General considerations on data publications PIDs – lessons learned

MARIA JOHNSSON & MONICA LASSI
LUND UNIVERSITY LIBRARY

PUBLICATION WORKFLOWS FOR DATA CURATION, MAX IV 2018-07-03

Parsing the purpose of the workshop

”to work out (1) **the right strategy** to (2) **publish data**, i.e. mint DOIs, either by data becoming public (3) **through policy or by user choice**”

Some aspects:

- The **technical**: trustworthy repository, flows of data and metadata, sustainability…
- The **people**: who are driving and resisting, who will be affected…
- The **organisations**: RIs, principals, management, departments…
- The **external actors**: funders, journals, policymakers…

LUND UNIVERSITY

(2) Publishing data

- Why publish data?
- What’s in it for the users?
- Finding the balance between effort and benefits

LUND UNIVERSITY

3) Driven by policy or user choice?

- *And/or other actors' requirements? Funders, journals, policymakers…*
- Practices of other RIs? What are the users' expectations?
- Do users read (and adhere to) policies? (Raising awareness, making it easy to follow rules)
- Involving users in co-creating policy and services (Using a consultation process)

LUND UNIVERSITY

1) The (right) strategy

- What problem or need does a publishing strategy solve?
- Who is the driving force behind the strategy?
- What is the message to stakeholders, in particular the users?
- Who delivers the message, from whom, to whom? (How to avoid killing the messenger)

LUND UNIVERSITY

A few more lessons learned re policies

- Follow the genre: mixing vision and how-tos creates confusion
- A support infrastructure (tools, experts, information) should be in place or under development
- Communication strategy: making the policy (and services) known
- Ensuring compliance – how?!

LUND UNIVERSITY

Persistent identifiers - PIDs

Current experience from:

- ICOS Carbon portal – reserach infrastructure holding a system for persistent identifiers to datasets, https://www.icos-cp.eu/
- ENVRIplus, sub-project ”Work Package 6” on data identification and data citation, some recent reports, http://www.envriplus.eu/

LUND UNIVERSITY

Persistent identifiers - PIDs

System for persistent identifiers in a research infrastructure

Main objective:

- Internal system for PIDs?
- System to be interoperable with other, external systems?
- Needs and preferences from the users (researchers)?

Technical and administrative:

- On what granual level should PIDs be issued/attributed to datasets?
- PIDs also for non-data objects, such as instruments, stations, people, organizations etc?
- Landing-pages in human-readable and/or machine readable format?

LUND UNIVERSITY

Persistent identifiers - PIDs

PIDs hot topic nowadays

- **Many service providers for PIDs**, such as DataCite, CrossRef, California Digital Library – they are developing and introducting more features
- **Developing infrastructure for PIDs**, i.e. FREYA, Research Data Alliance, CODATA
- **Upcoming need for measuring data citations** in similar ways as measuring publication citations, projects like ”Make Data Count”
- **”PIDapalooza”**, conference or ”festival” solely on different matters on PIDs

LUND UNIVERSITY

Persistent identifiers - PIDs

Citations best practices for RIs (from ENVRIplus WP6, deliverable 6.1)
Technically

A. All datasets intended for citation have a globally unique PID that can be expressed as an unambiguous URL
B. A PID expressed as a URL resolves to a landing page for a dataset
C. The landing page of a dataset is both human-readable and machine-readable and contains the dataset's PID
D. PIDs for datasets support multiple levels of granularity
E. Datasets are described with rich metadata to track provenance information and to create meaningful citations
F. Metadata are accessible even if a dataset is no longer accessible
G. RIs provide a robust resolver and registry for resolving PIDs and for data discovery
H. Metadata protocols and standards are used, that ensure interoperability with related stakeholders, e.g. cataloguing and indexing services
I. Data are published with a clearly defined data usage license

LUND UNIVERSITY

Persistent identifiers - PIDs

Citations best practices for RIs (from ENVRIplus WP6, deliverable 6.1)

In general

- RIs actively promote data citation to users, publishers and other stakeholders in their research community, e.g. by providing documentation, and by providing common citation formats to users

- Citation methods are flexible to support each community while still ensuring interoperability across communities

LUND UNIVERSITY

Persistent identifiers - PIDs

Data publishing at PANGAEA:

- PANGAEA, data publisher in earth and environmental science: https://www.pangaea.de/
- Example of dataset: https://doi.pangaea.de/10.1594/PANGAEA.888371
- Interoperability of PANGAEA's services: https://www.pangaea.de/about/services.php

LUND UNIVERSITY

Persistent identifiers - PIDs

Links to websites, projects, publications

- ICOS Carbon Portal, https://www.icos-cp.eu/
- ENVRIplus, http://www.envriplus.eu/
- ENRVRIplus, Work Package 6, deliverable reports 6.1 and 6.2, go to: http://www.envriplus.eu/deliverables/
- FREYA, https://www.project-freya.eu/en
- FREYA Knowledge Hub, Persistent Identifier Platform, https://project-thor.readme.io/
- RDA Research Data Alliance, https://www.rd-alliance.org/
- RDA, ”Research data needs of the Photon and Neutron Science community IG”, https://www.rd-alliance.org/groups/research-data-needs-photon-and-neutron-science-community.html
- CODATA, http://www.codata.org/
- DataCite, https://www.datacite.org/
- PANGAEA, https://www.pangaea.de/
- Make Data Count, https://makedatacount.org/

LUND UNIVERSITY

Thanks

Maria Johnsson
maria.johnsson@ub.lu.se
+46 46 222 90 40
+46 72 568 53 18

Monica Lassi
monica.lassi@ub.lu.se
+46 46 222 01 06
+46 70 647 10 91

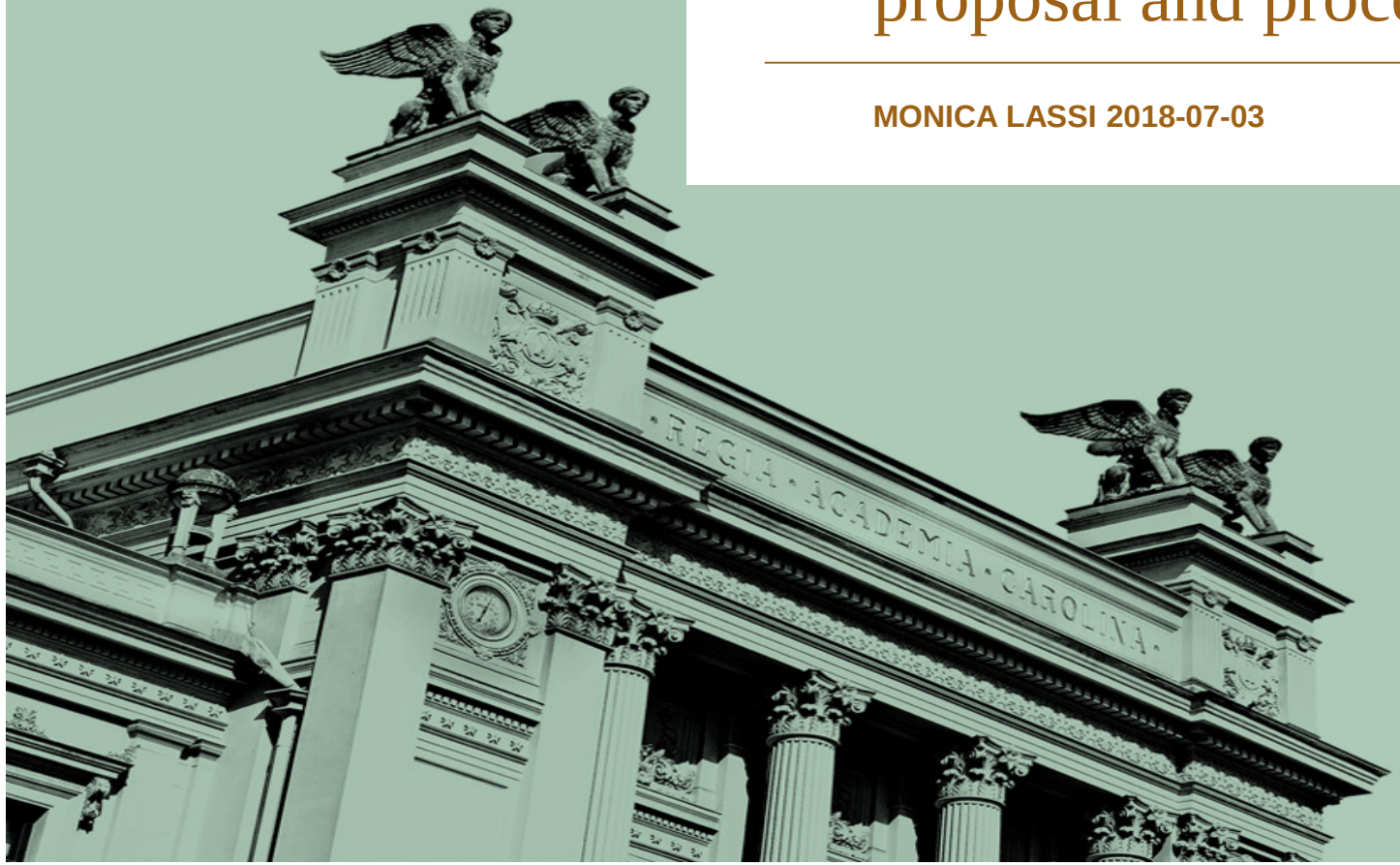
LUND UNIVERSITY

Extrabilder om policyn följer

LUND UNIVERSITY

Lund University research data policy: proposal and process

MONICA LASSI 2018-07-03



Background

- The Research Board established a working group for research data Sept 2016
  - ~10 members representing research and support
- First mission: To draft a policy for the management of research data

LUND UNIVERSITY

The process

- *Sept 2016*:
  - Analysis of ~20 RDM policies
  - Result: a proposal on a very detailed level
- *Spring 2017*: Presentation to Research Board
- *May-Oct 2017*: Presentation and discussion at faculties and infrastructures
- *Feb 2018*: Presentation of revised version to the Research Board
- *June 2018*: Consultation process
- *Late autumn 2018*: Expected formal decision by vice-chancellor

LUND UNIVERSITY

First version

Lund University policy on management of research data

This research data management policy for Lund University describes the aims of the university in promoting and supporting the highest standard of research data management for adhering to good scientific practice. Lund University will provide researchers with the support needed for research data management to the highest standards. This policy describes necessary aspects of research data management, and the roles and responsibilities of researchers and Lund University respectively in contributing to sound research data management.

Lund University acknowledges the value of research data for the scientific community and society. This is described in the Lund University research strategy 2017-2021, which states that Lund University will "Work proactively with issues concerning research data in digital form and the management of research data, particularly concerning storage and sharing of data" (Research strategy for Lund University 2017-2021, p. 4[1])

Research data are defined as all recorded information (regardless of the form or the media in which they may exist) necessary to support or validate observations, findings or outputs of research.

1. Applicability

This policy applies to researchers and other staff employed at or affiliated with Lund University who manage research data in any form.

Lund University recognises that intellectual property rights as well as ethical and legal regulations can set special requirements on the management of research data. Obligations to other stakeholders may also apply and take precedence over this policy.

2. General principles

Research data produced at Lund University should, whenever possible, comply with the so-called FAIR *Guiding Principles* for scientific data management and stewardship[2]. According to these principles research data should be made findable, accessible, interoperable, and reusable.

3. Ownership of research data

Lund University asserts ownership of research data created or collected as part of research conducted during employment or studies at the University. The University grants the creator or collector of research data rights to use and distribute those data.

4. Data Management Plan

At the start of a research endeavour, a Data Management Plan (DMP) should be set up and registered at Lund University. A DMP is a formal document that describes how research data will be managed both during the research and when it is concluded.

5. Metadata

Research data should be described with rich metadata in order to comply with the *FAIR Guiding Principles*. Metadata should be registered in the Lund University metadata registry and linked to the research data, regardless of where the research data are deposited.

6. Accessibility to research data

Research data should be made openly accessible for use whenever possible, unless prevented by legislation or ethical, contractual, or confidentiality obligations.

Lund University encourages depositing of research data in the commonly used research data repositories within their research community, in the Lund University research data repository, or in another research data repository deemed relevant.

7. Retention of research data

Research data records must be retained in the University Archives in accessible formats for at least 10 years in accordance with the Swedish Archives Act, or for 17 years, for EU funded research.

At the end of the retention period, a decision on the further retention of the research data must be made. Disposal or extended retention is decided on in consultation with the researcher, Lund University Archives, and other expertise on research data management at the University.

8. Roles and responsibilities

Individual researchers are responsible for:

- Complying with all relevant legislation, as well as with ethical, contractual, and confidentiality obligations regarding research data management.
- Setting up a Data Management Plan complying with this policy and registering it at Lund University.
- Registering metadata in the Lund University metadata registry.
- Making research data openly accessible whenever possible while complying with legislation and ethical, contractual, and confidentiality obligations.

Head of the department, where the research is managed, is responsible for

- Naming a delegate who can take over all responsibilities of the research data in case the researcher responsible for the research data leaves Lund University.

Lund University is responsible for:

- Providing researchers with access to training, support, and advice in research data management.
- Providing access to services and facilities for the backup, storage, deposit, and retention of research data that allow researchers to meet their requirements under this policy and those of the funders of their research.
- Providing a central repository for Data Management Plans.
- Providing a metadata registry for research data.
- Providing a research data repository.

LUND UNIVERSITY

[1] http://www.medarbetarwebben.lu.se/sites/medarbetarwebben.lu.se/files/forskningsstrategi-for-lunds-universitet-2017-2021.pdf [Accessed 2017-02-21] [Note: The research strategy is only available in Swedish]

[2] Wilkinson, Mark D., et al. (2016). "The FAIR Guiding Principles for scientific data management and

First version

Lund University policy on management of research data

This research data management policy for Lund University describes the aims of the university in promoting and supporting the highest standard of research data management for adhering to good scientific practice. Lund University will provide researchers with the support needed for research data management to the highest standards. This policy describes necessary aspects of research data management, and the roles and responsibilities of researchers and Lund University respectively in contributing to sound research data management.

Lund University acknowledges the value of research data for the scientific community and society. This is described in the Lund University research strategy 2017-2021, which states that Lund University will "Work proactively with issues concerning research data in digital form and the management of research data, particularly concerning storage and sharing of data" (Research strategy for Lund University 2017-2021, p. 4[1])

Research data are defined as all recorded information (regardless of the form or the media in which they may exist) necessary to support or validate observations, findings or outputs of research.

1. Applicability

This policy applies to researchers and other staff employed at or affiliated with Lund University who manage research data in any form.

Lund University recognises that intellectual property rights as well as ethical and legal regulations can set special requirements on the management of research data. Obligations to other stakeholders may also apply and take precedence over this policy.

2. General principles

Research data produced at Lund University should, whenever possible, comply with the so-called FAIR *Guiding Principles* for scientific data management and stewardship[2]. According to these principles research data should be made findable, accessible, interoperable, and reusable.

3. Ownership of research data

Lund University asserts ownership of research data created or collected as part of research conducted during employment or studies at the University. The University grants the creator or collector of research data rights to use and distribute those data.

4. Data Management Plan

At the start of a research endeavour, a Data Management Plan (DMP) should be set up and registered at Lund University. A DMP is a formal document that describes how research data will be managed both during the research and when it is concluded.

5. Metadata

Research data should be described with rich metadata in order to comply with the *FAIR Guiding Principles*. Metadata should be registered in the Lund University metadata registry and linked to the research data, regardless of where the research data are deposited.

6. Accessibility to research data

Research data should be made openly accessible for use whenever possible, unless prevented by legislation or ethical, contractual, or confidentiality obligations.

Lund University encourages depositing of research data in the commonly used research data repositories within their research community, in the Lund University research data repository, or in another research data repository deemed relevant.

7. Retention of research data

Research data records must be retained in the University Archives in accessible formats for at least 10 years in accordance with the Swedish Archives Act, or for 17 years, for EU funded research.

At the end of the retention period, a decision on the further retention of the research data must be made. Disposal or extended retention is decided on in consultation with the researcher, Lund University Archives, and other expertise on research data management at the University.

8. Roles and responsibilities

Individual researchers are responsible for:

- Complying with all relevant legislation, as well as with ethical, contractual, and confidentiality obligations regarding research data management.
- Setting up a Data Management Plan complying with this policy and registering it at Lund University.
- Registering metadata in the Lund University metadata registry.
- Making research data openly accessible whenever possible while complying with legislation and ethical, contractual, and confidentiality obligations.

Head of the department, where the research is managed, is responsible for

- Naming a delegate who can take over all responsibilities of the research data in case the researcher responsible for the research data leaves Lund University.

Lund University is responsible for:

- Providing researchers with access to training, support, and advice in research data management.
- Providing access to services and facilities for the backup, storage, deposit, and retention of research data that allow researchers to meet their requirements under this policy and those of the funders of their research.
- Providing a central repository for Data Management Plans.
- Providing a metadata registry for research data.
- Providing a research data repository.

LUND UNIVERSITY

[1] http://www.medarbetarwebben.lu.se/sites/medarbetarwebben.lu.se/files/forskningsstrategi-for-lunds-universitet-2017-2021.pdf [Accessed 2017-02-21] [Note: The research strategy is only available in Swedish]

[2] Wilkinson, Mark D., et al. (2016). "The FAIR Guiding Principles for scientific data management and

Revised version, out on consultation

26 June 2018

Lund University policy on research data – PROPOSAL

The strategic plan for Lund University 2017–2026 stipulates that research findings are to be made openly accessible. This is in line with the research strategy bill *Kunskap i samverkan – för samhällets utmaningar och stärkt konkurrenskraft* (*Knowledge in collaboration – for societal challenges and strong competitiveness* prop. 2016/17:50) and with the requirements set by increasing numbers of research funding bodies. Lund University strives for research and research findings to be disseminated efficiently and in compliance with ethical guidelines.

In order to enable continued positive development of research at Lund University, the University's research strategy states that research is to be supported by research infrastructure that is both accessible and fit for purpose. A central part of the University's investment in research infrastructure is to work proactively on issues concerning research data in digital form and the management thereof, in particular with regard to data storage and accessibility. In the long term, open access to research data provides added value in the form of increased quality, utilisation and visibility of research, etc.

This document describes Lund University's policy for the management of research data – its aim, general principles and issues of responsibility.

1. Definition of research data
Research data means all forms of stored information, both data gathered through primary research and obtained secondarily from research sources, that is analysed for research purposes and forms the basis for the results and conclusions generated by research.

2. Objective
The policy is to promote high-quality and appropriate management of research data, by providing guidance to researchers and other employees at Lund University who manage research data.

3. General principles

(a) The FAIR principles
The management of research data should follow the international FAIR principles[1] (*F - Findable, A - Accessible, I - Interoperable, R - Reusable*). According to these principles, research data should be made searchable, accessible, interoperable (i.e. processable), and suitable for further development and reuse for new research endeavours. Lund University strives to make research data and associated metadata searchable, accessible and possible to process for research purposes for a long time, which generally means over several generations of the hardware and software required to process them.

(b) Open access to research data
Lund University works actively to promote open access to research findings and research data, in compliance with both the strategic plan and the research strategy for the University. Research data at Lund University should be made openly accessible for use in future research without inappropriate fees or other requirements, unless there are legal obstacles to open access. Lund University safeguards the researcher's freedom to decide on the time and ways of making research data accessible. However, research data are generally covered by the Public Access to Information and Secrecy Act and are to be released on request.

(c) Processing personal information
The processing of personal information for research purposes is only permitted if there is a legal basis for it. Personal information can, for example, be pseudonymised in the process, which means that the personal details can no longer be attributed to a specific person unless complementary data are used. Complementary data that enables identification, e.g. in case a person whose details are included in a data set requests to be removed from the data processing, are to be stored separately and protected through technical or organisational measures. Research data including personal details are to be reported to the University's data protection officer.

(d) Verification of research findings
Research data are to be accessible so that they can be freely used to support, verify and, if required, defend the methods and results of the research.

(e) Archiving and destruction
Research data are to be stored in a way that facilitates access for as long as it is deemed to be of value for continued research, and are also to be prepared for long-term storage and archiving in compliance with the University guidelines. Research data may only be destroyed if there is a rule allowing for this. Destruction is to follow Lund University's local application decisions on the destruction of research documents.

4. Responsibility
As the principal for research, Lund University has overall responsibility for ensuring that research data are reliable and managed legally, securely, and in alignment with good practice in research. All research projects conducted at Lund University are to appoint a person responsible for research data who ensures compliance with the University's current policy on research data. If the person responsible leaves Lund University or for some other reason is unable to fulfil their task, the responsibility passes to the head of the department where the research project is based, until a new person can be appointed. If research data are handed over to another part of Lund University, the responsibility is also passed on to that organisational unit.

[1] Wilkinson, Mark D., et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". Scientific data 3, Article number 160018. http://doi.org/10.1038/sdata.2016.18

- Focus is more on *why*, connecting to policy documents and strategies
- General principles:
  - FAIR data
  - Open access to research data
  - Processing personal information (e.g. GDPR compliance)
  - Verification of research findings

Guides for different types of data, disciplines and scenarios + best practices and use cases will complement the policy

LUND UNIVERSITY

Lessons learned so far

- The first version was a mix of vision/goals and how-tos. The practical stuff will be moved to guides, e.g.:
  - DMPs – Data management plans
  - Ownership – it's not clear who owns data – different actors have different answers. A national concern.
  - Registering metadata in a catalogue
- The full wish list for support infrastructure (tools, experts) won't be realised before the policy is decided by the vice-chancellor (e.g. longterm storage, archive for digital research data, DMP tool)

LUND UNIVERSITY

Reflections

- What is the role(s) of the policy? Does that reflect the stakeholders' expectations on a policy?
  - Policy as a tool for setting a common direction for LU. Existing initiatives and dedication are extremely important – these complement each other
- Consultation process important to collect views and connect with the broad range of stakeholders at LU
  - Sent out in Swedish and English
- Guides, examples of best practice RDM and use cases will complemen the policy
- Important to update the policy regularly (or check that it's up-to-date) to keep up with the rapid development in open science

LUND UNIVERSITY