What infrastructure is necessary for successful research data management (RDM) at universities?

Dr Heila Pienaar
Deputy Director: Strategic Innovation
Department of Library Services
University of Pretoria
CV: https://www.slideshare.net/heila1/dr-heila-pienaar-cv-aug2017/1

Acknowledgement

- On 19-20 March 2018 there will be a Symposium in Gottingen, Germany on the theme: 'The critical role of university RDM infrastructure in transforming data to knowledge'. More information at: http://www.eresearch.uni-goettingen.de/content/190320032018-pre-rda-symposium-göttingen-critical-role-university-rdm-infrastructure
- I thought it could be an interesting topic for the DIRISA track of the CHPC Annual National Meeting ….

CONTENT

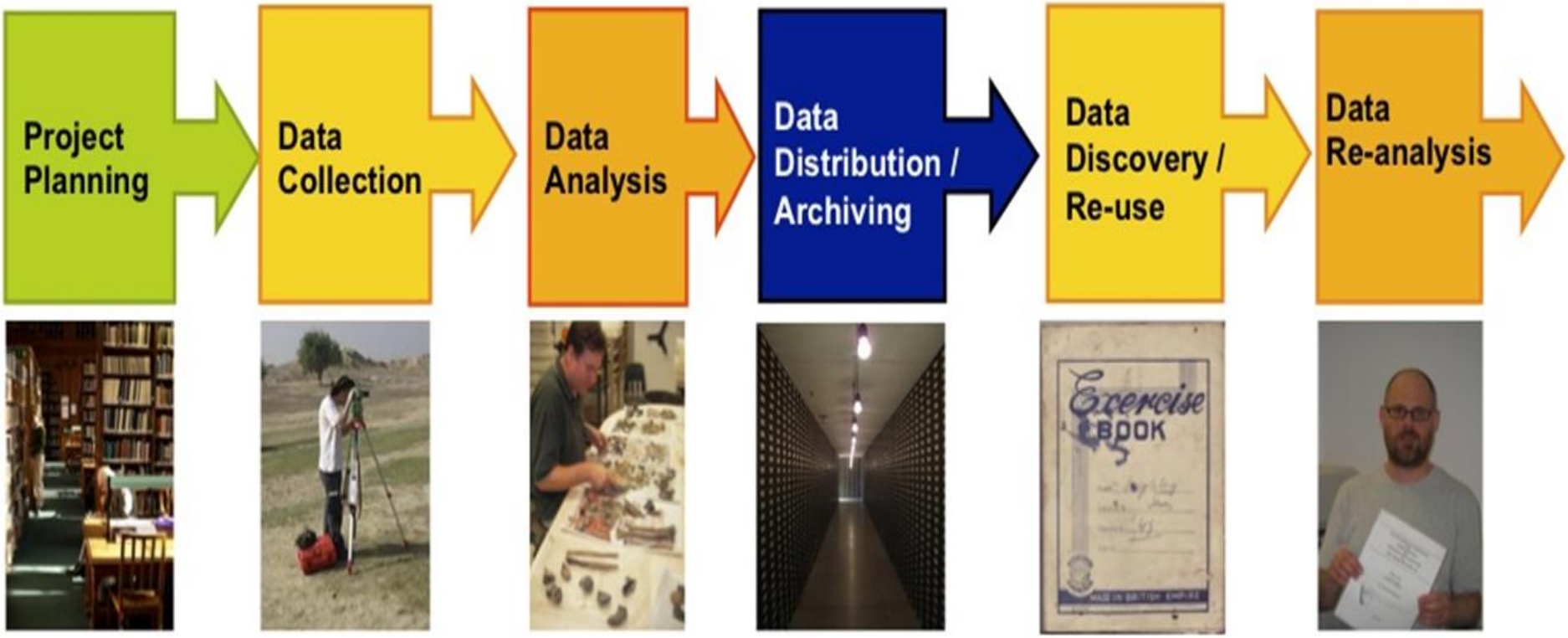
- RDM life cycle
- Research data elements in the research lifecycle
- What is RDM infrastructure
- IT infrastructure
- Library infrastructure
- Research office infrastructure
- Examples of 4 universities RDM service offerings
- Recommendations
- Bibliography

RDM LIFE CYCLE

University of Oxford Research Data Management Chart

http://www.admin.ox.ac.uk/rdm/

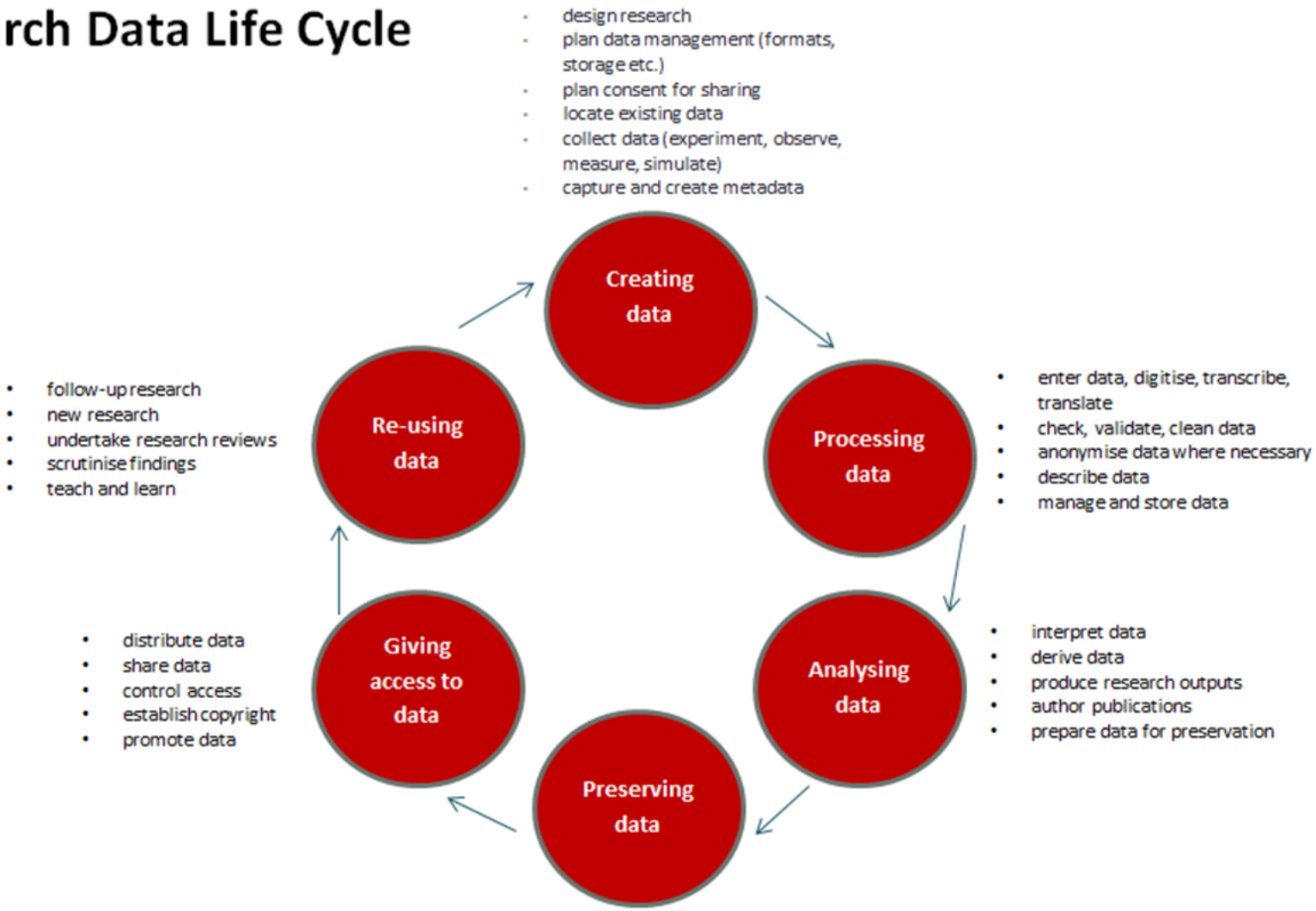
Data train, model 1



Created by Lindsay Lloyd-Smith as part of the JISC-funded Data Train Project based at the Cambridge University Library

UK DATA ARCHIVE

Research Data Life Cycle

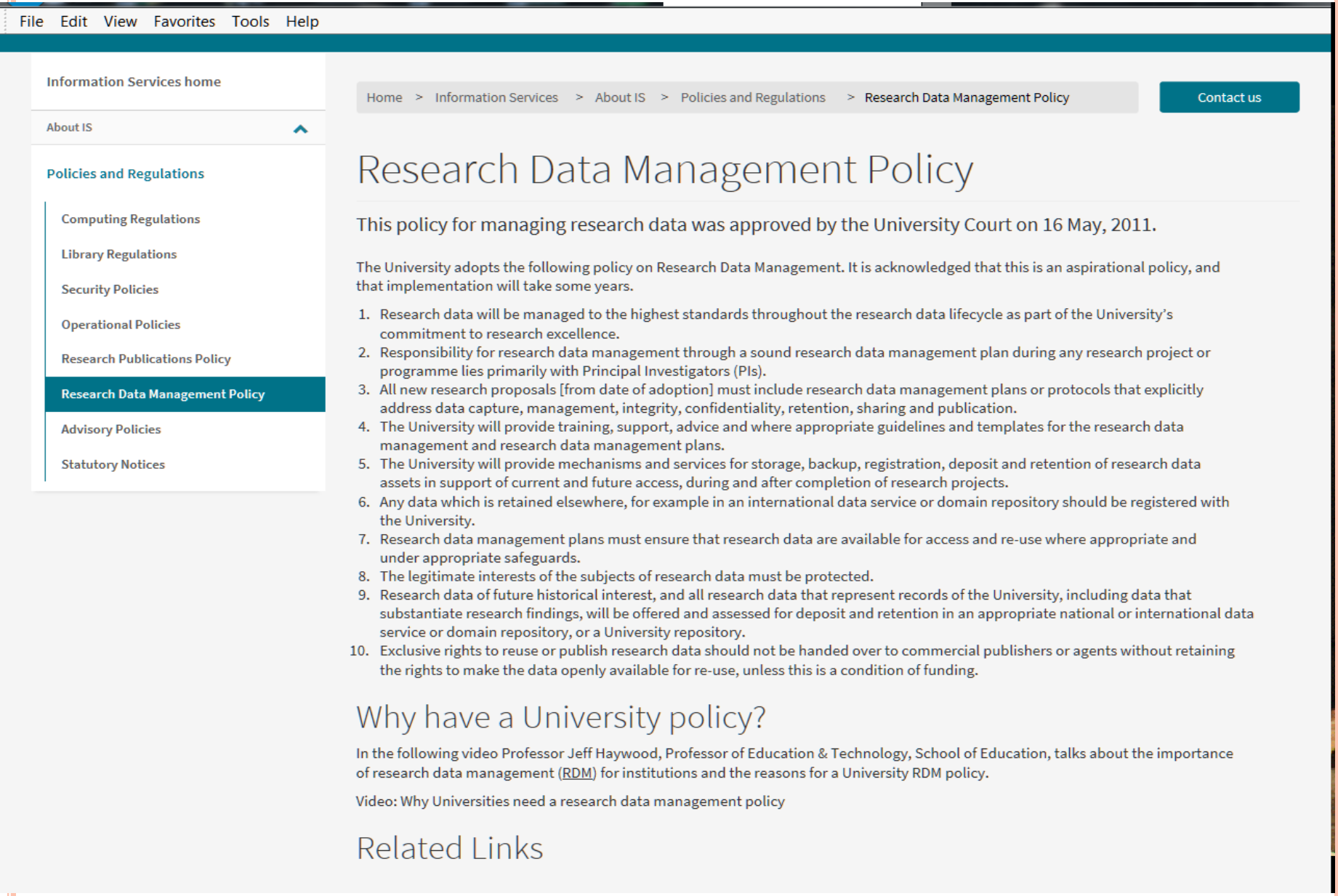


Based on UK Data Archive documentation:
http://www.hsrc.ac.za/Page-156.phtml
http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
http://www.data-archive.ac.uk/

Research data elements in the research lifecycle

- Adherence to RDM policy
- Writing data management plans (DMP's) – according to funding requirements
- Assignment of meta-data (for harvesting, search and retrieval)
- Assignment of persistent identifiers (e.g. DOI's; ORCID)
- Deposit of data in a data repository:
  - Active data
  - Final data
- Data must be safely archived
- Publishing data articles
- Data copyright and data citation
- Lecturing on research data management

RDM POLICY: UNIVERSITY OF EDINBURGH



File Edit View Favorites Tools Help

Information Services home

About IS

Policies and Regulations

- Computing Regulations
- Library Regulations
- Security Policies
- Operational Policies
- Research Publications Policy
- Research Data Management Policy
- Advisory Policies
- Statutory Notices

Home > Information Services > About IS > Policies and Regulations > Research Data Management Policy

Contact us

Research Data Management Policy

This policy for managing research data was approved by the University Court on 16 May, 2011.

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be protected.
9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

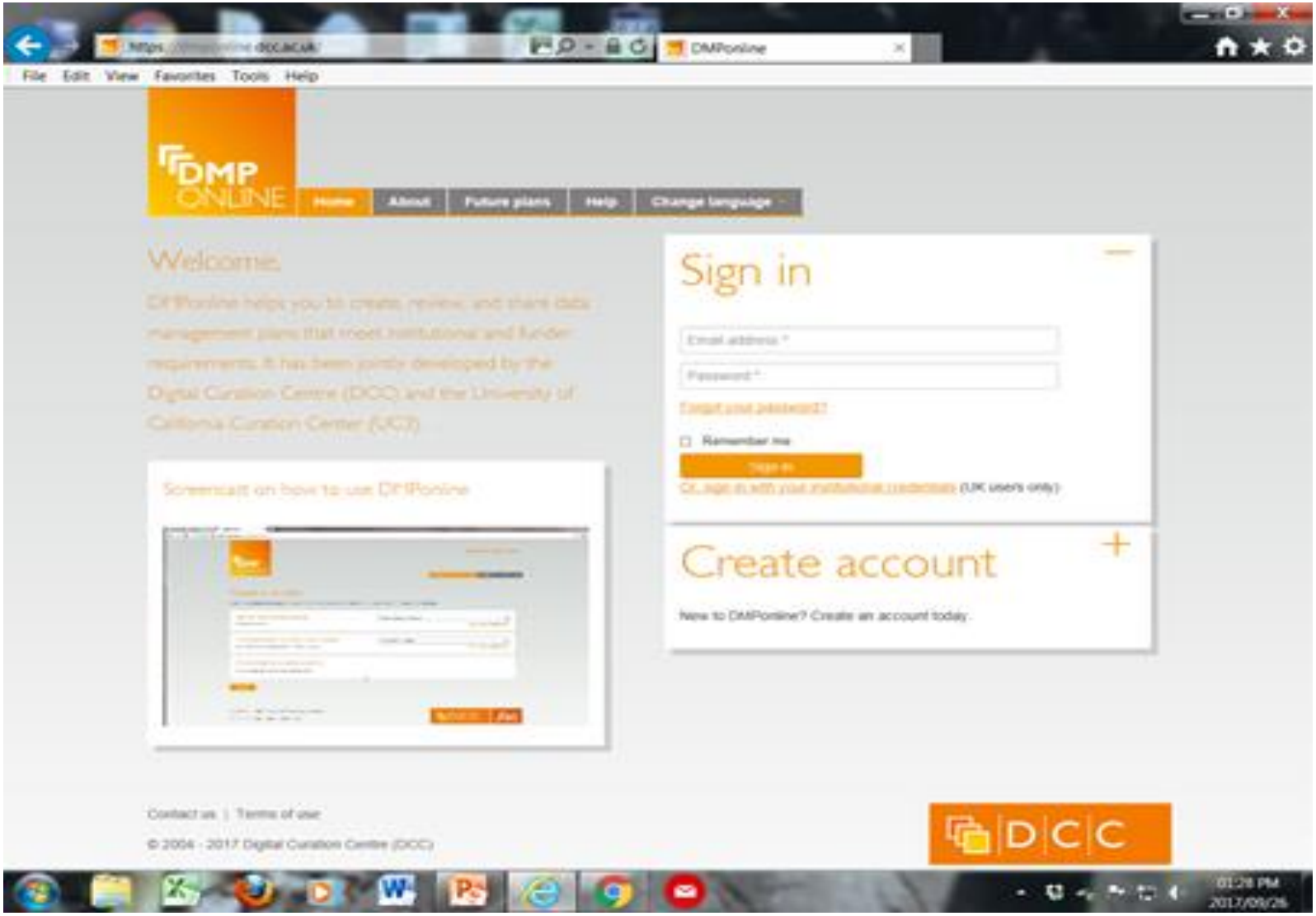
Why have a University policy?

In the following video Professor Jeff Haywood, Professor of Education & Technology, School of Education, talks about the importance of research data management (RDM) for institutions and the reasons for a University RDM policy.

Video: Why Universities need a research data management policy

Related Links

Data management plan: DMP online



DATA DEPOSIT

PURDUE UNIVERSITY **Purdue University Research Repository** PURR

Login | Register | Have a Question?

Home | Publications | Projects | Get Started | Policies | About | Search

You are here: Home › Publications › Datasets › Source code for Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision › About

Source code for Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision

By Haiguang Wen[1], Junxing Shi[1], Yizhen Zhang[1], Kun-Han Lu[1], Jiayue Cao, Zhongming Liu[2]

1. Purdue University 2. Weldon School of Biomedical Engineering, School of Electrical and Computer Engineering, Purdue University

This document includes the main source code (Matlab or Python) related to our study.

Listed in Datasets | publication by group Laboratory of Integrated Brain Imaging

Download Bundle →

Additional materials available

Version **1.0** - published on 25 Sep 2017
doi:10.4231/R7V98675 - cite this

Content may change until committed to the archive on 25 Oct 2017

0 review(s) (Review this)
0 questions (Ask a question)
0 citation(s)
58 total view(s), 5 download(s)

Share:

About | Supporting Docs | Versions | Reviews | Questions | Citations | Usage

Description

Convolutional neural network driven by image recognition has been shown to be able to explain cortical responses to static pictures at ventral-stream areas. Here, we further showed that such CNN could reliably predict and decode functional magnetic resonance imaging data from humans watching natural movies, despite its lack of any mechanism to account for temporal dynamics or feedback processing. Using separate data, encoding and decoding models were developed and evaluated for describing the bi-directional relationships between the CNN and the brain (Wen et al., 2017). Here, we provided the main source codes (Matlab or Python) that are related to this study.

Reference: Wen, H., Shi, J., Zhang, Y., Lu, KH., Cao JY. & Liu, Z. (2017). Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. Cerebral cortex. In press.

Cite this work

Researchers should cite this work as follows:

Wen, H., Shi, J., Zhang, Y., Lu, K., Cao, J., Liu, Z. (2017). **Source code for Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision**. Purdue University Research Repository. doi:10.4231/R7V98675

Tags

Biomedical Engineering | deep learning | FMR1 | LIBI | Matlab | natural vision | neural decoding | Neural encoding | Neuroscience | Python Code | source code

Laboratory of Integrated Brain Imaging

Data deposit: meta-data

ing and Sharing Data | **Metadata** | Data Management @ VU - Case Studies | Policies

What is the Best Metadata for My Data?

You may not like this answer, but there is no one size fits all for metadata. The metadata that would best describe your data depends on the nature of your data. However, due to the fact that specialised systems for description (know as schemas) take time and expertise to implement, there are some generalised schemas that have been developed. These schemas include what are considered essential elements to describe any data.

Some disciplines have their own metadata schema. Each will have their own specified elements and structure:

Discipline	Metadata standard
General	Dublin Core (DC) Metadata Object Description Schema (MODS) Metadata Encoding and Transmission Standard (METS)
Arts	Categories for the Description of Works of Art (CDWA) Visual Resources Association (VRA Core)
Astronomy	Astronomy Visualization Metadata (AVM)
Biology	Darwin Core
Ecology	Ecological Metadata Language (EML)
Geographic	Content Standard for Digital Geospatial Metadata (CSDGM)
Social sciences	Data Documentation Initiative (DDI)

http://libraryguides.vu.edu.au/researchdatamanagement/metadataschemasandexamples

Data citation

- Please cite as: Ball, A. & Duke, M. (2015). 'How to Cite Datasets and Link to Publications'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides
- **Contents**
  - Why cite datasets and link them to publications?
  - Principles of data citation
  - Data citation for authors
    - Ways of referencing data
    - Elements of a data citation
    - Digital Object Identifiers
    - Contributor identifiers
    - Granularity
    - Citing unreleased data
    - Citing physical data
  - Data citation for repositories
    - Tools and services
    - Example implementations
    - Manual and automatic use of citations
    - Granularity
    - Versioning and dynamic data
  - Notes
  - Further information
  - Acknowledgements

Data journals

- Data journals are publications whose primary purpose is to expose datasets. They enable the author to focus on the data itself, rather than producing an extensive analysis of the data which occurs in the traditional journal model. Fundamentally, data journals seek to:
  - promote scientific accreditation and re-use
  - improve transparency of scientific method and results
  - support good data management practices and
  - provide an accessible, permanent and resolvable route to the dataset
- **Examples of data journals**
  - Geoscience Data Journal - published by Wiley and established in 2012
  - Scientific Data - published by Nature and established in 2013
  - Journal of Open Archaeology Data - published by Ubiquity and established in 2011
  - Biodiversity Data Journal - published by Pensoft and established in 2013.
  - Earth System Science Data - published by Copernicus Publications and established in 2009

https://www.ands.org.au/working-with-data/publishing-and-reusing-data/data-journals

What is RDM infrastructure?

- Can include:
  - Regulatory environment: laws, policies, standards
  - Expertise in RDM
  - RDM training
  - Academic subject expertise
  - IT infrastructure: repository software; HPC; networks etc.

Research Infrastructure refers to **facilities**, **resources** and related **services** used by researchers to conduct research. As the complexities of the research environment grow, so too does the need for supporting infrastructures to enable researchers to operate effectively. https://www.jisc.ac.uk/sites/default/files/international-advances-in-digital-scholarship-report.pdf

IT INFRASTRUCTURE

- Input into institutional RDM policy and guidelines
- Input in the evaluation and implementation of a research data repository
- Make repositories / platforms available for final and active data
- Make adequate IT infrastructure available for the storage and long-term preservation of data
- Make tools and capacity available for the analysis of data

Library infrastructure

- Input into institutional RDM policy and guidelines
- Training in DMP tool
- Input in the evaluation and implementation of a research data repository
- Provide meta data guidelines
- Training to make data ready for deposit e.g. assignment of relevant subject meta-data
- Training on the deposit of data in a research data repository
- Provide advice on:
  - Data copyright
  - Data citation
  - Publishing data articles
- Data carpentry workshops on 'R', SQL, OpenRefine https://github.com/datacarpentry ; http://www.datacarpentry.org/

Research office infrastructure

- Input into institutional RDM policy and guidelines
- Keep researchers up to date with relevant research data funding requirements e.g. open data publishing
- Obtain business information e.g. statistics on the publishing and impact of research data
- **NB!!!** A RIMS (Research Information Management System) or CRIS (Current Research Information System) must not be confused with a RDM (Research Data Management) system

A different perspective: scoping the university service bundle

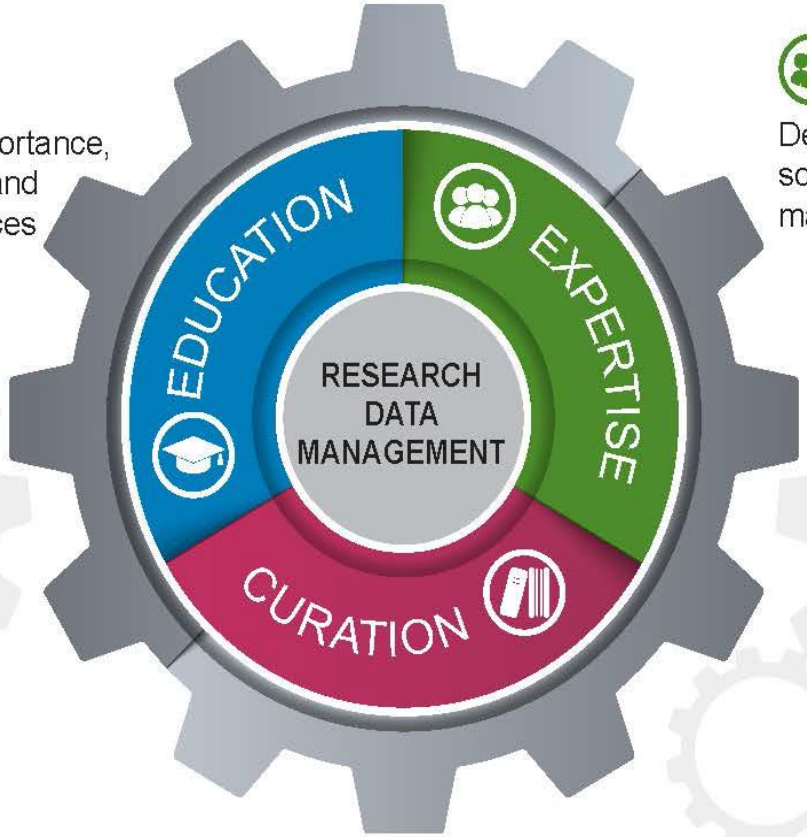
- Case studies of four top international research-intensive universities:
  - University of Edinburgh (UK)
  - University of Illinois at Urbana-Campaign (USA)
  - Monash University (Australia)
  - Wageningen University and Research (The Netherlands)

  *RDM is not a monolithic set of services duplicated across universities; it is a **customized solution** shaped by a range of internal and external factors operating on local decision-making.*

Research Data Management Service Categories

EDUCATION

Raise awareness of RDM's importance, encourage RDM skill-building, and disclose RDM tools and resources



EXPERTISE

Decision support for, and customized solutions to, specific research data management problems

CURATION

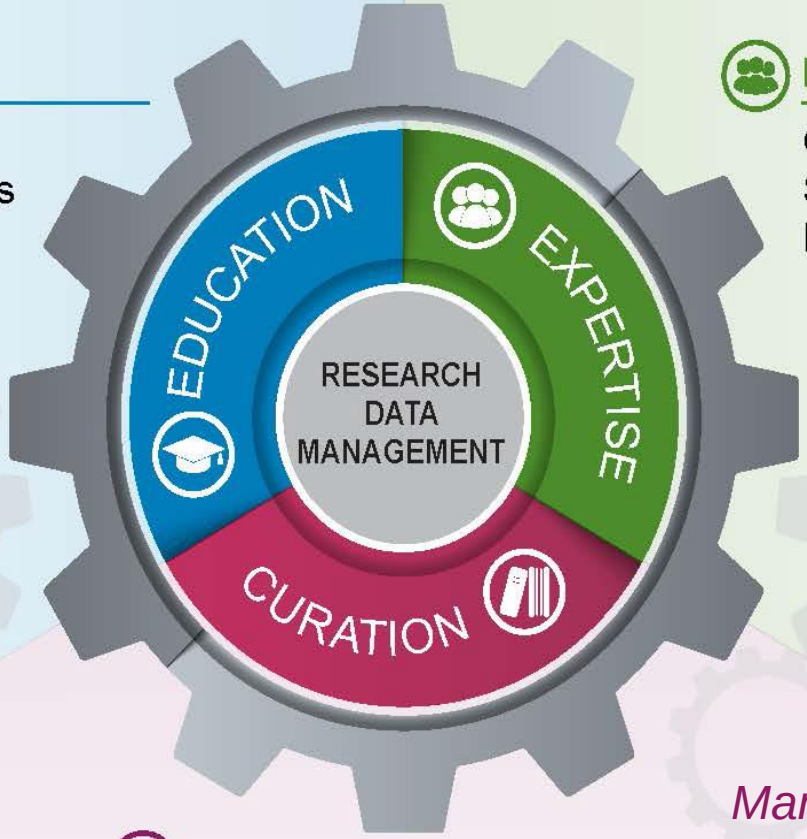
Technical infrastructure and related services that support data management throughout the research cycle

RDM Service Bundle Highlights
University of Edinburgh

EDUCATION

MANTRA
Courses & Workshops
RDM MOOC
Customized Training

MANTRA and the RDM MOOC are community resources, rather than targeted exclusively to Edinburgh's affiliated researchers



EXPERTISE

General Email Contact Point
Staff Contact Information
Digital Curation Centre

National centers of RDM capacity are important supplements to the local RDM service bundle

CURATION

DataStore
DataSync
DataShare
DataVault
Pure

Many research institutions choose to offer RDM Curation services in parallel with, rather than subsumed in, the institutional repository.

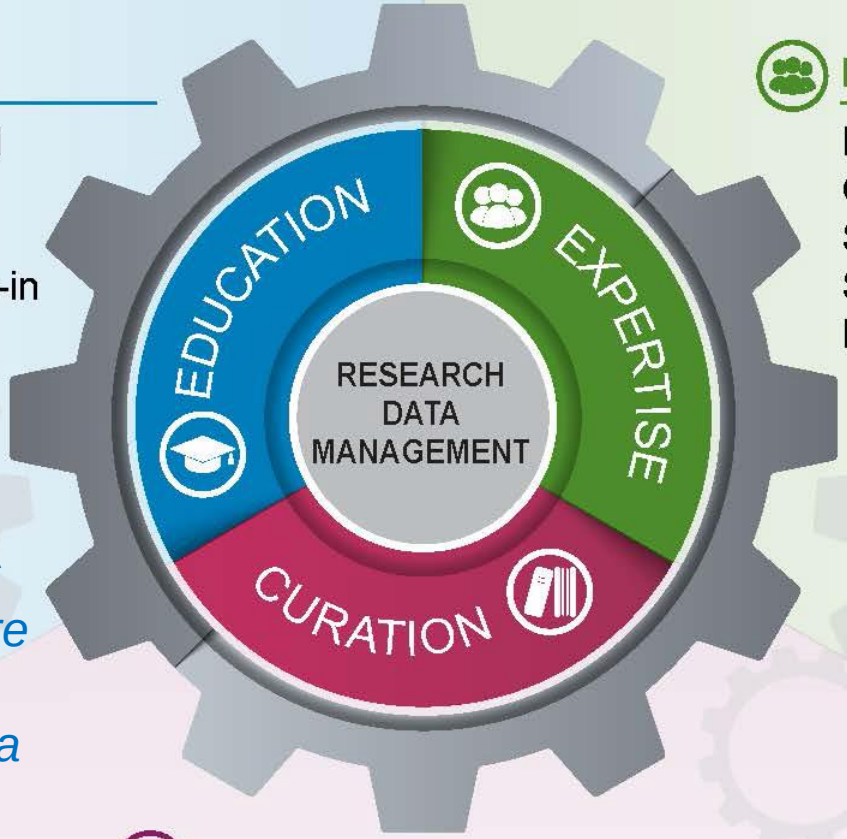
RDM Service Bundle Highlights
University of Illinois at Urbana-Champaign

EDUCATION

Customized DMPTool
RDM Workshops
Outreach
Data Help Desk Drop-in
Data Nudge

EXPERTISE

DMP Review/Feedback
General Email Contact Point
Staff Contact Information
Scholarly Commons
Digital Curation Network



*Use of the Data Bank is optional: researchers are advised that "[i]f your **discipline** already has a trusted **repository**, we recommend you deposit where your community knows to look"*

CURATION

Illinois Data Bank
U of I Box
Active Data Management Support

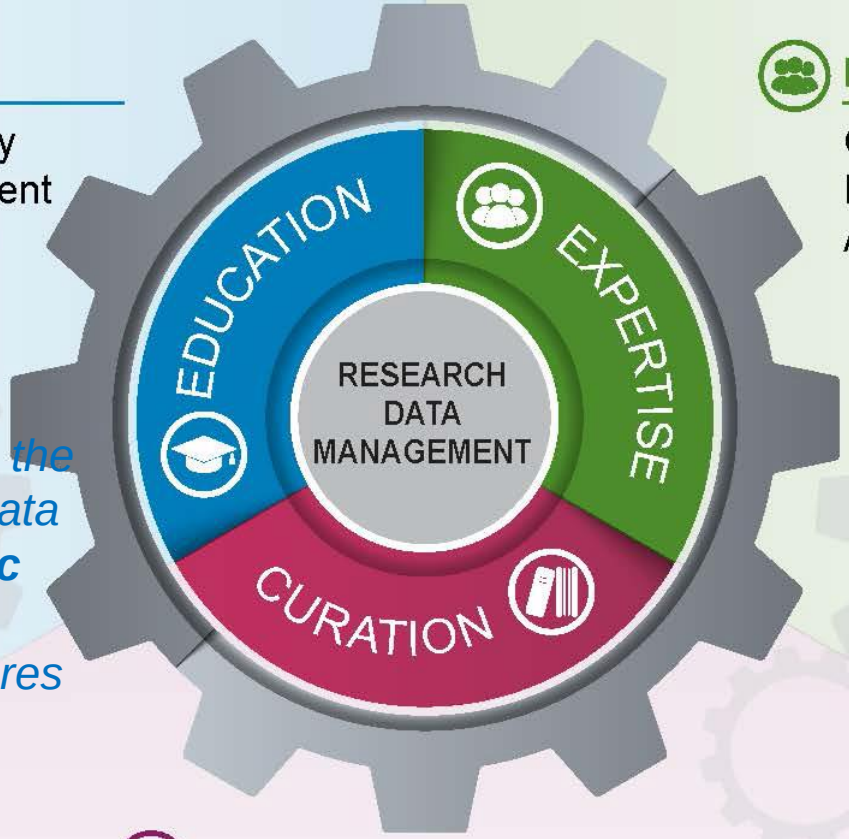
RDM Service Bundle Highlights
Monash University

EDUCATION

Online Resource Library
Professional Development Activities for Students & Researchers

EXPERTISE

General Email Contact Point
Referrals to Campus Units
Australian National Data Service



*Store.Monash supports the capture and storage of data generated from **scientific instrumentation**. The OzFlux Repository captures and stores **atmospheric data**. LabArchives is for managing and storing **active** research data online, as well as Monash Google Drive for smallscale **active** data storage.*

CURATION

Monash.Figshare
VicNode
Store.Monash

Research Data Australia
DOI/Handle service

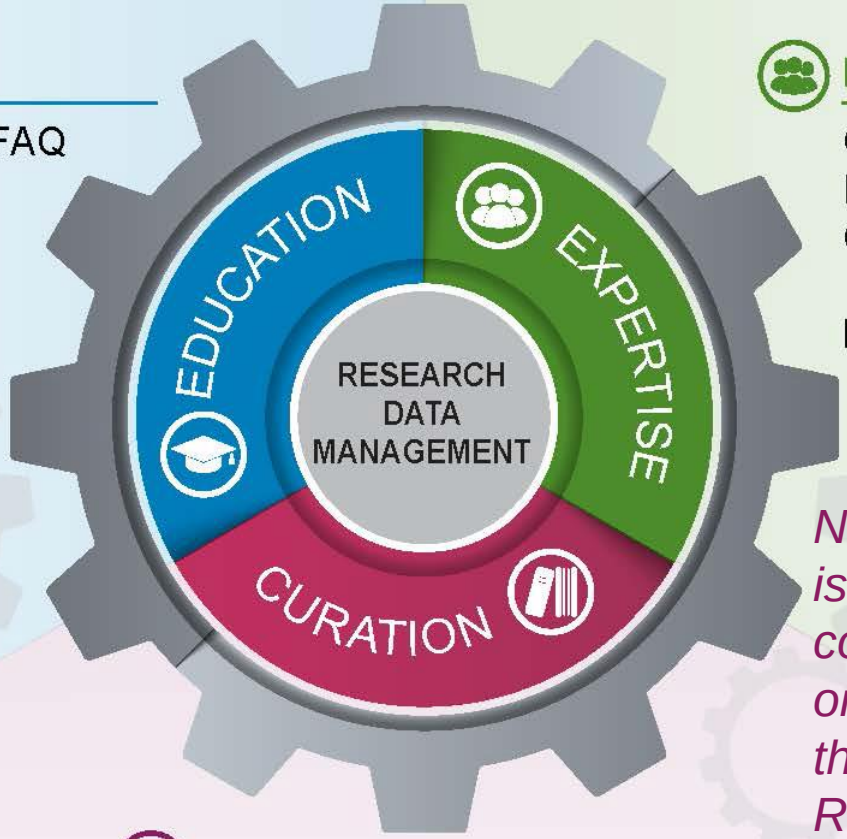
RDM Service Bundle Highlights
Wageningen University & Research

EDUCATION

"Browse by Subject"/FAQ
Online Resources
Training Courses
News Feed

EXPERTISE

Online Contact Form
DMP Review/Feedback
Consultation for Data and Metadata Preparation
Mediated Deposit



The Data Management Support group is composed of experts from the Library, IT Services, Document Management and Logistics, and Corporate Governance and Legal Services departments.

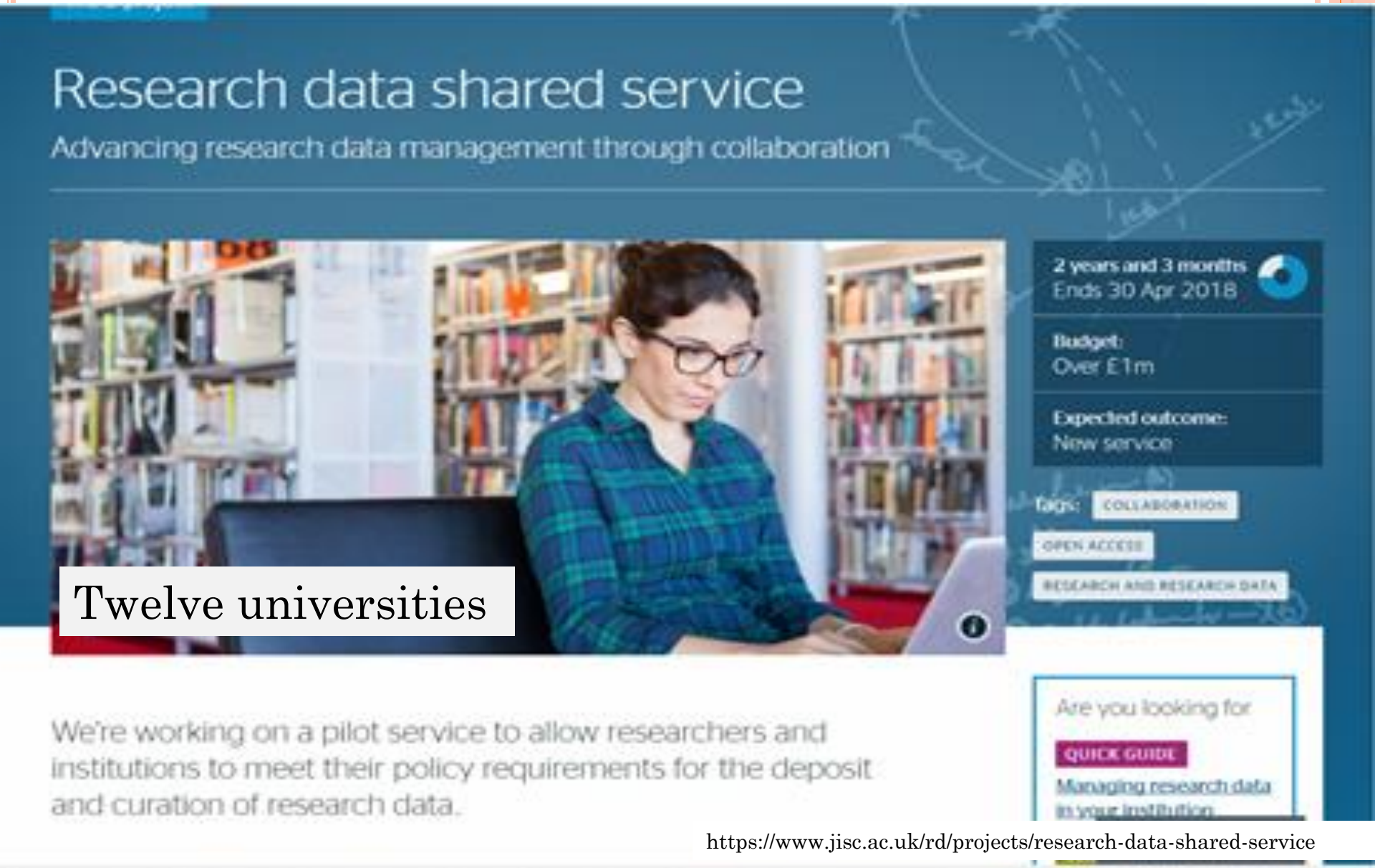
No RDM service bundle is an island — all are connected, to a greater or lesser degree, to the broader, external RDM service ecosystem.

CURATION

DANS-EASY
4TU.ResearchData
Git@WUR
Pure/NARCIS

Source code is managed at a GitLab implementation

Last but not least: research data shared service project (UK JISC)



https://www.jisc.ac.uk/rd/projects/research-data-shared-service

CONCLUSIONS

- In general, the **Library** is mainly responsible for RDM
- **Close working relationships** amongst university role players is essential: University management; Faculty members; Library; IT; Research office etc.
- Must design and deliver an **adequate and seamless service portfolio / offering** (including internal & external services) covering the research data lifecycle
- Close working relationship between research institutions in **South Africa** is also critical – NeDICC (Network of Data and Information Curation Communities) is a good example of a national Community of Practice on RDM
- Close working relationship with **national** initiatives e.g. DIRISA is very important
- **International** relationships and co-operation can be leveraged to leapfrog RDM service offerings

BIBLIOGRAPHY

- Bryant, Rebecca, Brian Lavoie and Constance Malpas. 2017. *Scoping the University RDM Service Bundle*. The Realities of Research Data Management, Part 2. Dublin, OH: OCLC Research. doi:10.25333/C3Z039.