

# Developing Research Data Management Services and Repository Certification

Dr. Tomasz Miksa

SBA Research & TU Wien

[tmiksa@sba-research.org](mailto:tmiksa@sba-research.org)

# Agenda

- Research Lifecycle models
- Developing Research Data Management (RDM) Services
  - Policies
  - Costs
  - Data Management
  - Infrastructure
- Repository Certification

# Introduction



- Perspective change
  - 'DMP lecture'
    - Researcher / data scientist perspective
      - How should I manage my data?
      - How to comply with policies?
  - This lecture
    - You need to design a solution at an institution supporting data management and preservation
      - Taking into account technical, organisational, cultural, political problems...
    - **What do you do?!**

# The size and governance structure of institutions has an impact

## Large, hierarchical institutions

- Move slowly
- Require a lot of advocacy
- Have more resource
- Economies of scale



## Smaller institutions

- More agile
- Simpler communication
- More focussed vision
- Less resource



# Establish a long term governance group

- Good mix of representatives from operational units
- Senior management leadership
- Avoid relying on one or two key individuals
- Keep the group active



Image credit: <http://www.executive-coaching-services.co.uk/executive-coaching/leaders.jpg>

# Be aware of existing infrastructure

In many  
HEIs  
services are  
beginning to  
be  
embedded  
but aren't  
joined up  
effectively

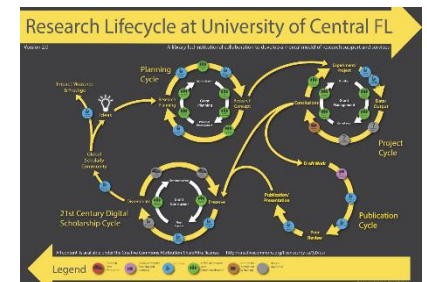
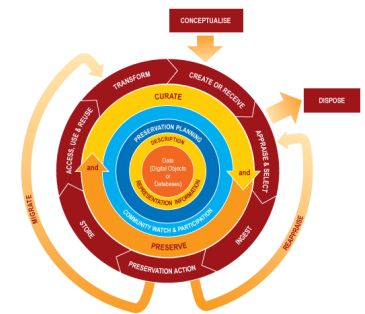
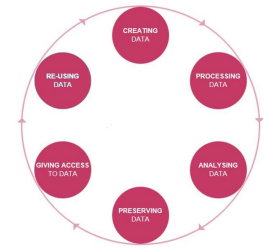
Research Organisation Description		
RDM profile component	Record HEI Link - insert your URLs in the space provided	Guidance
Means of raising staff awareness of funders' research data requirements		Provide a link to an information page on funders' policies. This could be internal or external (E.g., DCC's policy overview table)
Research data policy		Provide a link to research data policy or aspirational statement
Strategy or implementation plan for research data services		Provide a link to research data strategy page or roadmap
RDM advice and support services		Provide a link to page describing data management planning guidance and/or support services at this organisation
Active data storage		Provide link(s) to active research data storage information page(s). There may be multiple options at Research Group/School/College/Central levels.
Data register or catalogue		Provide a link to your internal research data registration homepage. This may be provided via the data repository and/or CRIS.
Persistent identification for datasets		Provide a link to any page(s) detailing schemes used to identify digital data items (e.g., DataCite).
Data access procedures		Provide a link to any information provided about research data access.
Secure data access		Provide a link to any information provided about secure data access and governance.
Institutional publications repository (if it includes research data or metadata)		Provide a link to your institutional repository homepage
Data repository for longer term access and preservation		Provide a link to your research data repository homepage. This may be an extension of your publications repository, a separate data repository or a pointer to an external data repository service (E.g., Zenodo).

<http://www.dcc.ac.uk/projects/opd-for-rdm>

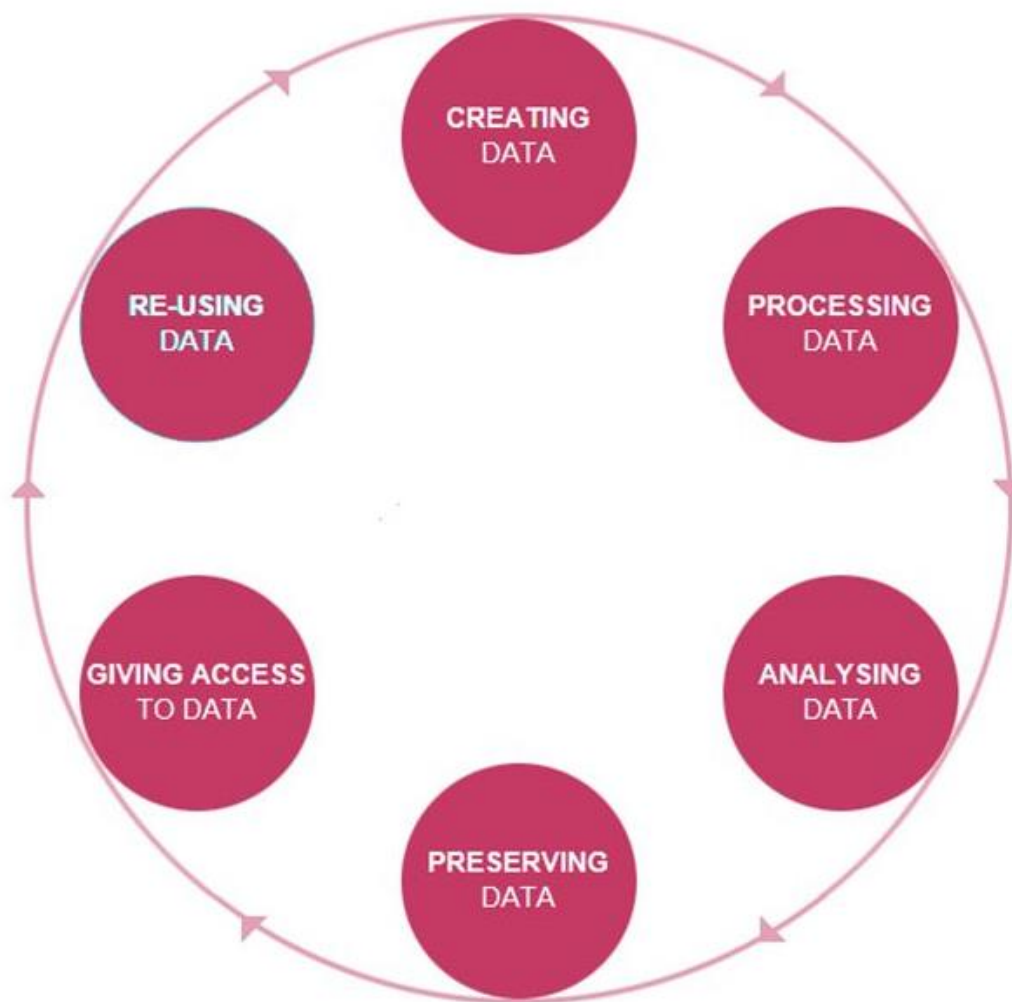
# Understand processes

## ■ Research Data Lifecycle Models

- Describe roles of stakeholders
- Help in
  - tailoring services
  - identifying responsibilities
  - defining infrastructure
- NOT to be used by researchers
- Examples
  - UK Data Archive
  - Digital Curation Centre
  - University of Central Florida

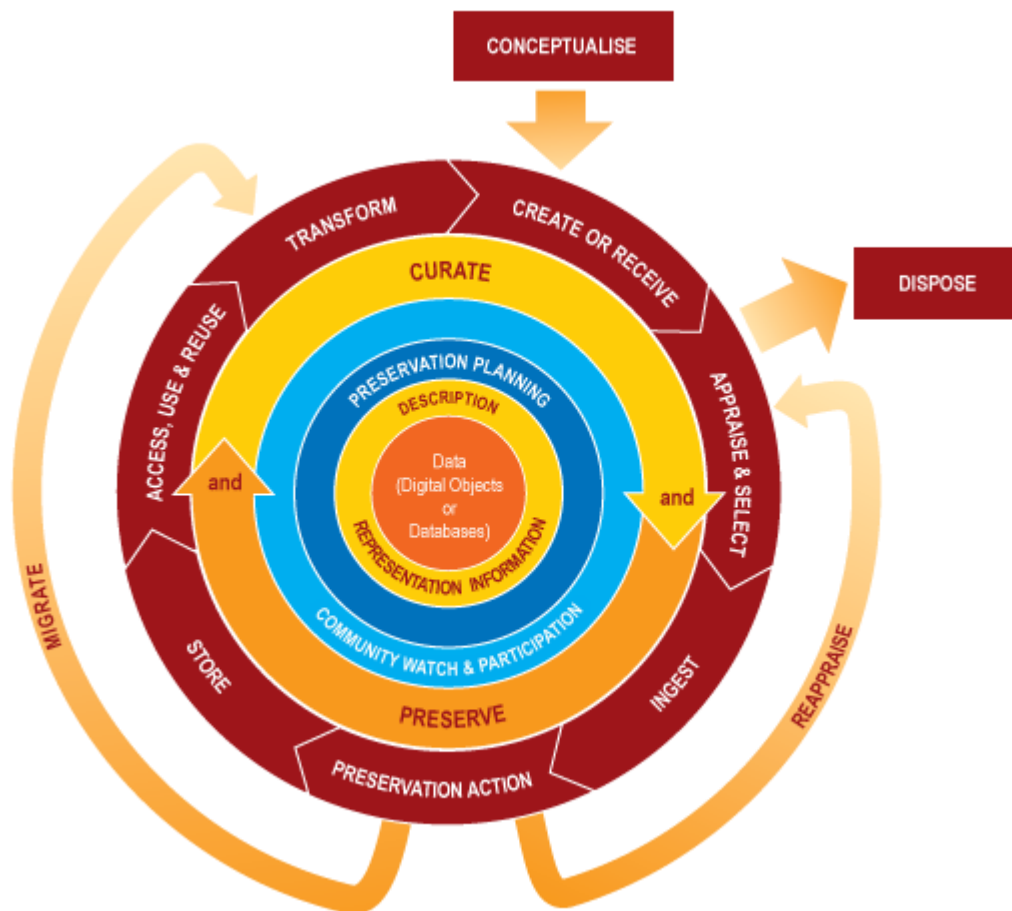


# UK Data Archive Lifecycle model



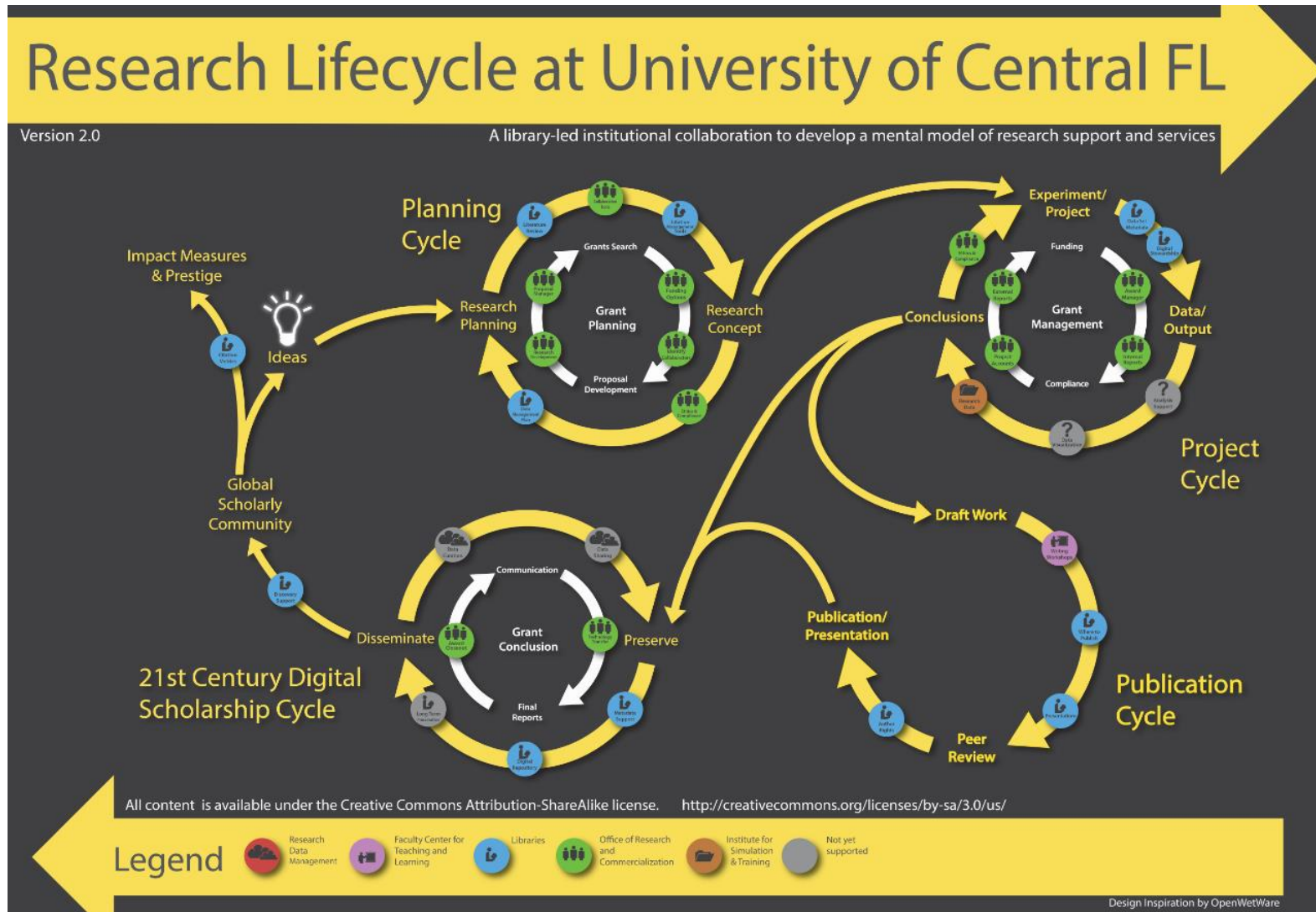


# Digital Curation Centre Lifecycle Model



<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

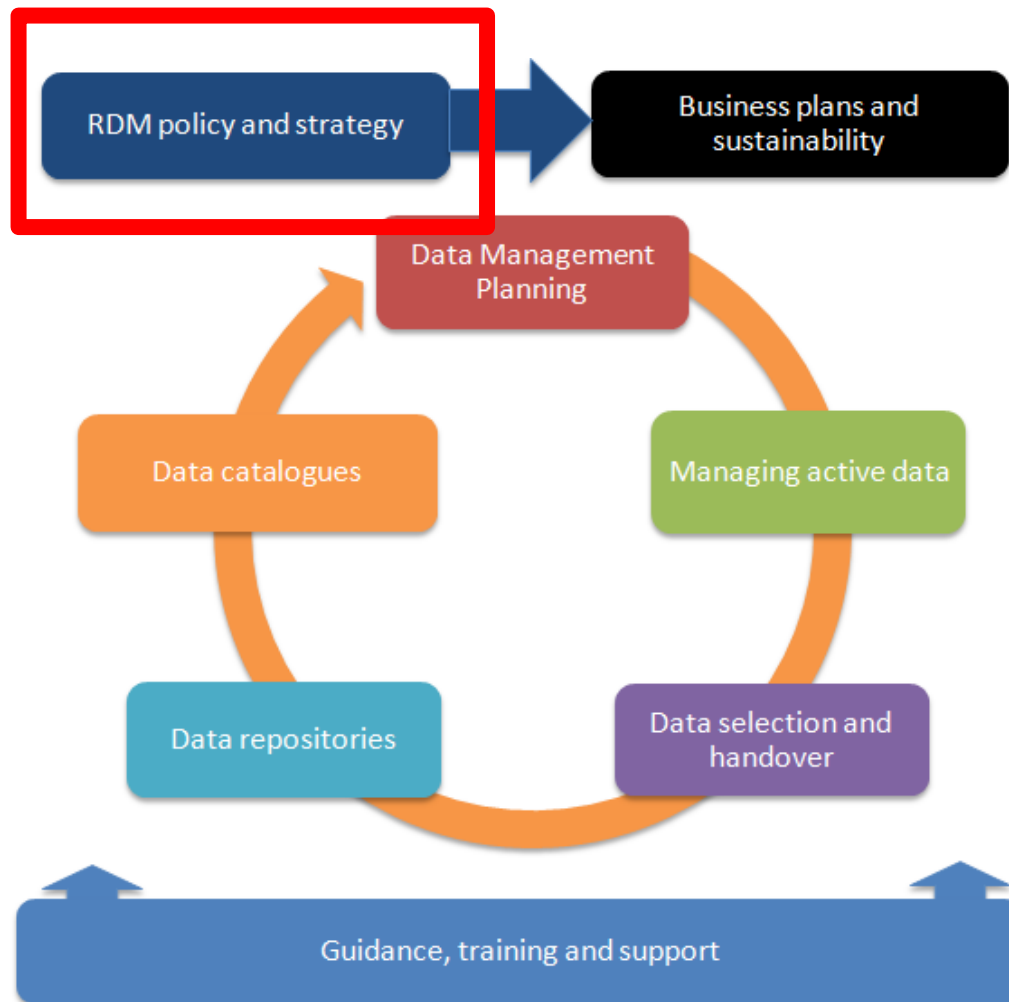
# University of Central Florida Lifecycle Model



# Developing Research Data Management (RDM) Services

- Components of an RDM service infrastructure
  - RDM policy and strategy
  - Business plans and sustainability
  - Guidance, training and support
  - Data management planning
  - Managing active data
  - Data selection and handover
  - Data repositories
  - Data catalogues

# Developing Research Data Management (RDM) Services



# Understanding policies

1

Starting with some *Taboos*

2

Going over to related *Principles*

3

Going over to the creation of a *Policy*

4

Going over to *Rules, Legislations* and *Regulations* (canons, norms, guidelines)

# Taboo

A **taboo** is something, which is **forbidden** or **disapproved of**, or placed under a social prohibition.

*“Thou shalt not delete scientific data“*

*“Thou shalt not destroy infrastructures“*

Usually a **negative** assertion.

In society and academic environment taboos are accepted only if they are just **a few**.

# Principle

- A principle is a fundamental truth or proposition that serves as the foundation for a system of belief

*Research **data** are to be preserved*

*Research **data** are to be kept **FAIR** - Findable, Accessible, Interoperable, Reusable.*

*Research data **infrastructures** are to be kept accessible*

- Format: positive assertion

A policy is...

- a **course of action** or **principles** adopted by an organization

*“The Institution [name XY] will preserve its research data infrastructure always accessible and free to its members according to the FAIR principles”*

General assumptions concerning policies:

- A single Policy
  - the policy is a single entity, it should not be in competition with other policies
- Creators of Policy do not want to modify it
- Policy is usually accepted after a while
- Policy offers the frame for the generation of Rules



# Rules, Regulations

- Rules are prescribing conducts or actions
- They are generated by the founder of “orders”
- Characteristics of rules are:
  - There may exist “**lots of rules**”
  - Rules are not always **clear**  
(they often need interpretation according to the situation)
  - Rules are **usually accepted**, but **often imposed** procedures
  - It is allowed to **modify Rules**
  - The Law is an expression of rules

Example:

*“Our University will maintain accessible our infrastructure each day from 9:00 a.m. to 12:00 a.m and offer support only on Friday from 7:00 a.m. to 8:00 a.m. The research data, that are publicly funded are to be kept free and accessible to all members of our University each Sunday, from 9:00 to 12:00 a.m.”*

Taboos	Principles	Policies	Rules
<p>Negative assertion</p> <p>few</p> <p><i>“You shall not delete scientific data”</i></p> <p><i>“Youl shall not destroy infrastrcutures”</i></p>	<p>Positive assertion</p> <p>more than „few“</p> <p><i>“Research data are to be kept FAIR - Findable, Accessible, Interoperable, Reusable.”</i></p> <p><i>“Research data infrastructures are to be kept accessible”</i></p>	<p>A course or principle of action. Policy offers the frame for the generation of Rules, should not be in competition with other policies</p> <p><i>“The Institution [name XY] will preserve its research data infrastructure always accessible and free to its members according to the FAIR principles”</i></p>	<p>Rules prescribe conducts or actions; define who what when and where should be done according to the Policy</p> <p><i>“Our University will maintain accessible our infrastructure each day from 9:00 a.m. to 12:00 a.m and offer support only on Friday from 7:00 a.m. to 8:00 a.m. “</i></p>

## Why these differentiations?

- It is important to identify the different semantic levels
- Understanding of the semantic hierarchy is useful in order to produce appropriate guidelines

# Policy – LEARN project

- Leader Activating Research Networks
- **Research Data Management Policy**
  - can be tailored by any University or Research Institution
  - based on existing European policies



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

# Outreach example: Austria

- Adaptation to needs of
  - five Austrian art universities (started)
  - three Medical Universities
  - TU Wien



e-infrastructures  
austria

## Policy sections

1. Preamble
2. Jurisdiction
3. Intellectual Property Rights
4. Handling Research Data
5. Responsibilities, Rights, Duties
  - 5.1 Researchers are responsible for:...
  - 5.2 The [name of research institution] is responsible for:
6. Validity



<http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf?pdf=RDMToolkit>

# Policy section examples

- **5.1 Researchers are responsible for:**

- Collecting, documenting, archiving, providing access to and storing or ensuring the proper destruction of research data and records. (...) Such information should be included in a Data Management Plan (DMP) that explicitly defines the approach to matters of data collection, administration, integrity, confidentiality, storage, use and publication.

- **5.2 The Institution is responsible for:**

- Designing and deploying mechanisms and services for the storage, safekeeping, registration and deposition of research data to support current and future access to research data during and after the completion of research activities;

- **Tricky points:**

- How do you define research data? (publications, source code, raw data..)
- Who are researchers? (students, employed formally, etc.)



# Homework – read TUW policy

Research > RTI support > Research data > Research data management > Policy

## Policy for Research Data Management (RDM) at the TU Wien

The TU Wien sees itself as playing an important role in the expansion of knowledge and technology transfer of research results and thus encouraging innovation and ultimately benefitting economy and society. A key to supporting academic research activities lies in the institution's ability to systematically manage, preserve and make available scientific output from different disciplines for reuse. In its Policy for Research Data Management, the TU Wien affirms the value of research data for research and teaching and the potential of their reuse by society.

↓ Policy\_for\_Research\_Data\_Management.pdf PDF 548 KB

### 1. PREAMBLE

### 2. SCOPE

### 3. RIGHTS OF USE

### 4. RESEARCH DATA AND RESEARCH DATA MANAGEMENT

### 5. HANDLING RESEARCH DATA

### 6. RESPONSIBILITIES, RIGHTS AND DUTIES

### 7. VALIDITY

### Definitions

### Recommendations

# RDM policy and strategy

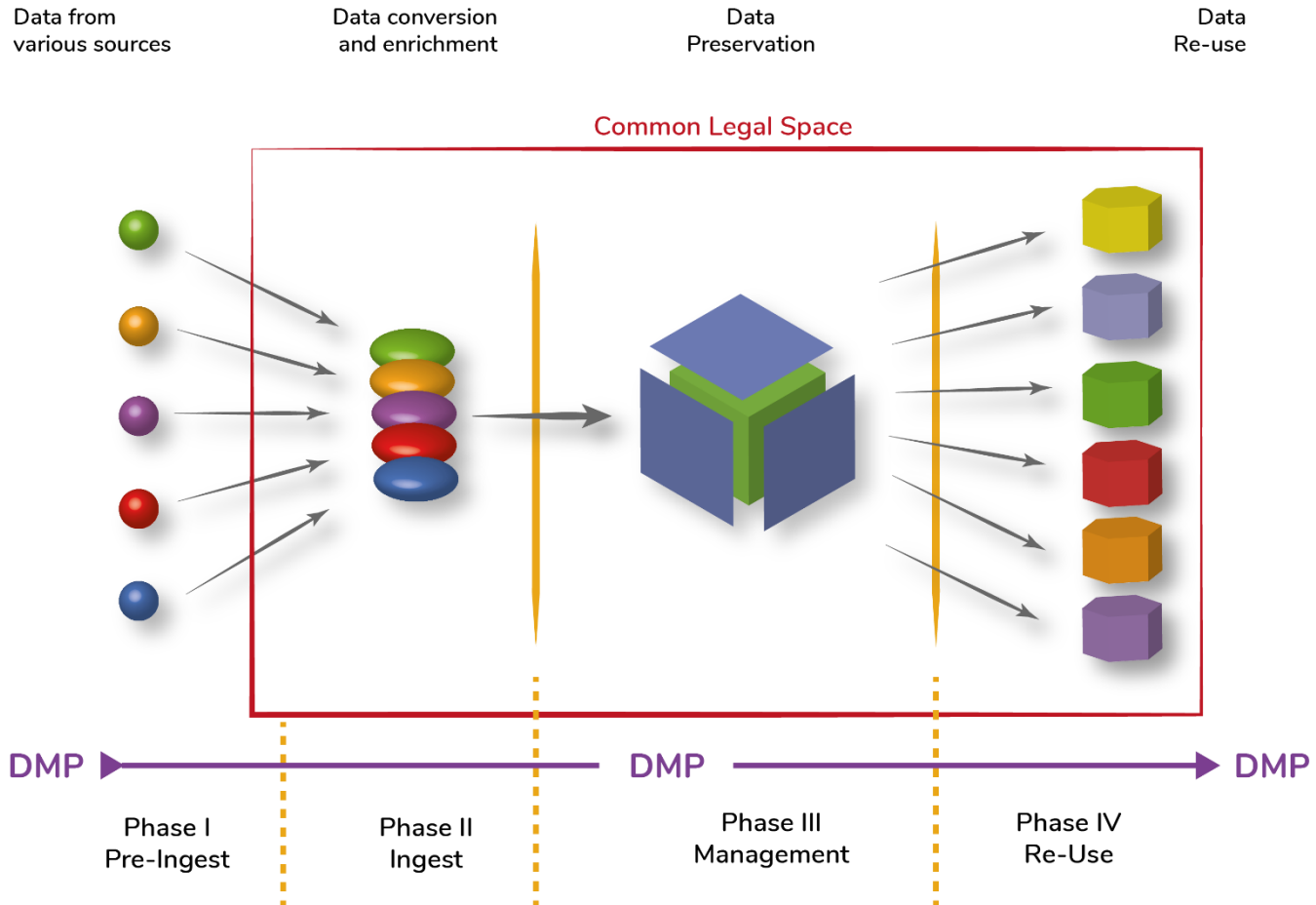
- Develop a strategy
  - Understand your current position and where you want to be to define your strategy
  - Map out a programme of activity to deliver infrastructure and services
- Develop a policy
  - Draft a policy based on external drivers and local context to establish your core RDM principles
  - Ratify the policy then undertake advocacy work and pilot studies to aid implementation
  - Consult broadly to gain consensus and secure support





# BUSINESS PLANS AND SUSTAINABILITY

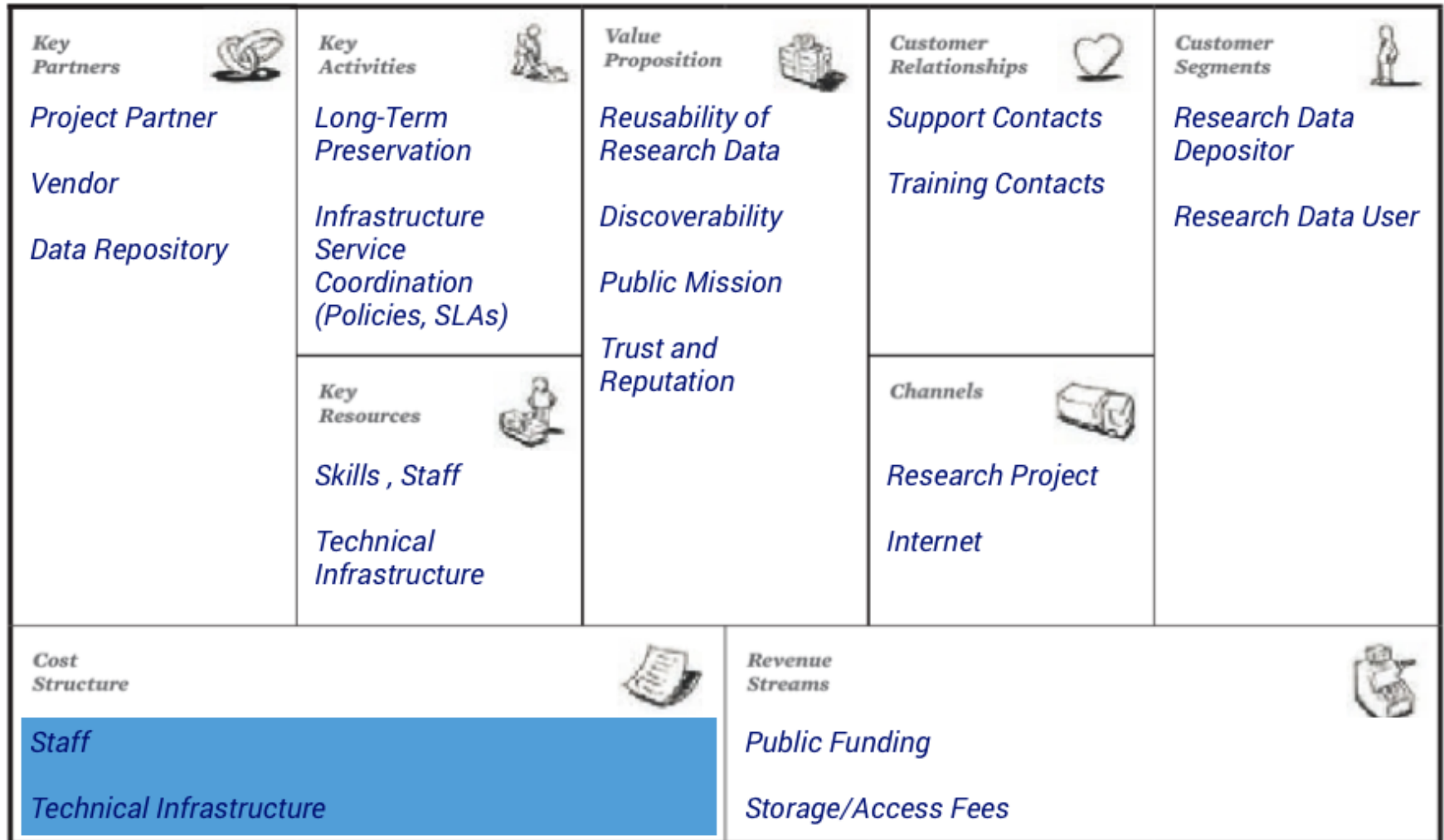
# Who pays for what?



Digital workflow model for data management  
Raman Ganguly, University of Vienna 2017

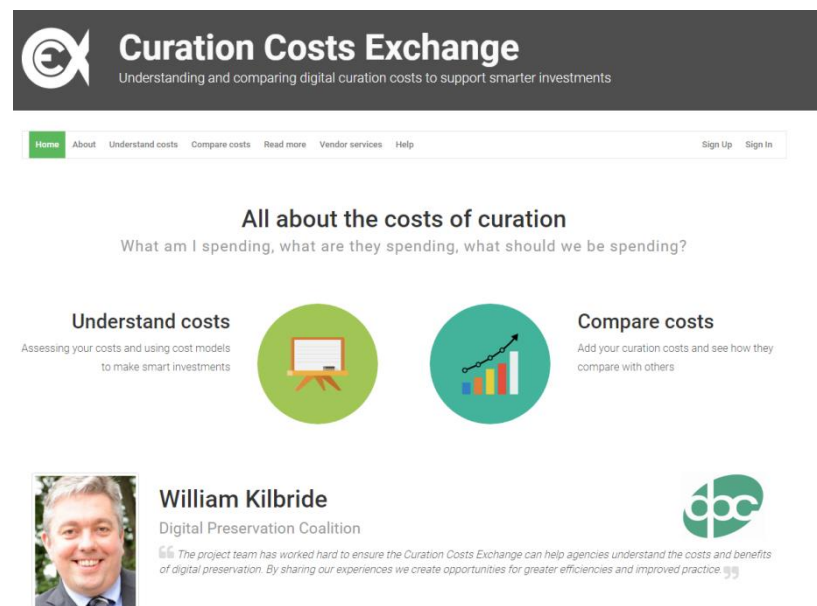
# Identify the cost structure

## Sample Research Data Organisation - Business Model Canvas



# Understanding costs of curation

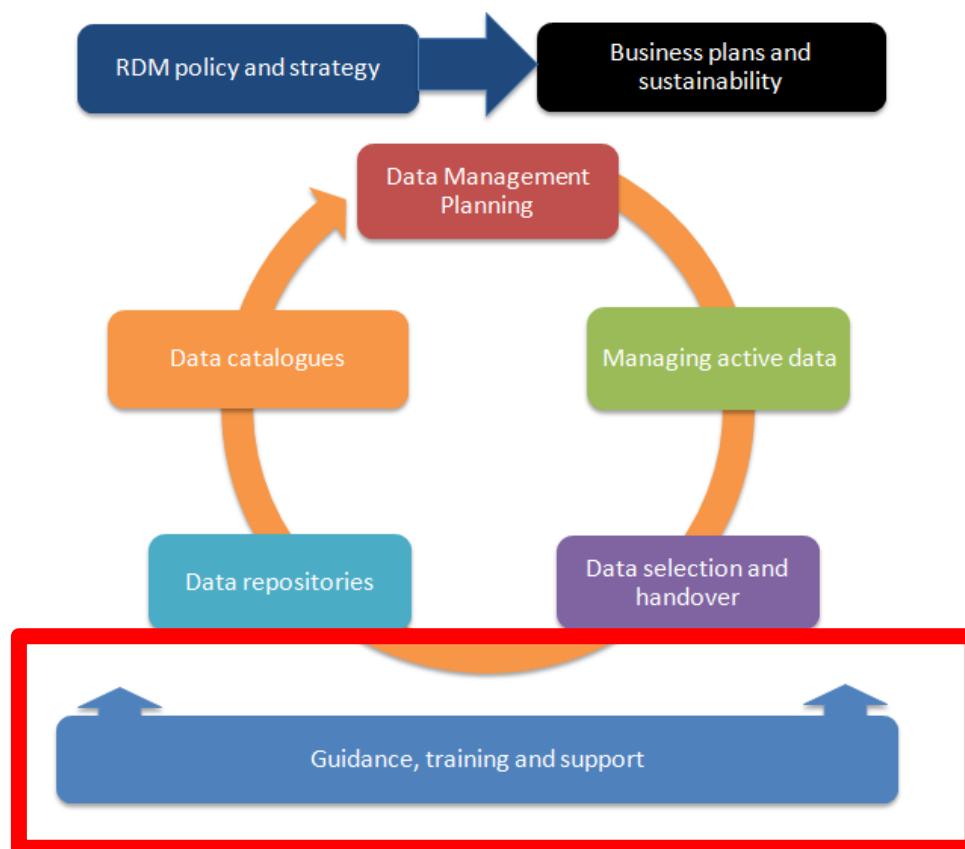
- Curation Cost Exchange (CCEX)
  - platform to understand and compare costs
- 4C analysed existing cost models
  - 10 evaluated
  - 78 assessment criteria
  - you decide which is best for you
- Described cost drivers
  - direct
    - personnel
    - infrastructure
  - indirect
    - quality and trustworthiness
    - reputation and risk



<http://www.curationexchange.org/>

# Business plans and sustainability

- Based on your strategy, develop a phased **business plan** covering projections for 3, 5 and 10 years
- Identify predicted **costs** and planned expenditures, together with an indication of the financial year in which expenditure will occur
- Consider whether any costs can be recouped, for example by **charging for services**
- Undertake a **cost/benefits analysis** to help make the case for investment
- Address **sustainability** issues and the associated **long-term costs**



# GUIDANCE, TRAINING AND SUPPORT



## ■ Single Point of Information

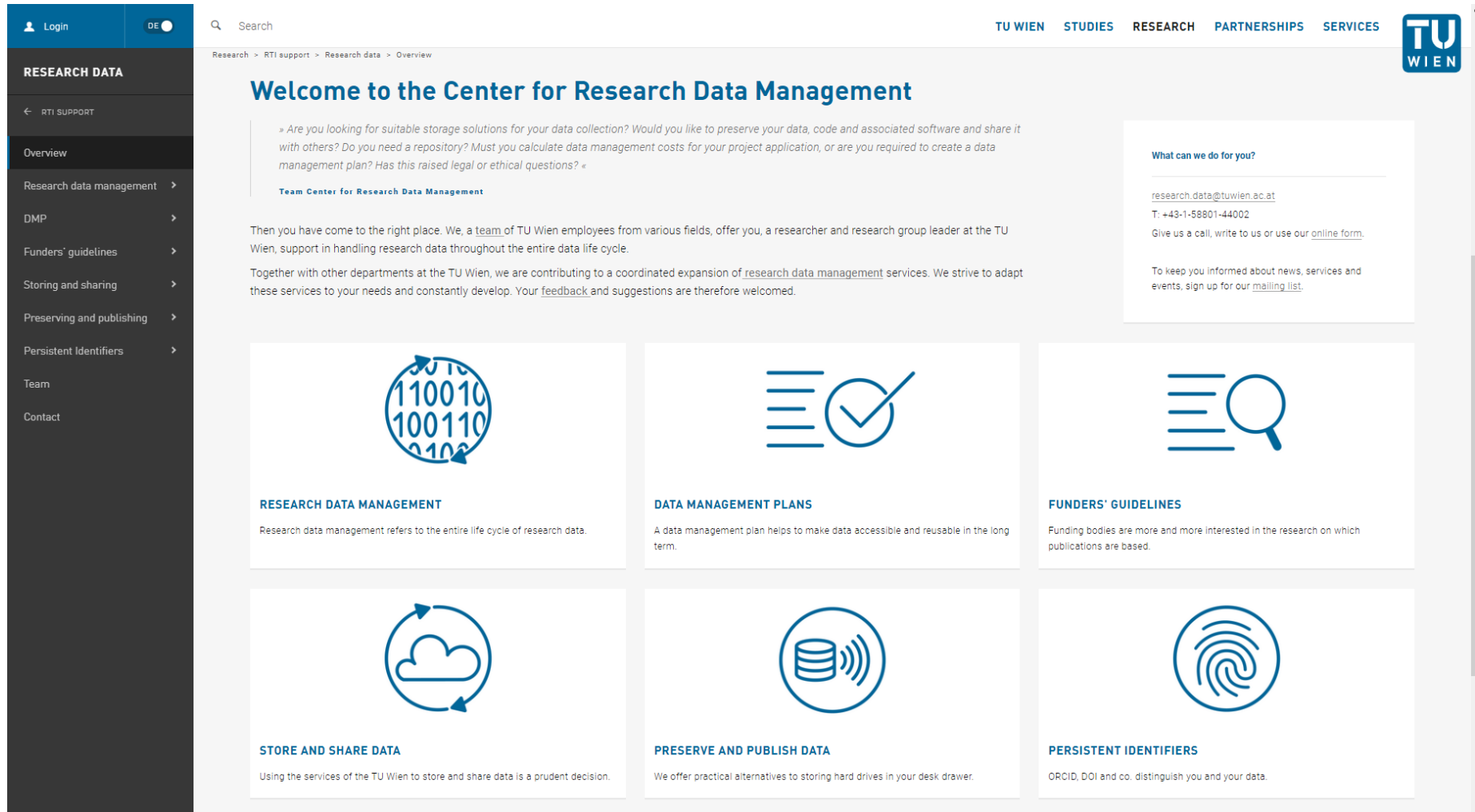
- Website
  - Collate details of existing institutional support to provide a single, joined-up RDM guidance website
- Helpdesk
  - Coordinate the provision of support, either through a central helpdesk or well-signposted contacts

## ■ Trainings and Consultations

- Provide RDM training to support researchers and reskill support staff
- Offer more in-depth consultancy services



# Singe point of information - example



The screenshot shows the 'Center for Research Data Management' website at TU Wien. The page has a dark blue sidebar with navigation links: Login, DE, RESEARCH DATA, RTI SUPPORT, Overview, Research data management, DMP, Funders' guidelines, Storing and sharing, Preserving and publishing, Persistent Identifiers, Team, and Contact. The main content area is white with a blue header bar containing 'TU WIEN', 'STUDIES', 'RESEARCH', 'PARTNERSHIPS', and 'SERVICES'. The page title is 'Welcome to the Center for Research Data Management'. Below the title is a quote: '» Are you looking for suitable storage solutions for your data collection? Would you like to preserve your data, code and associated software and share it with others? Do you need a repository? Must you calculate data management costs for your project application, or are you required to create a data management plan? Has this raised legal or ethical questions? «'. The text then states: 'Then you have come to the right place. We, a team of TU Wien employees from various fields, offer you, a researcher and research group leader at the TU Wien, support in handling research data throughout the entire data life cycle. Together with other departments at the TU Wien, we are contributing to a coordinated expansion of research data management services. We strive to adapt these services to your needs and constantly develop. Your feedback and suggestions are therefore welcomed.' The page features six service tiles: 1. RESEARCH DATA MANAGEMENT (icon: binary code in a circle) - 'Research data management refers to the entire life cycle of research data.' 2. DATA MANAGEMENT PLANS (icon: checklist with a checkmark) - 'A data management plan helps to make data accessible and reusable in the long term.' 3. FUNDERS' GUIDELINES (icon: magnifying glass over a list) - 'Funding bodies are more and more interested in the research on which publications are based.' 4. STORE AND SHARE DATA (icon: cloud with arrows) - 'Using the services of the TU Wien to store and share data is a prudent decision.' 5. PRESERVE AND PUBLISH DATA (icon: database cylinder with signal waves) - 'We offer practical alternatives to storing hard drives in your desk drawer.' 6. PERSISTENT IDENTIFIERS (icon: fingerprint) - 'ORCID, DOI and co. distinguish you and your data.'

<https://www.tuwien.at/en/research/rti-support/research-data/overview/>

## Guidance, training and support

- Who can do this?
- What is the profile of the best candidate?

‘We need 500.000 respected data stewards to operate the  
European Open Science Cloud’

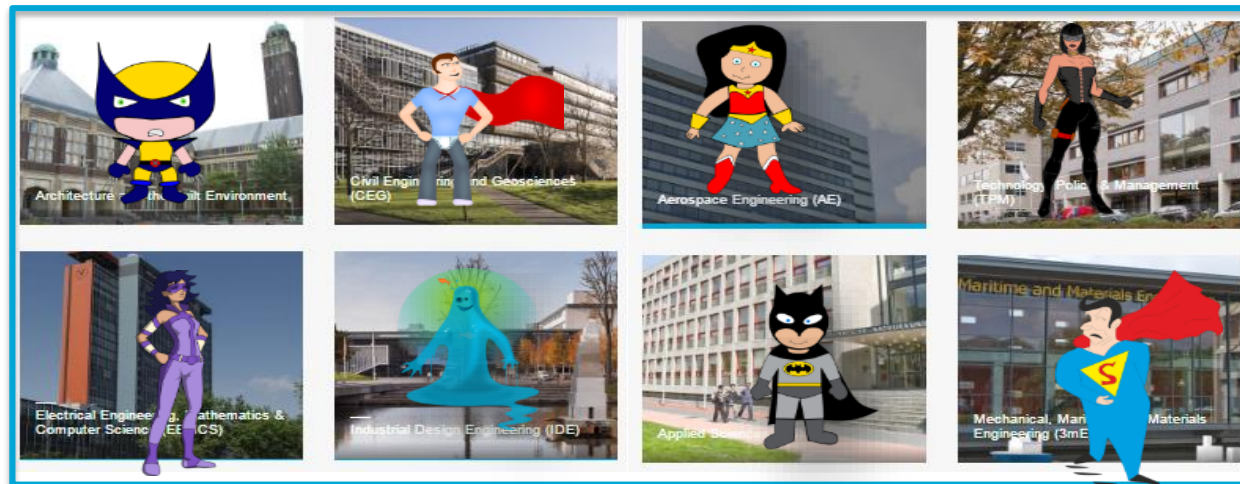
<http://e-irg.eu/news-blog/-/blogs/we-need-500-000-respected-data-stewards-to-operate-the-european-open-science-cloud>

- Data stewards



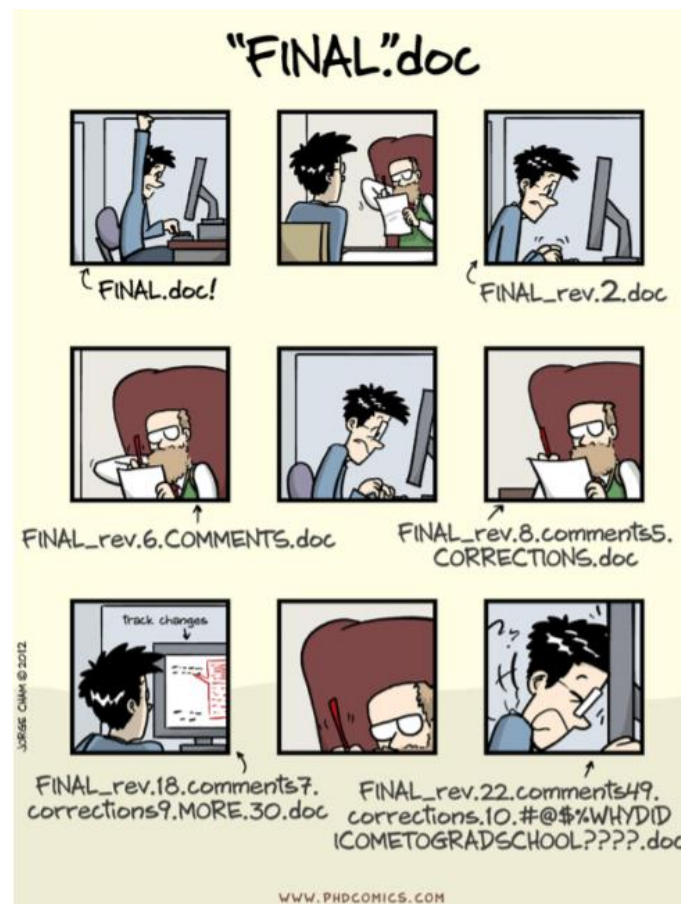
# Data Stewardship

- TU Delft in NL as an example
  - Subject-specific Data Steward at every Faculty
  - Half-time positions
  - PhD holders in a given area proffered
  - Strategic coordination from the Library



# What do data stewards do?

- Analyse data management needs
- Provide advice to researchers
- Train and inspire
- Help comply with funders' and journals' policies
- Develop faculty research data policies: roles and responsibilities of the different stakeholders
- Trusted point of contact for data management questions





FEBRUARY 23, 2018

# We are hiring (again!) – Data Steward position at TU Delft



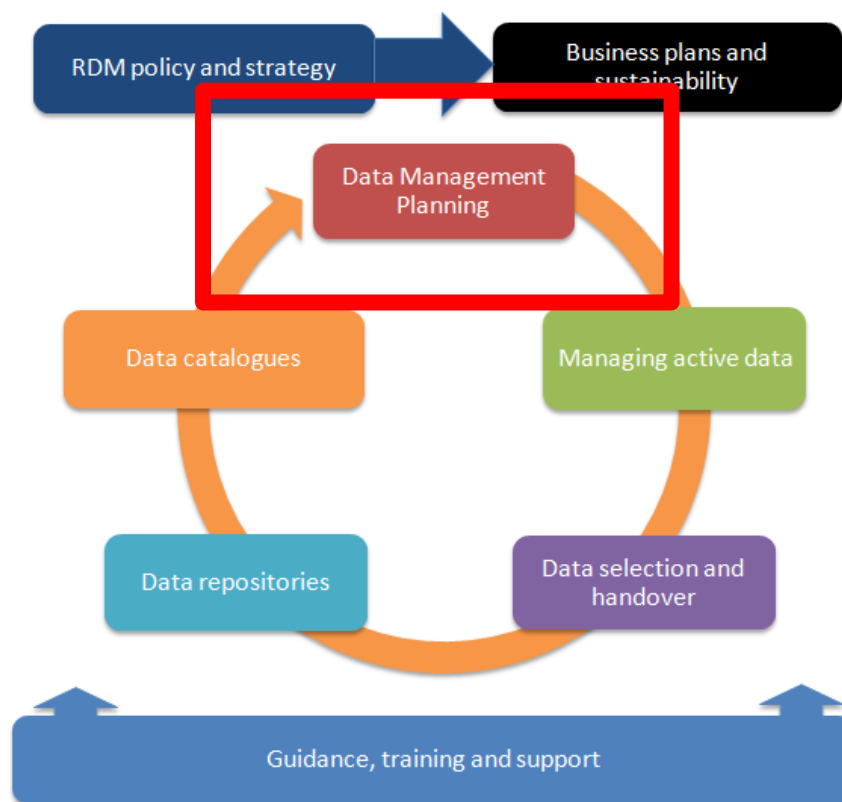
We have an exciting job opening for a Data Steward at TU Delft at the Faculty of Architecture & Built Environment and the Faculty of Industrial Design (joint appointment): <https://www.academictransfer.com/employer/TUD/vacancy/45483/lang/en/>

- Closing date: 15 March 2018
- Salary: up to € 4084/month
- We are looking for individuals enthusiastic about data management and who have a PhD degree in the relevant subject area (or equivalent experience).

This is a great chance to join the dynamically growing team of [Data Stewards at TU Delft](#) and to contribute to a cultural change in research data management in a disciplinary manner. The job is really about inspiring the research community and improving day to day practices, and not about policy compliance.

All informal inquiries can be directed to me: [M.Teperek@tudelft.nl](mailto:M.Teperek@tudelft.nl)

.....



# DATA MANAGEMENT PLANNING



# Data management planning

- Data Management Plan (DMP) describes:
  - how the data will be created
  - how it will be documented
  - who will be able to access it
  - where it will be stored
  - who will back it up
  - whether (and how) it will be shared & preserved





# Do you really need your own DMP template?

- 1. Does your institution encourage/require that researchers write Data Management Plans, even if their funder does not require one?  
If so, what information are researchers asked to provide?
- 2. Do DMPs submitted to research funders supply the information your institution asks for (if any), or are there additional questions that you want to ask?
- 3. What guidance, examples and suggested answers can you provide to help researchers write DMPs?

[<http://www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMPonline-customisation-guidelines.pdf>]

## If you need one...

- Identify who and why will need this information
- Browse existing checklists for help
- Provide a template
  - questions
  - guidance
  - possible answers
- Identify overlaps of your template (e.g. with Horizon 2020)
- Test it and incorporate feedback

## It is not only about DMPs...

- Design the whole data management process
  - do you need different DMP phases?
  - who receives the DMPs?
  - who reviews them?
  - who provides assistance?
  - etc.
- Provide a reviewers' guide
  - clear judgment criteria
- Offer default data location
  - if you require people to deposit data and describe in a DMP, provide a place for doing that

# It is not only about DMPs...

- Establish a policy
  - make a use of DMPs obligatory
  - consider incentives and penalties
- Provide support
  - tools
  - examples
  - trained staff for assistance in DMP creation
    - related problem: Where do you find people with expertise?
- Awareness and understanding
  - offer training material
  - organise workshops
  - related problem: How to teach correct data management practices at schools...



# MANAGING ACTIVE DATA

# Managing active data

- Host on your own



Open Science Framework



**GitHub**



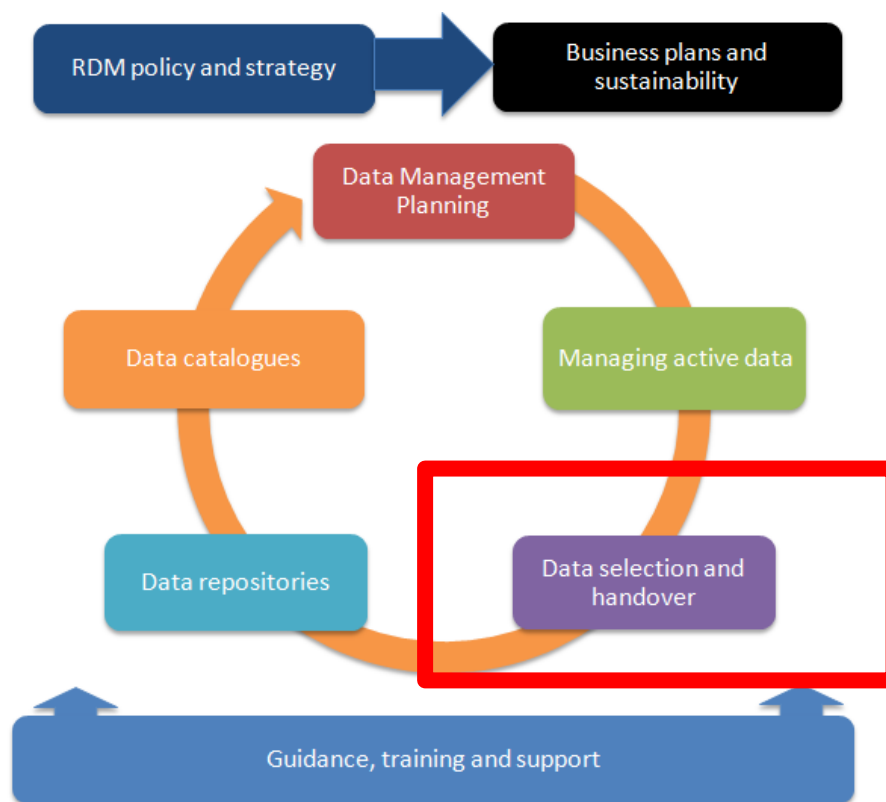
- Outsource?

- Sensitive data?
- Jurisdiction?



## Managing active data

- Review data holdings and RDM practices to see if the current infrastructure and systems are sufficient
- Where appropriate, make a case for investment to provide additional research data storage
- Develop procedures for the allocation and management of research data storage
- Provide flexible RDM systems to support the creation, management and sharing of data that meet a diverse range of research contexts and needs



# DATA SELECTION AND HANDOVER



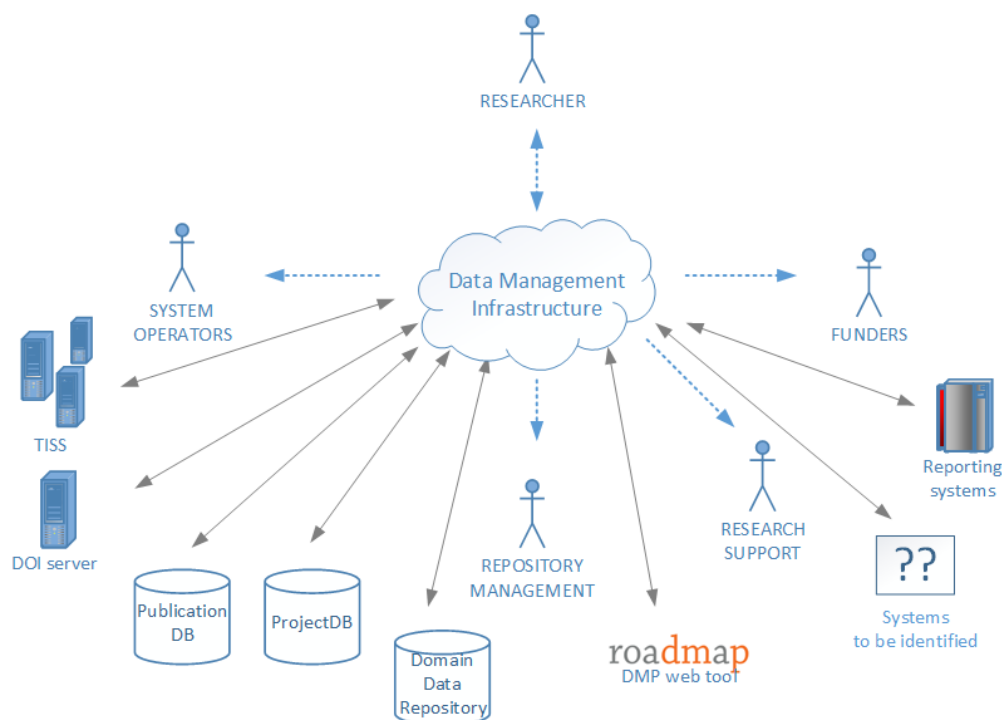
# Data selection and handover

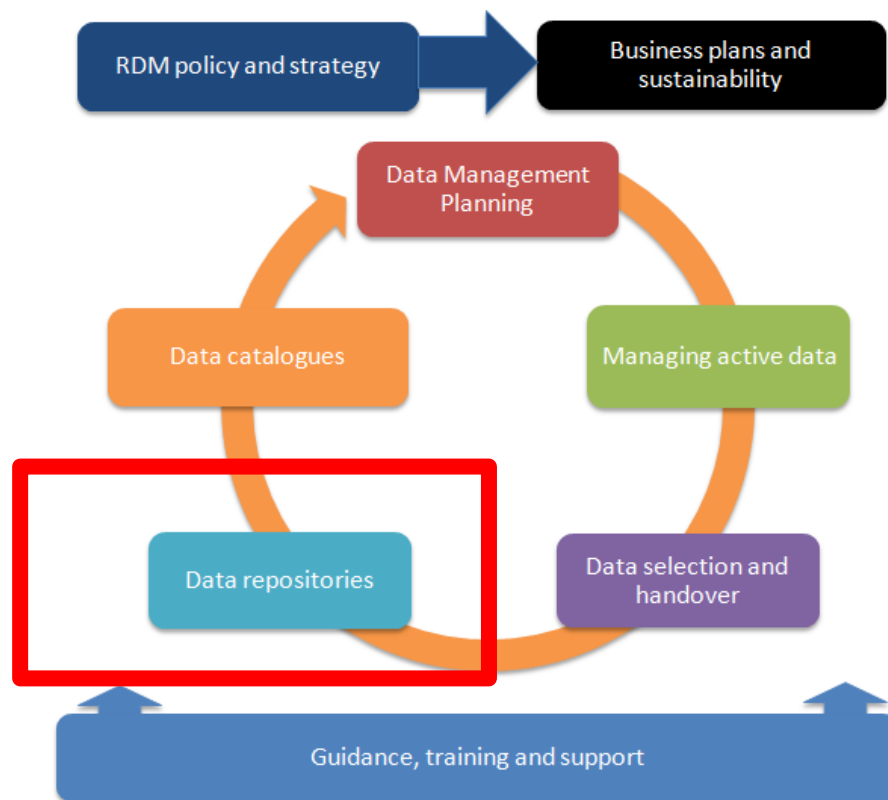
- Identify which data fall under your remit and establish criteria to guide decisions on what to keep
- If establishing a repository, develop deposit agreements and high-level guidance
- Develop or use existing deposit tools to ease the process of handover
- Advocate the benefits of data deposit to encourage uptake
- Support research groups to develop guidance and offer input to decisions



# Data selection and handover

- Data Management Infrastructure
  - Integrate existing systems
  - Seamless data flow
- More details on process monitoring and automated data capturing in the lecture on Reproducibility



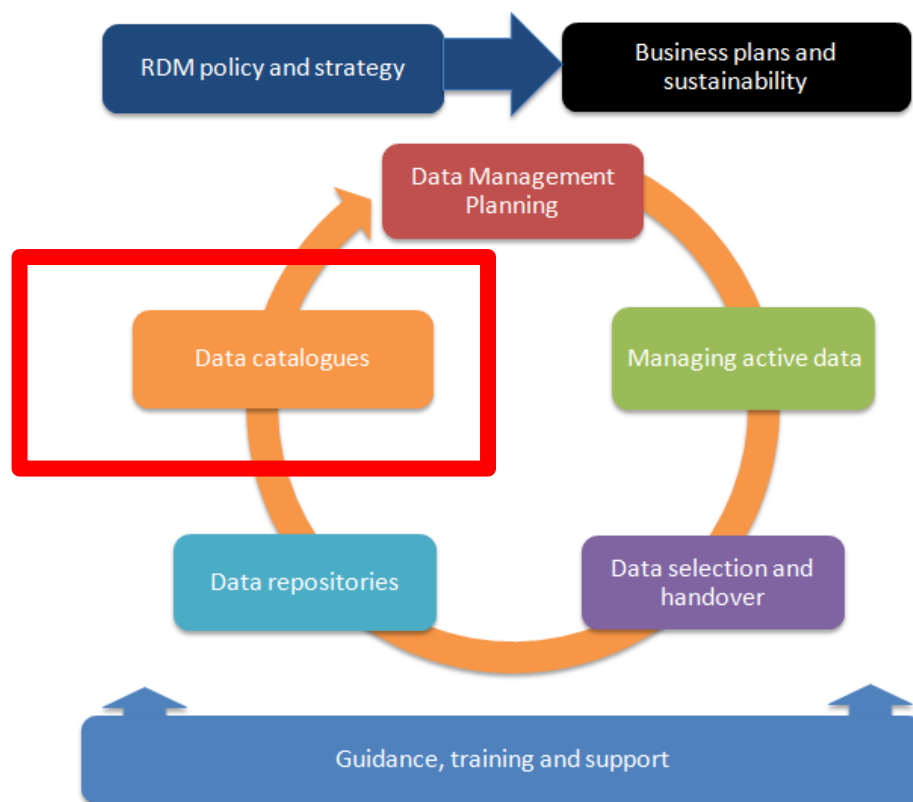


# DATA REPOSITORIES

# Data repositories

- Discussed in the lecture on Repositories
- Where required, develop and maintain an institutional data repository
- Explore options to align your repository with existing systems, for example for Research Information Management
- Decide what role external data repositories will play in your RDM strategy
- Provide guidance and support to direct researchers to relevant services





# DATA CATALOGUES

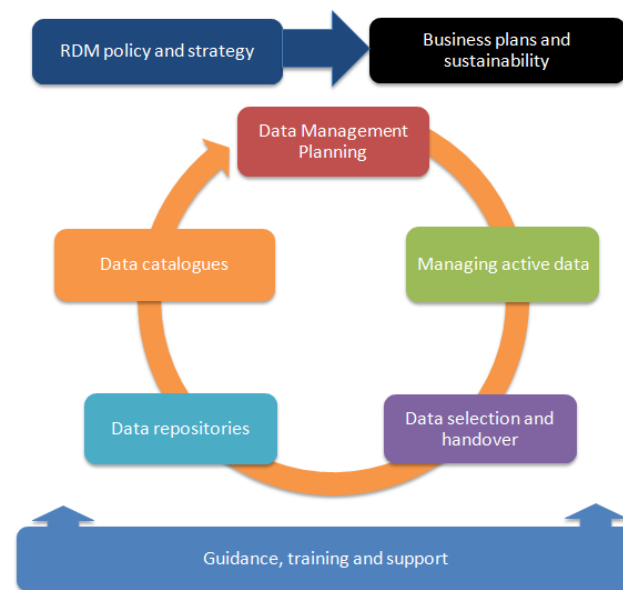
# Data catalogues

- Topic covered in the lecture on Repositories
- Define the metadata you need to record research datasets
- Establish a system for capturing and displaying a record of research data holdings
- Integrate systems to harvest data and embed metadata creation into existing workflows
- Expose metadata for inclusion in national catalogues or other relevant services



# Summary

- It is not only about technical solutions!
- Integration is crucial
  - systems
  - stakeholders groups
  - services
- Develop a vision and a plan
- Include all stakeholders
- Make incremental development



# REPOSITORY CERTIFICATION



# Motivation

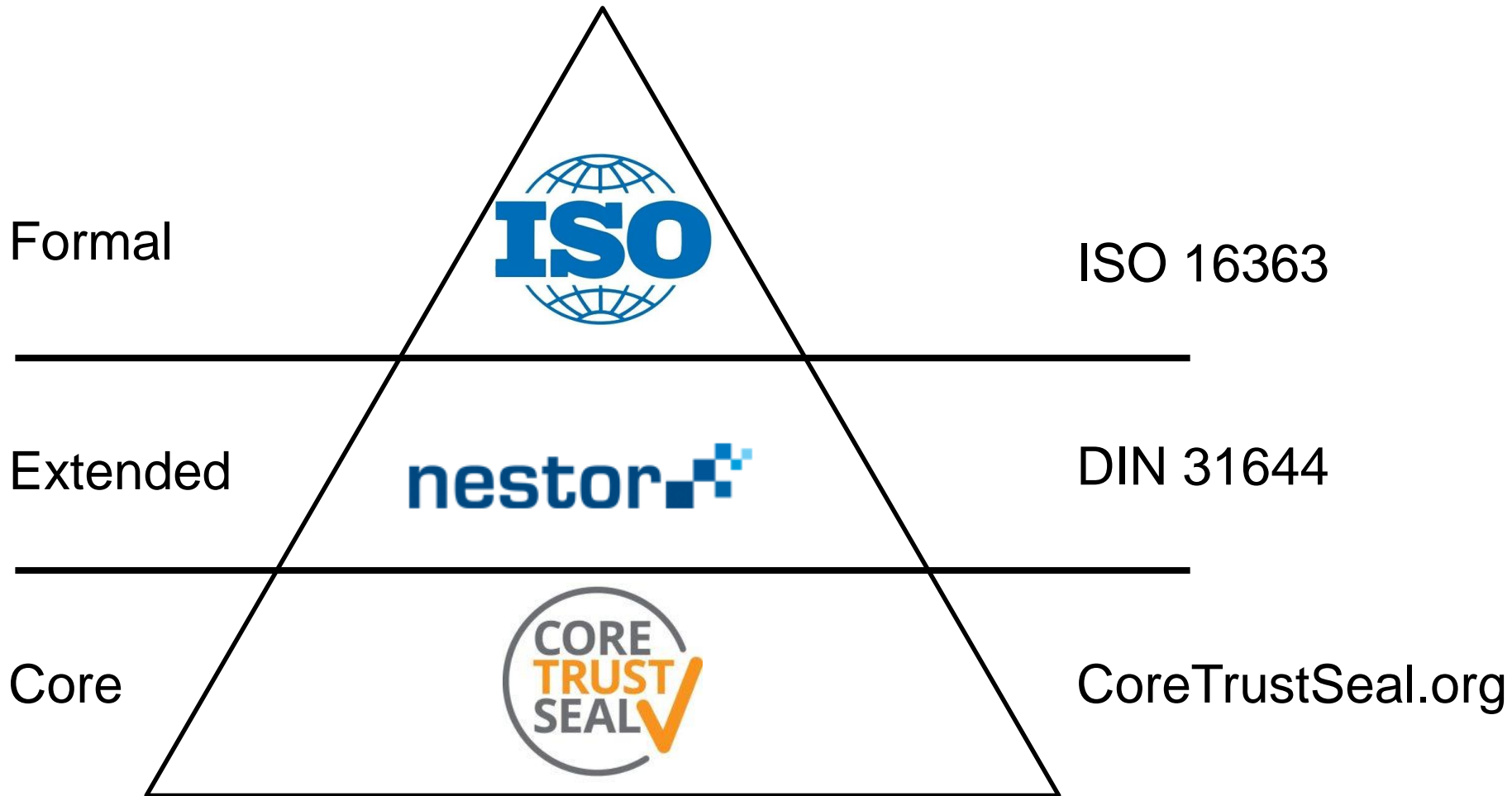
- **Trust** that data remains useful and meaningful into the future
  - Will the repository exist in the future?
  - How is data stored?
  - Is a preservation plan in place?
  - etc.

**re3data.org**  
REGISTRY OF RESEARCH DATA REPOSITORIES

Are all of them  
trusted repositories?



# Certification Standards



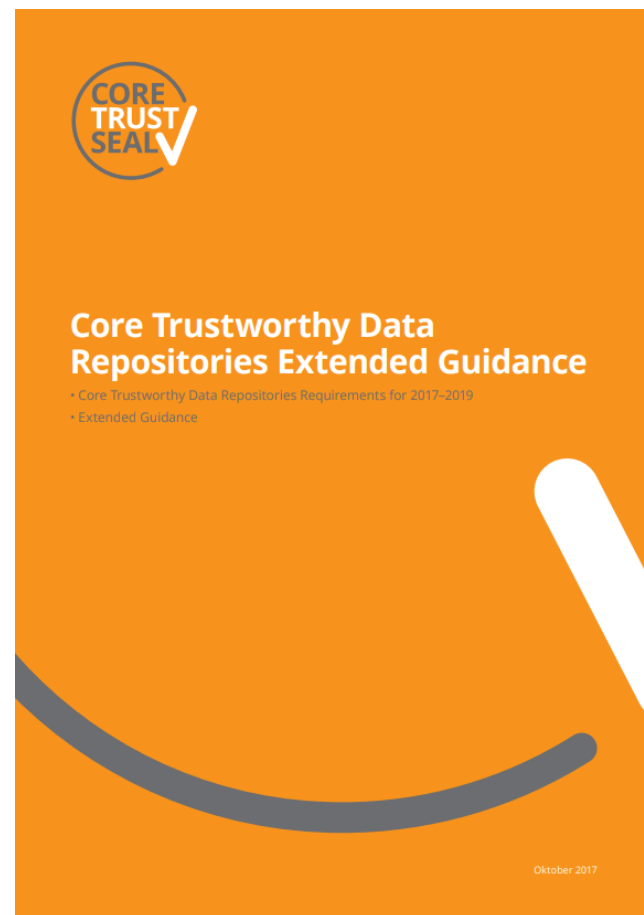
# Core Trust Seal (CTS) - procedure

- Self assessment based on a checklist
- Guidance
  - online tools
  - documents and webinars
- Review of the self assessment by two reviewers
- Assessments publically available
- Renewal every three years



# CTS Requirements

- Organisational Infrastructure (6)
- Digital Object Management (8)
- Technology (2)



<https://www.coretrustseal.org/wp-content/uploads/2017/01/20171026-CTS-Extended-Guidance-v1.0.pdf>

## Organisational Infrastructure (6)

- R1. The repository has **an explicit mission** to provide access to and preserve data in its domain.
- R3. The repository has a **continuity plan** to ensure ongoing access to and preservation of its holdings.
- R5. The repository has **adequate funding** and sufficient numbers of **qualified staff** managed through a clear system of governance to effectively carry out the mission.

- R7. The repository guarantees **the integrity and authenticity** of the data.
- R10. The repository assumes responsibility for **long-term** preservation and manages this function in a planned and documented way.
- R12. Archiving takes place according to **defined workflows** from ingest to dissemination.
- R13. The repository enables users to **discover the data** and **refer to them in a persistent way** through proper citation.

## Technical (2)

- R15. The repository functions on **well-supported operating system** and other core infrastructure software is using hardware and software technologies appropriate to service it provides to its Designated Community.
- R16. The technical infrastructure of the repository provides for **protection** of the facility and its data, products, services, and users.

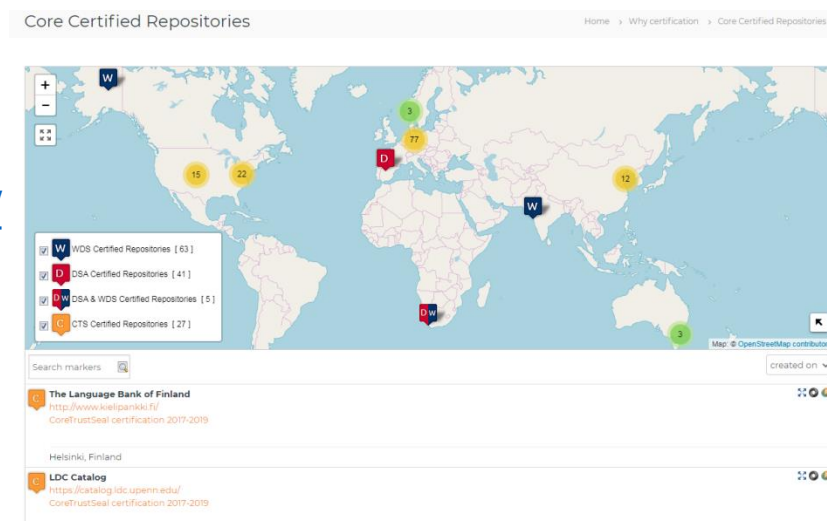
# Self-assessment examples

## ■ Austrian example

- A Resource Centre for Humanities Related Research in Austria
  - <https://arche.acdh.oeaw.ac.at>
- Self assessment
  - <https://www.coretrustseal.org/wp-content/uploads/2018/03/ARCHE.pdf>

## ■ More examples

- <https://www.coretrustseal.org/why-certification/certified-repositories/>





## X. Preservation plan

**R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.**

Compliance Level:

Response

Guidance:

The repository, data depositors, and Designated Community need to understand the level of responsibility undertaken for each deposited item in the repository. The repository must have the legal rights to undertake these responsibilities. Procedures must be documented and their completion assured.

For this Requirement, responses should include evidence related to the following questions:

- Is a preservation plan in place?
- Is the 'preservation level' for each item understood? How is this defined?
- Does the contract between depositor and repository provide for all actions necessary to meet the responsibilities?
- Is the transfer of custody and responsibility handover clear to the depositor and repository?
- Does the repository have the rights to copy, transform, and store the items, as well as provide access to them?
- Are actions relevant to preservation specified in documentation, including custody transfer, submission information standards, and archival information standards?
- Are there measures to ensure these actions are taken?

## 10. Preservation plan

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

As its primary preservation strategy, ARCHE performs migration of formats as opposed to providing software emulation. It aims to establish a high level of transparency with its depositors and users. Thus the deposition agreement and other relevant informative sections of our website highlight our responsibilities and our rights to copy, transform, store and provide access to the deposited items. All the actions relevant to preservation are specified in our documentation.

(this is just an excerpt of the full answer)

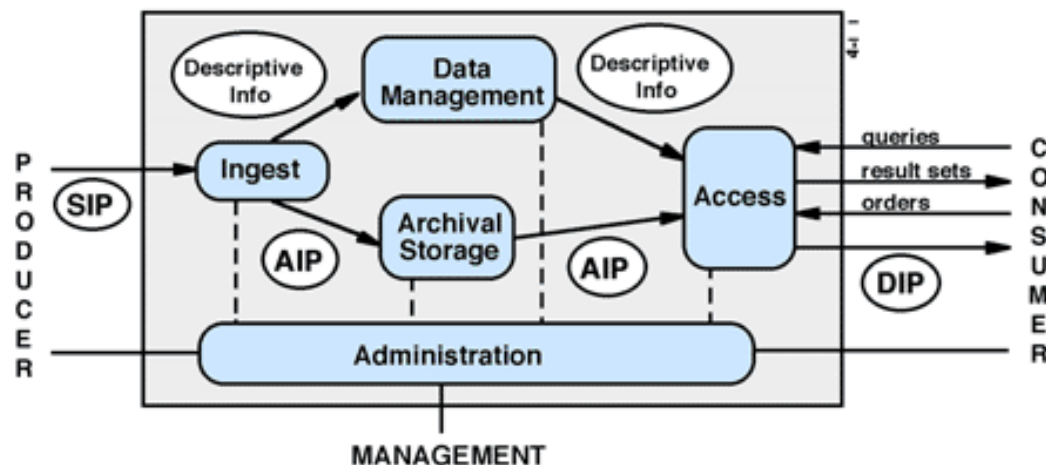
<https://www.coretrustseal.org/wp-content/uploads/2018/03/ARCHE.pdf>

- DIN 31644
  - Deutsches Institut für Normung
  - 'Criteria for trustworthy digital archives'
- Derives from Trustworthy Repositories Audit & Certification (TRAC)
  - document describing the metrics of an [OAIS](#)-compliant digital repository
  - TRAC is discontinued, replace by the ISO 16363
  - translated to German
- Uses OAIS terminology



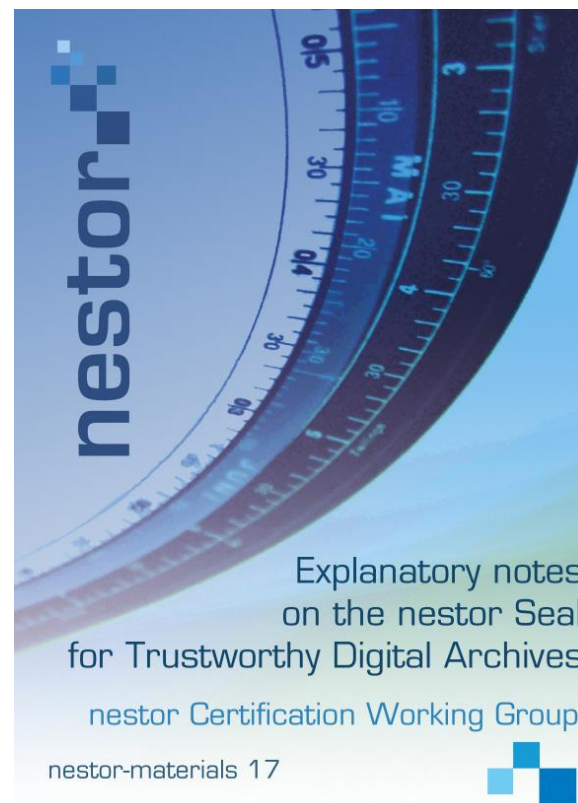
# Open Archival Information System (OAIS)

- Was discussed in a separate lecture
- OAIS
  - conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term
  - defines concepts, terminology
  - not to be instantiated!



# NESTOR requirements overview

- 34 criteria
  - 12 obligatory
  - other can be excluded when justified
- auxiliary questions
- self-assessment
- 2 external reviewers



<https://d-nb.info/1047613859/34>

**C27 Identification**

A digital archive should use internal identifiers to manage the information objects and their representations and, where applicable, their parts and relationships (part/totality, different variants, versions etc.), especially to ensure unique assignment of the content data to the metadata.

The use of externally visible, standardised persistent identifiers ensures reliable tracing of the information objects and their representations, and consequently also access.

**To what extent must the criterion be met?**

An average of 7 points must be achieved in the assessment of the applicable criteria C13 - C34.

**Explanation:** The information objects, representations and their parts are permanently linked to each other. These links can only be preserved through the use of persistent identifiers. The identifiers should not change over the course of time (i.e. be permanent) and should be created using uniform specifications. They should be recognisable to external users, producers and others. By entering the identifier, external users should be able to find and use the required object. Possible specific requirements for identifiers are described e.g. in DIN 13646 "Requirements for the long-term handling of persistent identifiers".

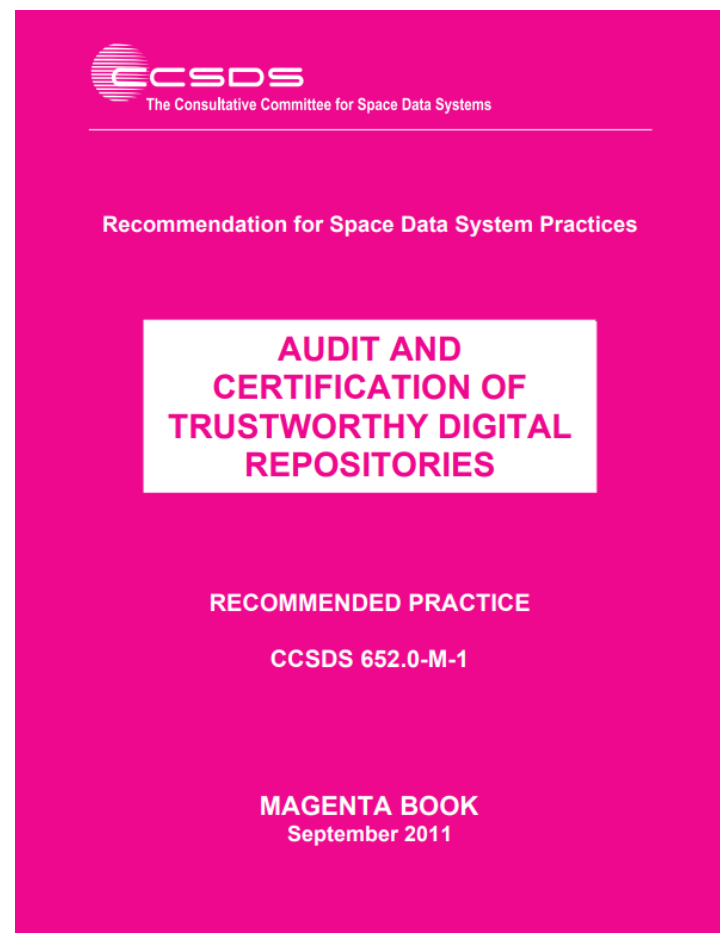
**Questions**

- Which identifiers does the digital archive use?
- Which procedure has been used to give unique identifiers to all information objects, representations and their parts, and to all content and metadata?
- How is the identifier-based assignment conducted?
- How is the permanence of the identifiers ensured?
- How are the identifiers made available to external users?

**Documents:** Specification of the internal and external identifiers



- Uses OAIS terminology
- Based on TRAC
- Over 100 metrics
- Detailed instructions
- Full external auditing process
- ONE repository  
certified since 2012!



<https://public.ccsds.org/Pubs/652x0m1.pdf>

- **Organizational infrastructure**
  - Governance and organizational viability
  - Financial sustainability
  - ...
- **Digital object management**
  - Ingest: acquisition of content
  - AIP Preservation
  - ...
- **Infrastructure and security risk management**
  - Security risk management



- Metrics and their structure:
  - Statement of requirement
  - Supporting text
  - Examples of Ways the Repository can Demonstrate it is Meeting this Requirement
  - Discussion

## **4.3 PRESERVATION PLANNING**

### **4.3.1 The repository shall have documented preservation strategies relevant to its holdings.**

#### **Supporting Text**

This is necessary in order that it is clear how the repository plans to ensure the information will remain available and usable for future generations and to provide a means to check and validate the preservation work of the repository.

#### **Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement**

Documentation identifying each preservation risk identified and the strategy for dealing with that risk.

#### **Discussion**

These documented preservation strategies will describe how the repository will act upon identified risks, as part of the preservation strategic plan. These preservation strategies and the preservation strategic plan will typically address the degradation of storage media, the obsolescence of media drives, and the obsolescence or inadequacy of Representation Information (including formats) as the knowledge base of the Designated Community changes, and safeguards against accidental or intentional digital corruption. For example, if

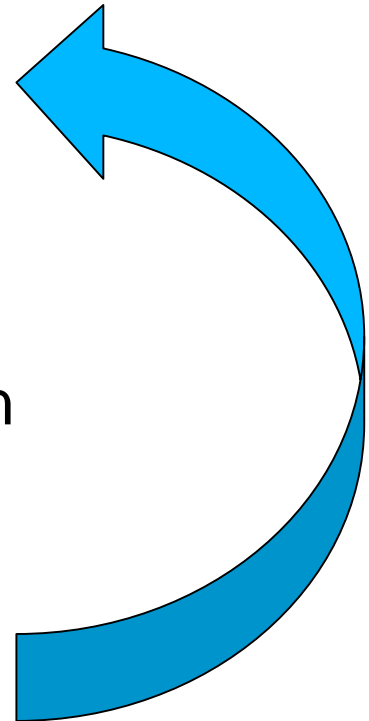
# Metrics – self assessment

4.3 PRESERVATION PLANNING					
	Metric	Supporting Text	Examples of Documents the Repository can use to demonstrate it is Meeting this Requirement:	Brief description of evidence (add rows if necessary to list all relevant documents for a metric) Use short titles for documents. Provide detailed	Explanation of how the repository addresses this metric
4.3.1	<b>THE REPOSITORY SHALL HAVE DOCUMENTED PRESERVATION STRATEGIES RELEVANT TO ITS HOLDINGS.</b>	This is necessary in order that it is clear how the repository plans to ensure the information will remain available and usable for future generations and to provide a means to check and validate the preservation work of the repository.	Documentation identifying each preservation risk identified and the strategy for dealing with that risk.		
4.3.2	<b>THE REPOSITORY SHALL HAVE MECHANISMS IN PLACE FOR MONITORING ITS PRESERVATION ENVIRONMENT.</b>	This is necessary so that the repository can react to changes and thereby ensure that the preserved information remains understandable and usable by the Designated Community.	Surveys of the Designated Community of the repository.		
4.3.2.1	<i>The repository shall have mechanisms in place for monitoring and notification when Representation Information is inadequate for the Designated Community to understand the data holdings.</i>	This is necessary in order to ensure that the preserved information remains understandable and usable by the Designated Community.	Subscription to a Representation Information registry service; subscription to a technology watch service, surveys amongst its Designated Community members, relevant working processes to deal with this information.		

[http://www.iso16363.org/?smd\\_process\\_download=1&download\\_id=30](http://www.iso16363.org/?smd_process_download=1&download_id=30)

# ISO Process for Audits

- Preparatory work by repository
- First audit and resulting certification
  - Identifies improvements needed
  - Repository prepares improvement plan
- Repository implements improvement plan
- Surveillance audit after a period
- Re-certification



# Why certify a repository (using any scheme)?

- Builds stakeholder confidence in the repository
  - researchers', funders', publishers', etc.
- Improves communication within the repository
  - roles and responsibilities
  - processes
- Ensures transparency
  - most assessments are public
- Enables comparison of repositories

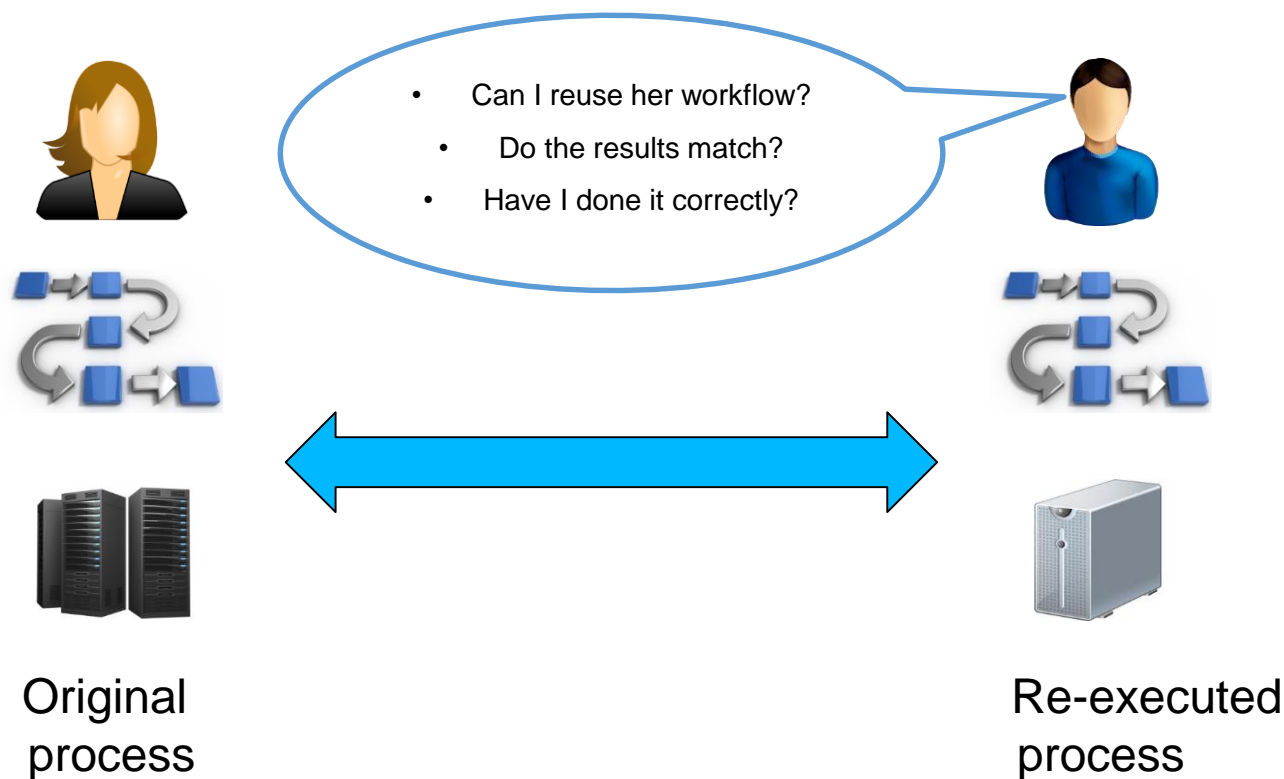
# Summary

# You should know

- What are the lifecycle models and how to use them?
- What are the components of an RDM service infrastructure?
- What is the scope of policies and how they drive DM activities and obligations?
- What are the cost drivers for data curation and how they impact design of RDM services?
- How do develop support services and who are data stewards?
- What to consider when implementing DMPs in an institution?
- What standards exist and why certification matters?
- What are the certification criteria and what is the process?

# Next lecture

- 27 May 2019
- Reproducibility





# Acknowledgments

- 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>
- Andreas Rauber. (2017, November). Research Outputs Management: what services are involved? File (Version 1.00). Zenodo. <http://doi.org/10.5281/zenodo.1063548LEARN>
- Teperek, Marta. (2017, November). Data Stewardship - addressing disciplinary data management needs at TU Delft. Zenodo. <http://doi.org/10.5281/zenodo.1064794>
- Budroni, Paolo. (2017, May). The LEARN Project Using the LEARN RDM Policy & Guidance. Zenodo. <http://doi.org/10.5281/zenodo.579993>