Indicators
 29
 Chris Pettit, Massimo Craglia, Piyushimita (Vonu) Thakuriah
- 
Wheat Data Interoperability WG
 27
 Esther Dzalé Yeumo, Richard Fulss

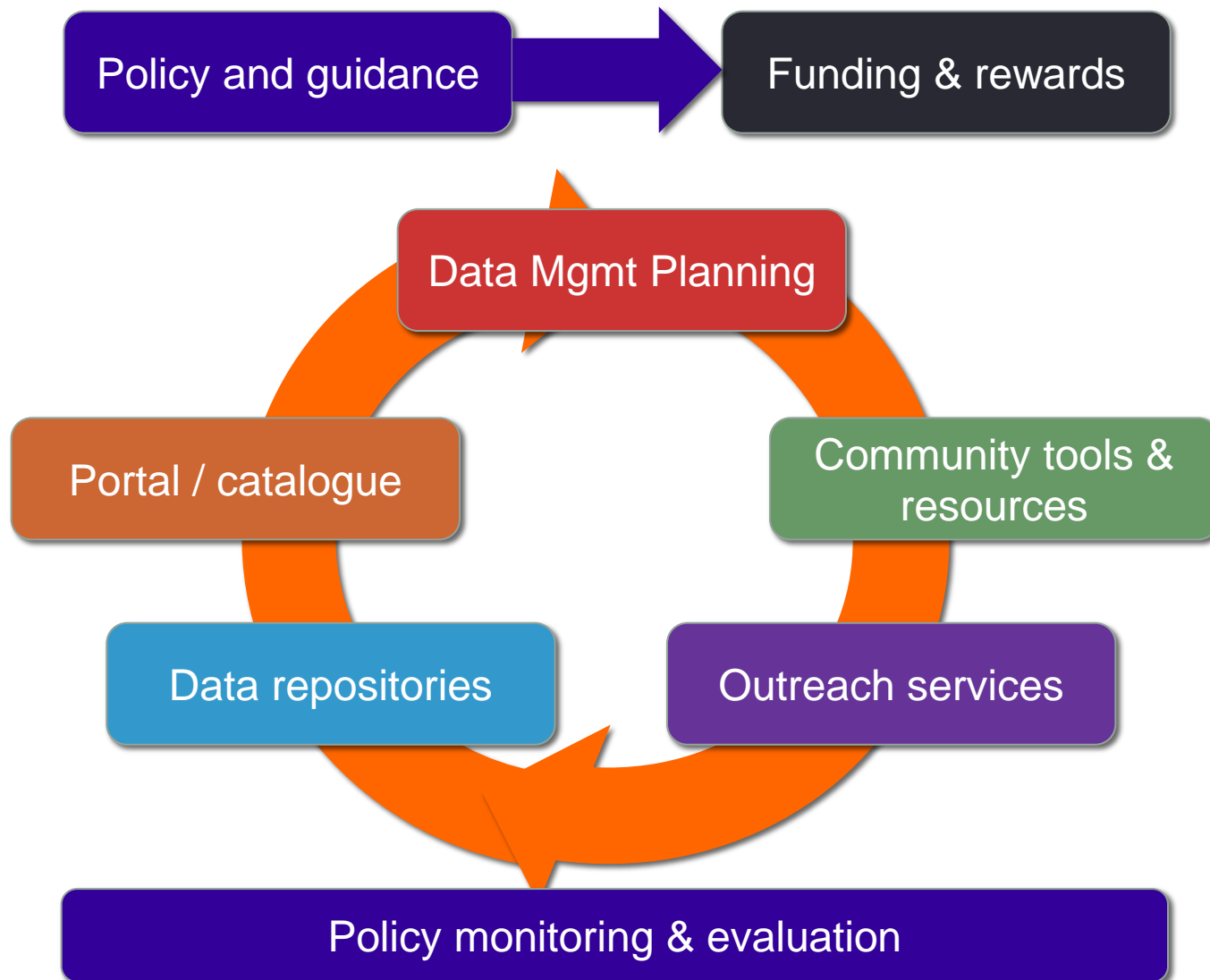
INTEREST GROUPS

- 
Research data needs of the Photon and Neutron Science community
 36
 Amber Boehnlein, Brian Matthews, Frank Schlünzen, Thomas Proffen
- 
Research Data Provenance
 59
 David Dubin, Sayeed Choudhury, Bridget Almas
- 
Service Management IG
 15
 Owen Appleton, Michael Brenner, Thomas Schaaf

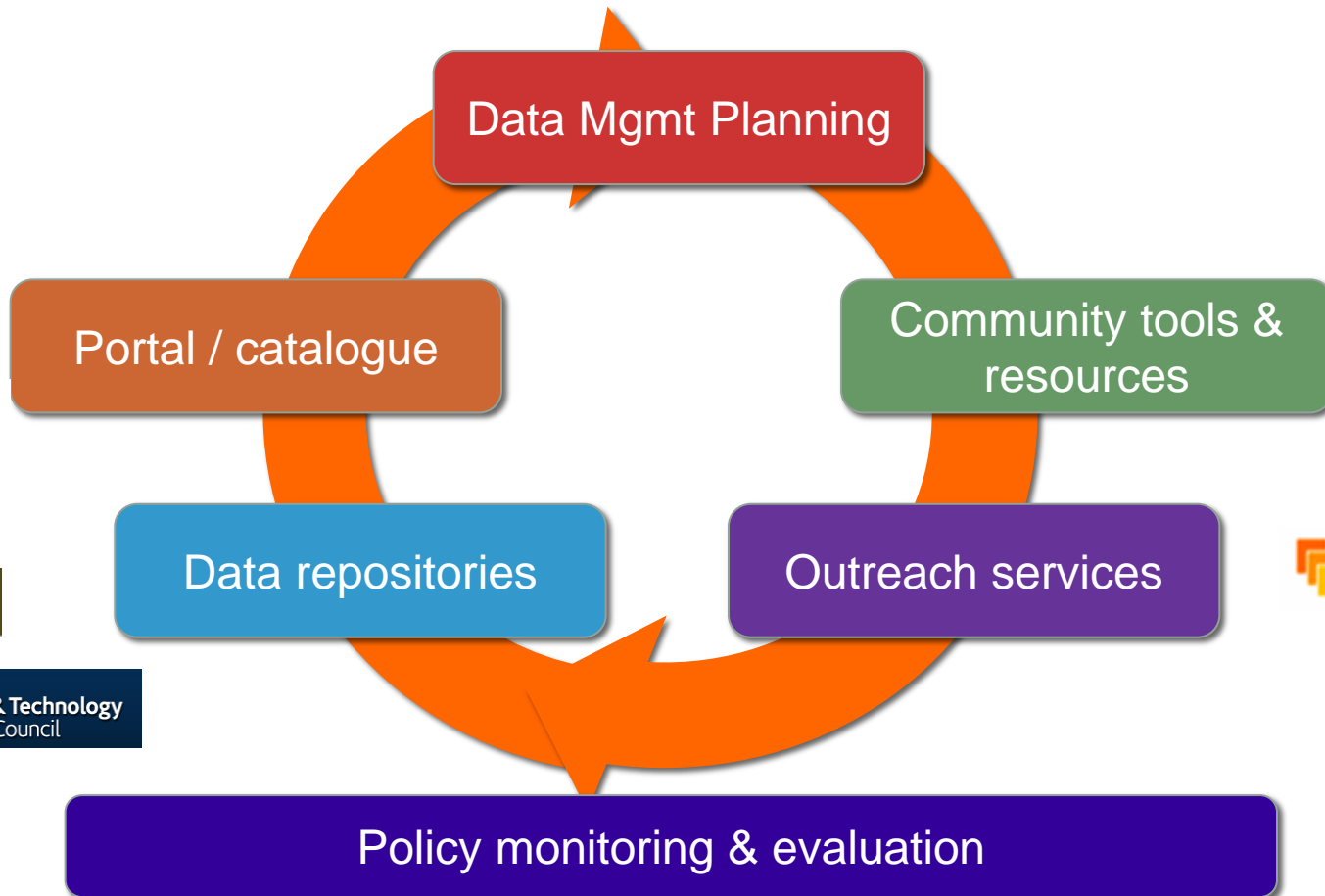
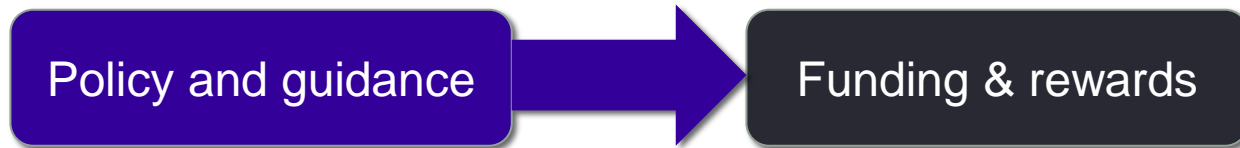
BOF GROUPS

- 
Data re-use, share your experiences
 21
 Odile Hologne
- 
Sustainability of eResearch / Cyberinfrastructure
 7
 Stefanie Kethers, Andrew Treloar

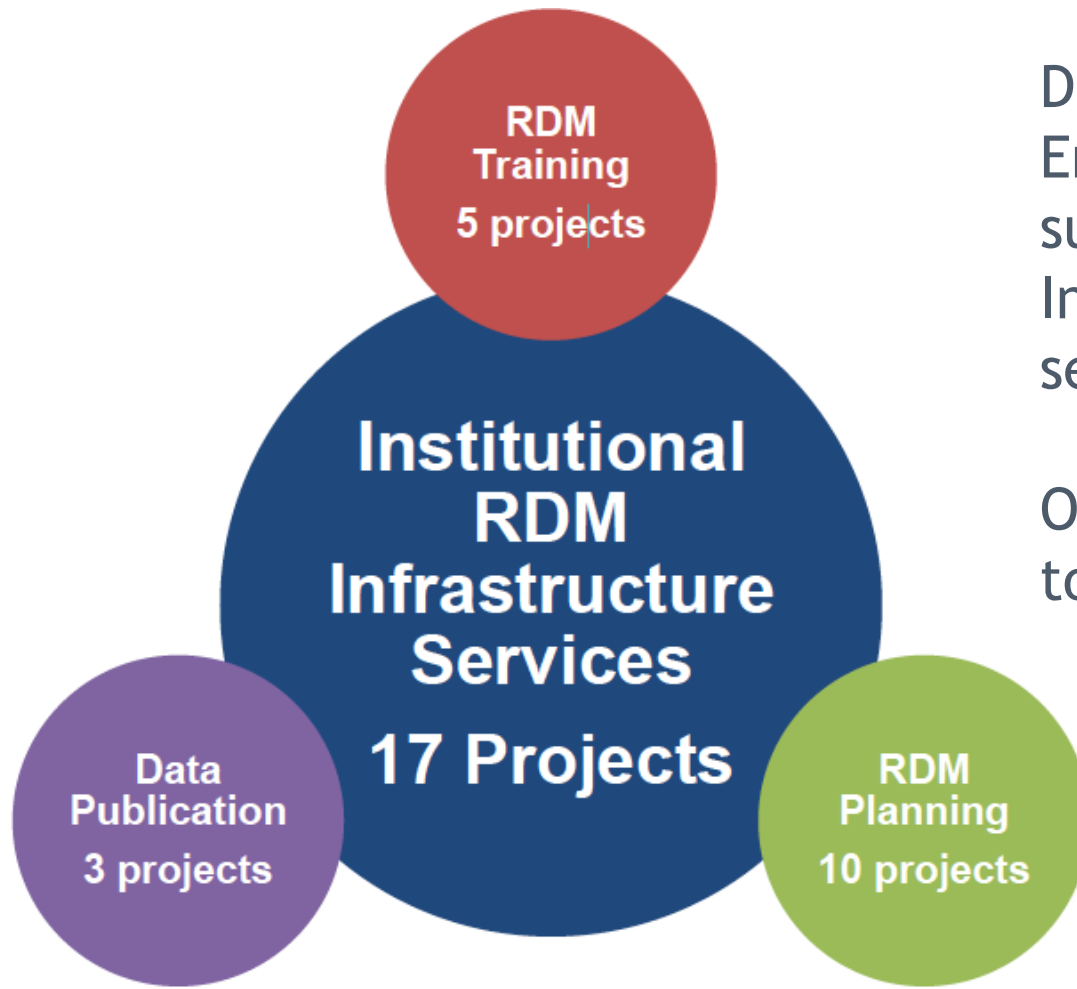
National infrastructure



National infrastructure



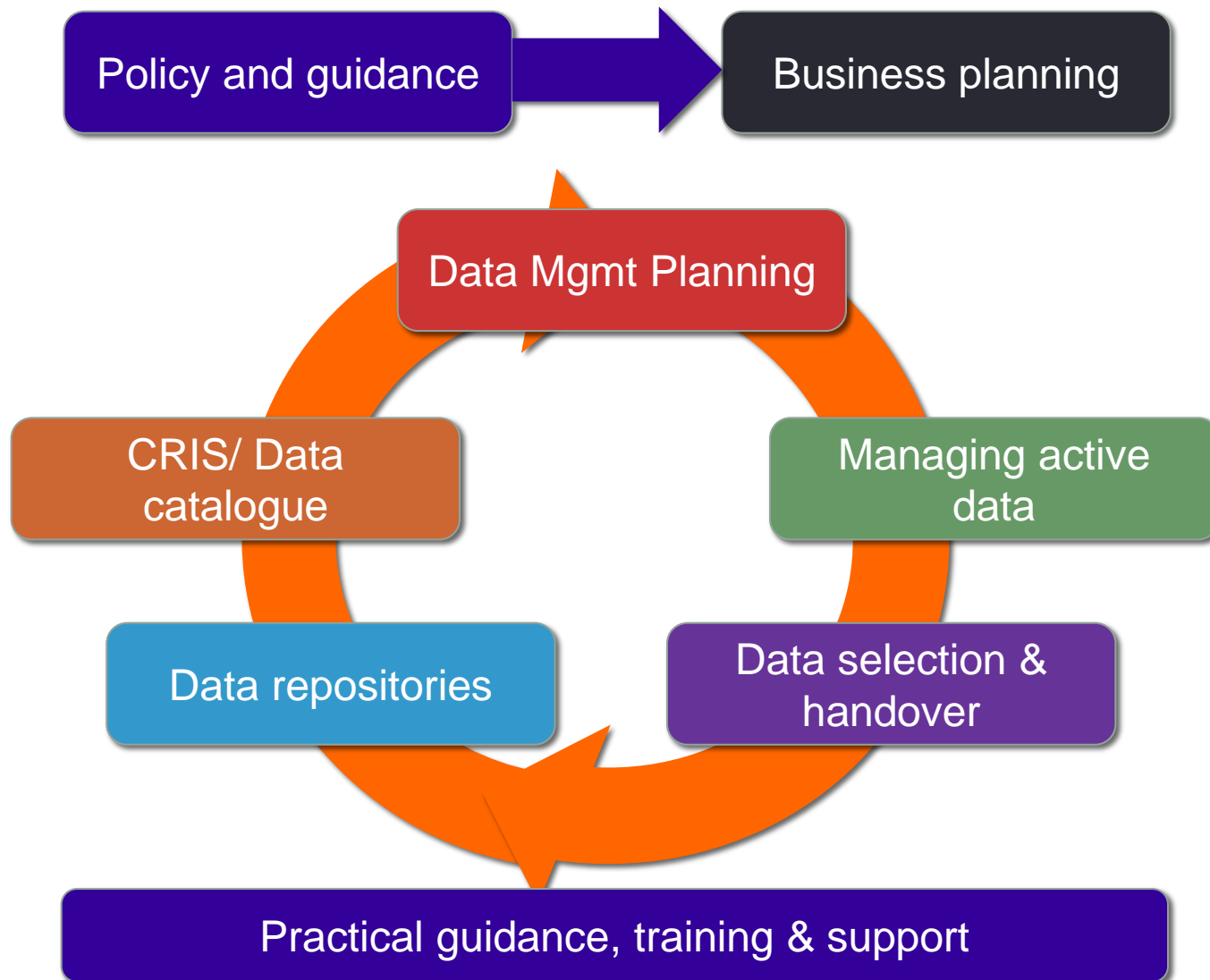
Jisc Managing Research Data Programme



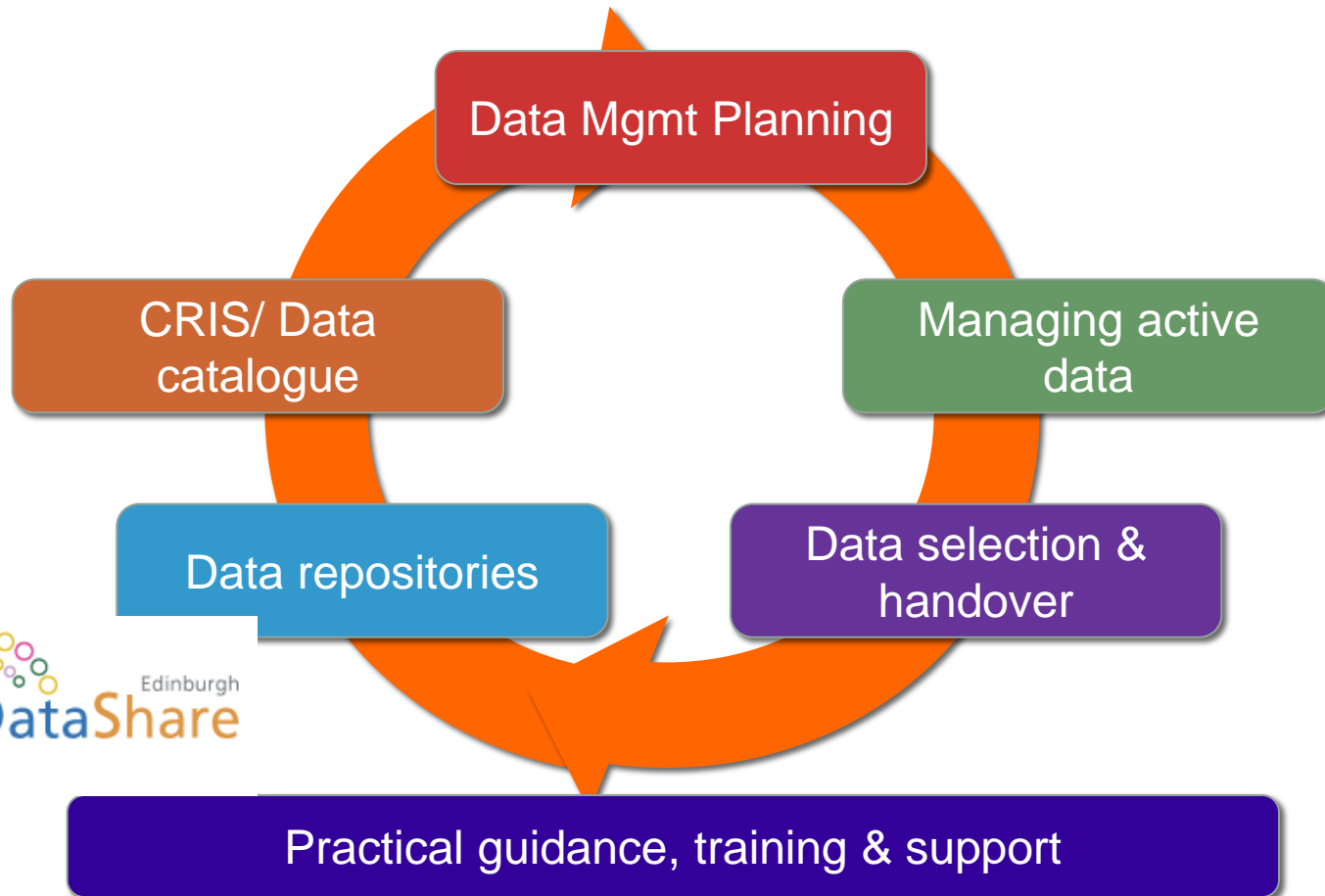
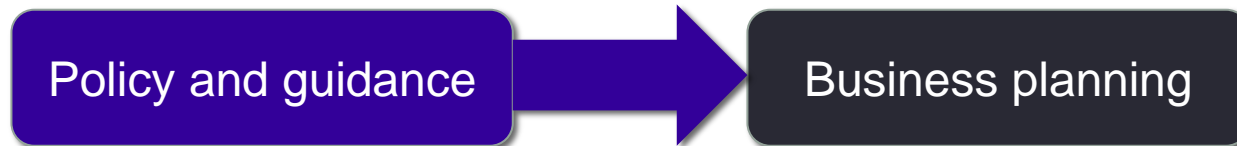
DCC Institutional Engagement in parallel - support to further 21 Institutions to set up services 2011-13

Ongoing tailored support to 38 since 2013

Institutional infrastructure



Institutional infrastructure



Information Services

[Contact Us](#) [Login](#)



[Edinburgh DataShare](#) > [Moray House School of Education](#) > [Centre for Educational Sociology \(CES\)](#) > [Scottish young people's surveys](#) > [View Item](#)



1979 national school leavers survey

[Download Dataset](#)

[Show full item record](#)

Search Edinburgh DataShare

Search Edinburgh DataShare
 This Collection

- Browse**
- Edinburgh DataShare
 - [Research Communities](#)
 - [Date Issued](#)
 - [Data Creators](#)
 - [Titles](#)
 - [Subjects](#)
 - This Collection
 - [Date Issued](#)
 - [Data Creators](#)
 - [Titles](#)
 - [Subjects](#)

My Account

[Login](#)
[Register](#)

| | |
|------------------------|--|
| Title: | 1979 national school leavers survey |
| Data Creator: | Ruus, Laine; Scottish Education Department. Scottish Office; Social Science Research Council; Manpower Services Commission |
| Date Available: | 2014-07-22 |

Citation: University of Edinburgh. Centre for Educational Sociology. (2014). 1979 national school leavers survey, 1979 [Dataset]. <http://dx.doi.org/10.7488/ds/58>.

Dataset Description (abstract): The 1979 survey covered young people who had left school in the session 1977/78. It was based on a smaller sample than the earlier 1977 survey: 10 per cent of O and H grade leavers, and 20 per cent of non-qualified * leavers, compared with a sample of 40 per cent of all leavers in 1977. The questionnaires were based on those used in 1977, and repeated most of the earlier items. Several new questions were added and others were revised, and the non-qualified leavers' questionnaire was substantially re-designed to include extra items on job-seeking, early employment experience and the Youth Opportunities Programme. In this way the 1979 survey has extended and updated the time series in the Scottish Education Data Archive. It also had four more specific objectives: (i) to obtain information on additional topics (covered by the new questions); (ii) to obtain information on non-Certificate leavers in Regions where they were not surveyed in 1977; (iii) to obtain information on changes between 1977 and 1979 with respect to the items common to both surveys; (iv) to provide contextual information for analyses which require large sample numbers and which will therefore need to be conducted on the 1977 data.

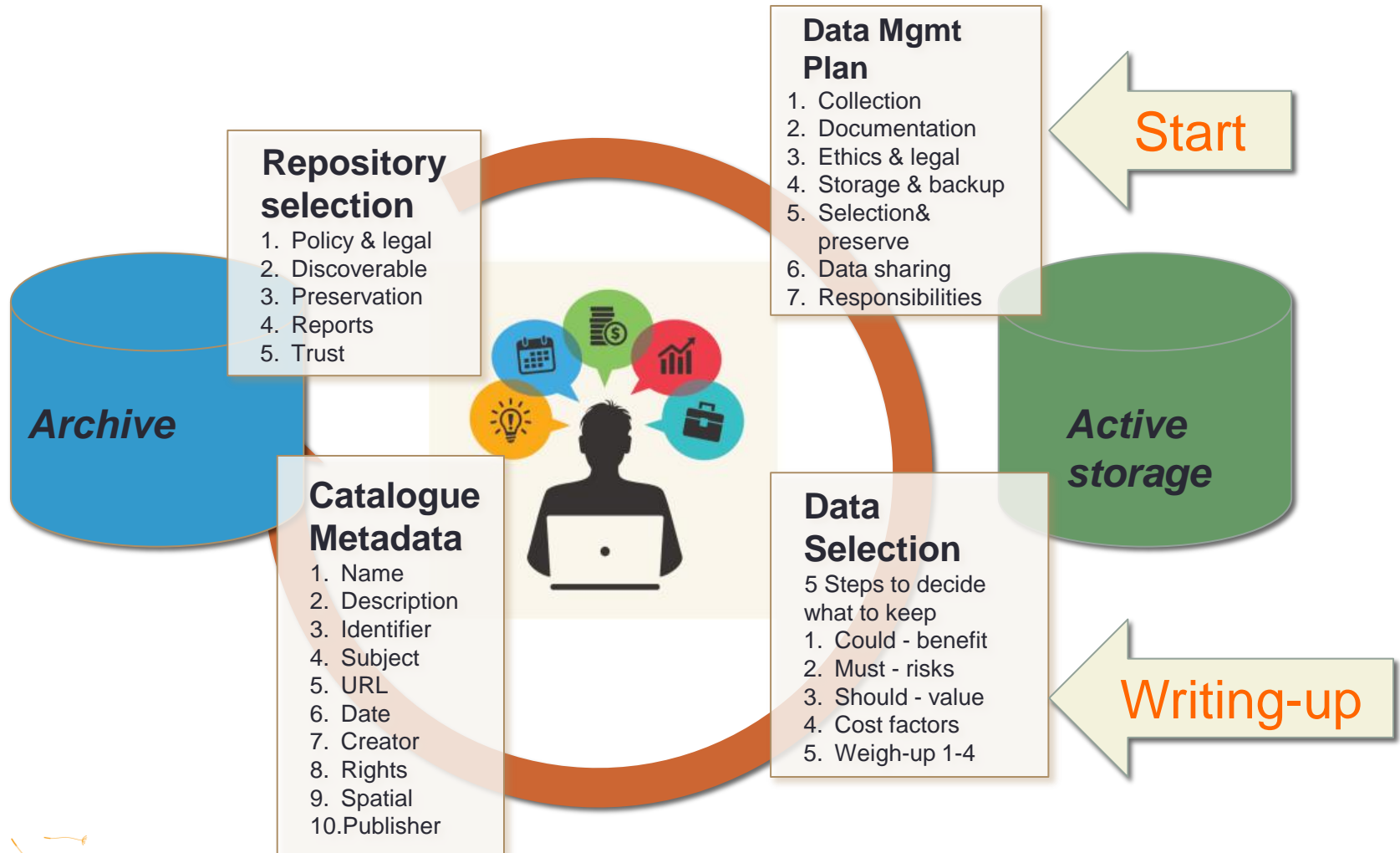
| Files | Size | Format | View | Description |
|---|---------|--------|--------------------------|----------------|
| Collaborative_Research_Dictionary_1979.pdf | 5.857Mb | PDF | Download | Codebook |
| Collaborative_Research_Questionnaire_1979.pdf | 1.698Mb | PDF | Download | Questionnaires |

Institutional infrastructure



Practical checklists

key points in research cycle



Data selection checklist

A Digital Curation Centre
'working level' guide



Five steps to deciding what data to keep

DCC Checklist for Appraising Research Data

Version 1

Please cite as: DCC (2014) 'Five steps to deciding what data to keep: a checklist for appraising research data v.1 Edinburgh: Digital Curation Centre. Available online: www.dcc.ac.uk/resources/how-guides

Digital Curation Centre, October 2014

This work is licensed under Creative Commons Attribution BY 2.5 Scotland, except Section 4, which is adapted under licence CC-BY-NC-SA, from UK Data Archive (2013) *Data management costing tool and checklist*. Available at: www.data-archive.ac.uk/create-manage/planning-for-sharing/costing



<http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>

Straightforward steps

- ① **Could** this data be re-used
- ② **Must** it be kept to manage compliance risk
- ③ **Should** it be kept for its potential value and...
- ④ Consider **costs**
- ⑤ Will ✓ or **won't** ✗ it be kept, shared on what terms



Step 1 (?) What 'must' be kept?

Research record includes data as evidence for
e.g. ...

- Audit purposes
- Health & Safety (Lab book)
- Contractual requirement

Compliance also about data that **won't** be kept,
or only shared with approved researchers...

Research Ethics, Duty of Confidentiality, Data Protection Act, Human Rights Act,
Statistics & Registration Services Act. UK Data Archive:

<http://www.data-archive.ac.uk/create-manage/consent-ethics/legal>

Step 1 (?) What 'must' be kept?

Data may be part of research records for compliance

- Audit
- Health & Safety (e.g. Lab book)
- Contractual obligations

Compliance also about data that **won't** be kept, or only shared with approved researchers...

Even where there are legal requirements to keep, or dispose of data, investigator makes initial selection of data that fulfils the research purpose

Step 1 (?) What 'must' be kept?

Funder & journal data policies expect some value judgement

“Data with **acknowledged long-term value**” *Research Councils UK
Common Principles on Data Policy*

“Data, information and other electronic resources **of long-term interest**” *Economic Social Research Council UK Data Archive
Collections Development Policy*

“An inherent principle of publication is that others should be able to **replicate and build upon the authors' published claims.** [Nature](#)”

Step 1 (?) What 'must' be kept?

Funder & journal data policies expect **value judgement**

“Data with **acknowledged long-term value**” *Research Councils UK Common Principles on Data Policy*

“Data, information and other electronic resources **of long-term interest**” *Economic Social Research Council UK Data Archive Collections Development Policy*

“An inherent principle of publication is that others should be able to **replicate and build upon the authors' published claims.** [Nature](#)

So the first step should be establishing

1. what are my **community's expectations** on verification/ replication (journals, repositories, societies)
2. what **other purposes** could the data be reused for?

Step 2-1 What *could* it be reused for?

Any purposes (potential benefits) not already considered?

1. Verification (**community expectations**)
2. Further analysis (**data linking, collaboration**)
3. Visibility (**impact, citation, credit**)
4. Resource development (**funding**)
5. Further publications /data articles (**citation**)
6. Learning and teaching materials (**credit**)
7. Private reference (**exploitation**)

Step 2-1 What *could* it be reused for?

Any purposes (potential benefits) not already considered?

1. Verification (community expectations)
2. Further analysis (data linking, collaboration)
3. Visibility (reputation, citation, credit)
4. Resource development (funding)
5. Further publications /data articles (citation)
6. Learning and teaching materials (credit)

From what is **desirable**, what is likely to be **feasible**? What data, documentation, software, other resource would you need to keep to make it happen later?

Step 3 What data *should* have value

Does it meet any two of these criteria?

1. Good quality - data and description

complete, accurate, reliable, valid, representative etc

2. High demand

known users, integration potential, reputation, recommendation, appeal

3. High effort to (re)produce

difficult, costly, or impossible to reproduce

4. Low barriers to reuse

legal/ ethical, copyright non-restrictive terms and conditions

5. Rarity value

unique copy or other copies at risk

Step 4 Consider cost factors

Costs already met may **add** to data value

Question is, can you afford to do the minimum to ensure that value is not lost when research ends?

1. Creation, collection & cleaning
2. Short-term storage & backup
3. Short-term access & security
4. Team communication & development
5. Preservation & long-term access



Step 5 Bring it all together

Balance risks, costs and value

Document the choices made

1. Dataset name, contributors, description, sensitivity - metadata
2. Reuse purposes and value - the 'reuse case'
3. Risk of non-compliance and costs shortfall
4. Justification to keep or dispose
5. Actions to prepare for preservation or disposal



Thank you, any questions?

Any further thoughts on preparation for sharing openly?

