Developing a Research Data Policy

Core Elements of the Content of a Research Data Management Policy

This document may be useful for defining research data, explaining what RDM is, illustrating workflows, describing benefits and informing about funder requirements. Additionally, legal and ethical issues are considered, as well as the relationship with publishers.

Elements	Content	Issues that may be addressed
Definitions of Research Data	Define the scope of the policy and the definition of data according to the definition of the institution	Consider the elements of this list: • Raw data • Primary data • Inconclusive results • Negative results • Positive results • Processed data • Released data • Shared data • Published data • Open access published data
Institutional elements / aspects		
Roles	Role of: The institution / The University IT services Scientific Library Research Support Service Legal services The researcher / the PI Data scientist Students in PhD degree Funding body	The policy could define roles, in order to foster / promote the scientific process. Concerning the **University**, consider these issues: • The University will provide training, support, advice, and when appropriate, guidelines and templates for the research data management and research data management plans. • The University will provide procedures, mechanisms and services for storage, backup, registration, deposit and retention of research data assets to support current and future access, during

and after completion of research projects.

Consider also these issues:

- Any data stored externally, for example in an international data service or domain repository should be registered with the University.

- Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate policies and conditions.

- Should the University / the funding body maintain ownership of all data produced by its "own" researchers within its "own frame"?

Concerning the **IT-Services**, consider these issues:

- Maintenance of the system(s)
- Access to the system(s)
- Preserving integrity of data
- Authenticity of data
- Ensure access according to the needs of data provider & rights management system
- Provide appropriate tools

Concerning the **Repository management**, consider these issues:

- Preservation of data
- Management of data
- Access to data
- Extraction of data
- Reuse of data
- Visualization of data

Concerning the **Libraries**, consider these issues:

- How can Libraries offer a stable environment in order to support research data management?

- How can they contribute in expanding digital collections and developing digital infrastructures for research data?

- How can libraries have a central role in coordinating the implementation of data management policies and procedures in institutions?

Concerning the **funding body**, consider these issues:

- Requirements of the Funder
- Costs related to long term archiving (deposit) of research data
- Costs related to further reuse of data (setting of interfaces and visualization issues)

Concerning the **data producer** / **PI** , consider these issues:

- Ensure quality of content
- Organization of data
- Clear file structure
- Consistent file naming and version control
- Choice of format
- Documentation of data & metadata

Responsibility for research data management through a sound research data management plan during any research project or program lies primarily with the Principal Investigators (PIs).

Research Data Lifecycle		
	- Creation of Data - Preparation and Preservation planning - Ingest and Deposit - Management - Assignment of PI - Integrity - Storage - Backup - Registration - Retention - Future access - Sharing - Extraction - Publication - Further reuse	

Appropriate storage systems for long term preservation	Who is providing the mechanisms and services for storage, backup, registration, deposit and retention of research data in support of current and future access?	Checklist of relevant issues: • Data loss • Recovery strategies • Training
Legal compliance **Legal and Ethical issues**	Who is responsible for legal compliance and security along all phases of research data lifecycle? This concerns data, metadata, related applications as well as technical systems.	The legitimate interests of the subjects of research data must be protected. 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. Checklist of relevant issues: • Copyright issues • Distribution rights • Exploitation rights (rights to re-use, rights to publish, etc.) • Privacy issues • Employee protection Consider also these issues: • Who is in charge of negotiations with publishers? • Who is in charge of negotiations with funders? • Who is providing licensing mechanisms and models (licenses)?
Sensitive data	Who is responsible?	Consider anonymization issues: • Disclosure agreements • Privacy issues Focus on: • The issue of re-use of data • The long-term preservation issue

Data Management Plans	Who is developing DMPs? Who is reviewing the DMPs? Who is in charge of the institutional DMPs? Who is responsible for DMP training? Who will run the first level support?	Introduce guidelines for submission of project proposals. Should all new research proposals (from date of adoption) include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication?
Metadata complex	Consider also: • Contextual information • Provenance • Relations between datasets • Annotations • Intellectual property rights of metadata • Technical issues • Terminology issues • E-accessibility issues	Information concerning the metadata
Standards	Who is responsible for the standardization of processes?	Research data may be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence. Checklist of relevant issues: • Quality assurance • Metadata • Technical processes
Visualization models	Who is responsible: • …for the definition of visualization models? • …for the extraction of data? • …for the implementation of the visualization?	
Academic based teaching	Who is in charge of academic based teaching related to online learning platforms concerning research data?	Is this an issue at the institutions (concerning research data management and the re-use of research data)?

Traditional knowledge	Who is in charge of safeguarding traditional knowledge of the institution?	Consider the integration of the organization's culture and traditions.
Further issues		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 storage and retention in an appropriate national or international data service or domain repository, or a University repository.
Content provider entities		
Quality of content of data	Who is responsible?	Usually the Principal Investigator (the data producer).
Quality of metadata of research data	Who is responsible?	Usually the Principal Investigator (the data producer).
Data access	Who is responsible?	Consider various levels: • from restricted access to open access Checklist of relevant issues: • Legal and ethical issues
Data sharing	Who is responsible?	Checklist of relevant issues: • Legal and ethical aspects • Technical aspects • Management aspects (e.g. metadata)
Data release	Who is responsible?	Checklist of relevant issues: • Legal and ethical aspects • Technical aspects • Management aspects
Selection of Long Term Archiving /Sharing / Release	Who is responsible? Who will take care of this data? Who will hand over this data to whom/where? Who will decide about sharing and releasing? Who will decide about the forms of visualization?	Focus on the issue of re-use and future visualization of data.

Re-use	Who is responsible: • …for the assignment of correct metadata? • …for clearing licenses of use? • …for ethical issues? • …for technical issues?	Consider the hand over to: • re-using party • additional software applications for further use (for example: Digital Humanities, E-learning systems, visualization and representation tools)
Data retention **Data preservation**	Who will decide about data retention? Who will take care of the preservation processes? Who is responsible for the technical support?	
General topics concerning technical and non-technical aspects		
Format quality and preservation	Concerning the management of the format : Who is responsible for it?	Are there any formats or content models recommended by the policy?
Quality assurance	Who is responsible for the digital objects / data during the different phases of the life cycle of the object from the technical and non-technical points of view?	See also the phases of a Digital Workflow Model (Image I: *Model of Workflow*) representing the phases and describing the different roles (production, preparation, pre-ingest, ingest, management inside an archiving system, extraction, export, visualization)

This graph shows a digital workflow model starting at the left side with the content producer and ending on the right side with the visualization of the content. Data management plans accompany the processes along the entire life cycle.
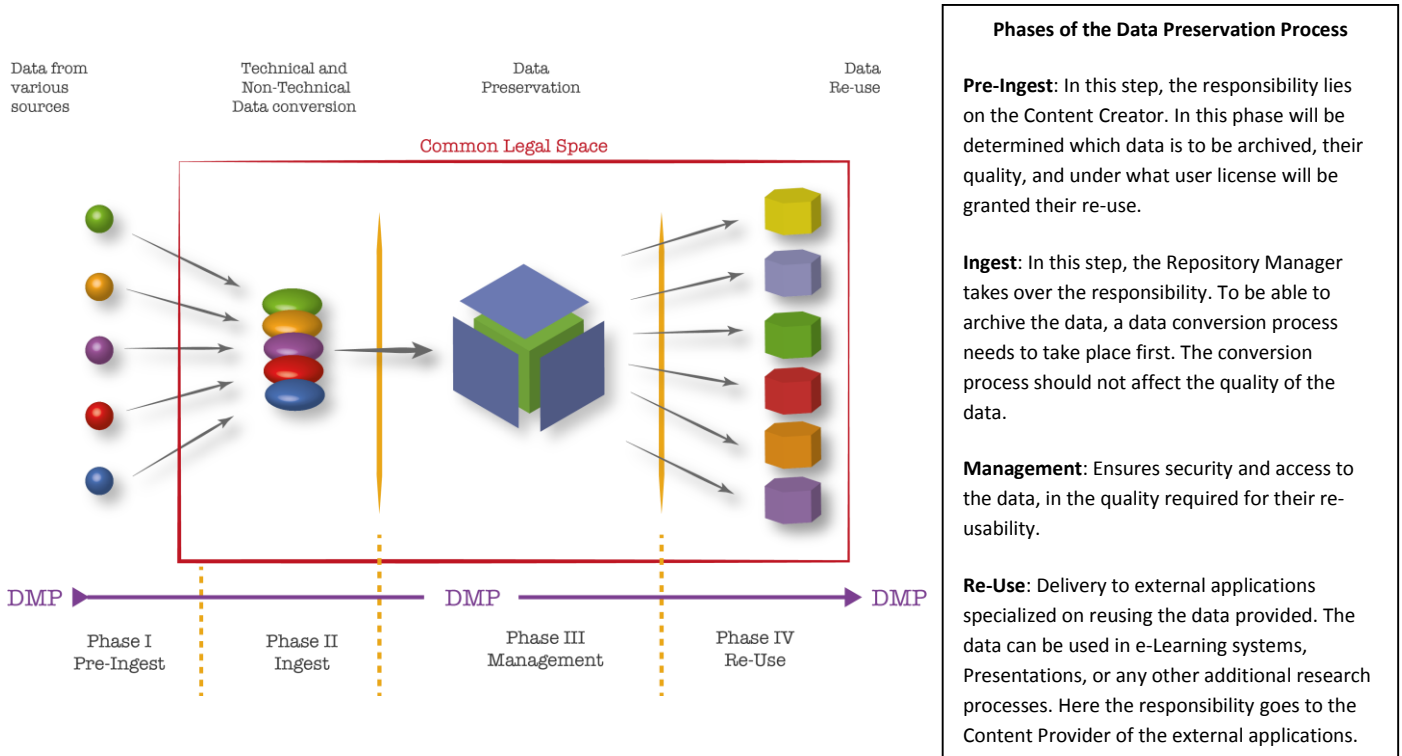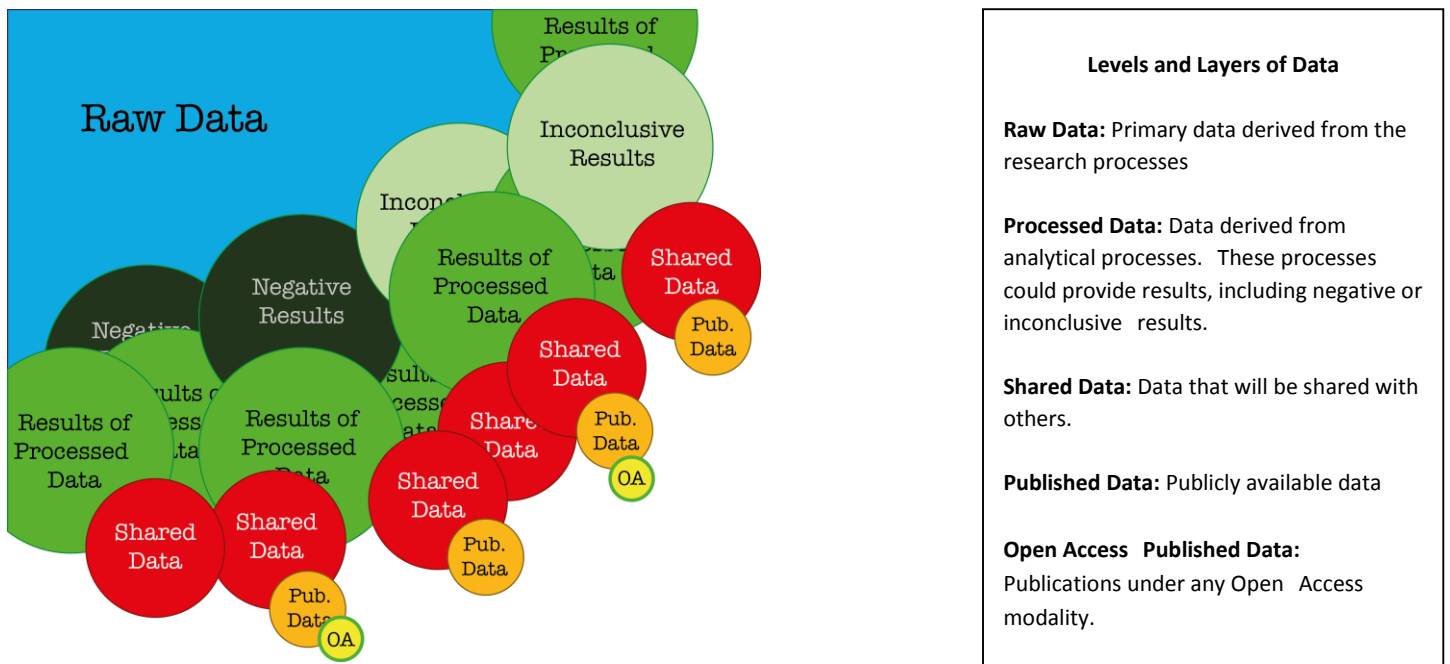

Phases of the Data Preservation Process

Pre-Ingest: In this step, the responsibility lies on the Content Creator. In this phase will be determined which data is to be archived, their quality, and under what user license will be granted their re-use.

Ingest: In this step, the Repository Manager takes over the responsibility. To be able to archive the data, a data conversion process needs to take place first. The conversion process should not affect the quality of the data.

Management: Ensures security and access to the data, in the quality required for their re-usability.

Re-Use: Delivery to external applications specialized on reusing the data provided. The data can be used in e-Learning systems, Presentations, or any other additional research processes. Here the responsibility goes to the Content Provider of the external applications.

Image II: Data from Research Processes: from raw data to open access published data, by Raman Ganguly (http://phaidra.univie.ac.at/o:387241)

Levels and Layers of Data

Raw Data: Primary data derived from the research processes

Processed Data: Data derived from analytical processes. These processes could provide results, including negative or inconclusive results.

Shared Data: Data that will be shared with others.

Published Data: Publicly available data

Open Access Published Data: Publications under any Open Access modality.

Useful links and further related documents:

The following list of useful links and further related documents has been created in order to deepen knowledge about Research Data Management. It comprehends references to Research Data Management policies, Data Management Plans.

[The University of Edinburgh Research Data Management Policy](#)

The list is work in progress