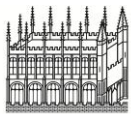


# A brief overview of metadata for datasets

Sally Rumsey

The Bodleian Libraries

University of Oxford



Bodleian Libraries  
UNIVERSITY OF OXFORD

# Structure

- Citation
- Discovery
- Documentation and re-use
- Timing of metadata creation
- WIP at Oxford

# 1. Citation: Why?

- Attribution
  - Recognition
  - Reward
  - Impact
- Discovery
  - Referencing
  - Avoid repetition
  - Potential collaboration
- Location
  - For access and re-use
  - Includes non-digital data
- Access
  - Verification
- Re-use
  - ‘Standing on the shoulders of giants’
  - Documentation of methods and descriptions of data
- Administration

# Citation: Key documents/principles

- **DataCite** organisation <https://www.datacite.org/>
- **Force11** Joint Declaration of Data Citation Principles. 8 principles covering cover purpose, function and attributes of citations. <https://www.force11.org/datacitation>
- **DCC** How to Cite Datasets and Link to Publications <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>
- **Out of cite, out of mind:** The Current State of Practice, Policy, and Technology for the Citation of Data. CODATA-ICSTI Task Group on Data Citation Standards and Practices, Yvonne M. Socha (Ed). CODATA *Data Science Journal*. 13 September 2013 [https://www.jstage.jst.go.jp/article/dsj/12/0/12\\_OSOM13-043/ article](https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/article)

# Basic citation: DataCite metadata kernel

## 1.3 The Metadata Schema

The DataCite Metadata Schema is a list of core metadata properties chosen for the accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. The resource that is being identified can be of any kind, but it is typically a dataset.

**Table 1: DataCite Mandatory Properties**

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
1	Identifier (with type sub-property)	M
2	Creator (with name identifier sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M

# Minimum reference

Creator (PublicationYear): Title. Publisher. Identifier

# Deluxe reference

Creator (PublicationYear): Title. Version.  
Publisher. ResourceType. Identifier

## 2. Discovery

- In addition to 5 mandatory elements, how might an interested person search for a dataset?
- May include (not exhaustive):
  - Subject/keywords
  - Resource type
  - Description text/abstract
  - GeoLocation
  - Methodology



# 3. Documentation & Re-use

**EPSRC**

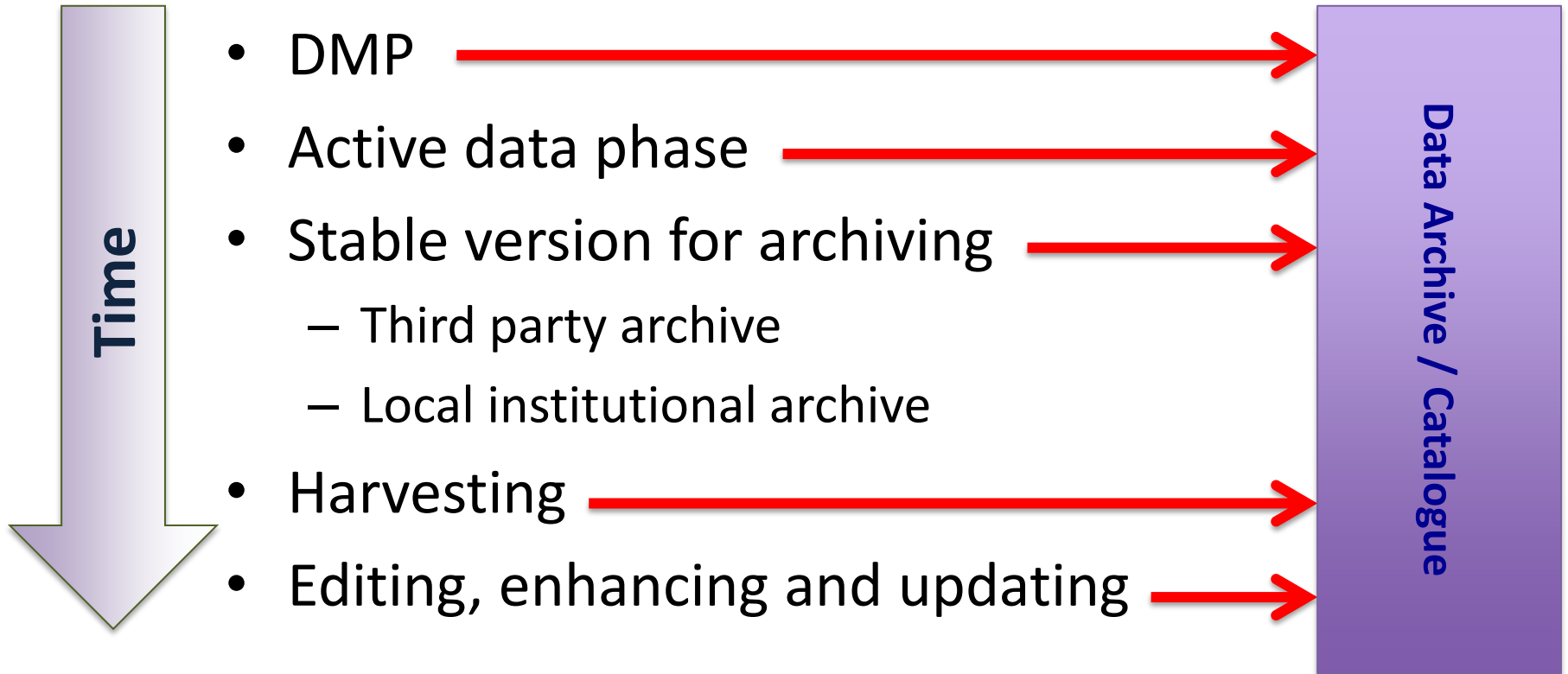
Engineering and Physical Sciences  
Research Council

EPSRC policy framework on  
research data: Principles

<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/principles.aspx>

“**Sufficient metadata** should be recorded and made **openly available** to enable other researchers to **understand** the potential for **further research** and **re-use** of the data. Published results should always include information on **how to access** the supporting data.”

# When are the metadata created?



# WIP at Oxford

- Library perspective: data archiving and cataloguing
- ORA-Data branding (Oxford University Research Archive)
- Phase 1: Basic service
- Hydra implementation <http://projecthydra.org/>
  - Work completed under JISC Damaro project
  - Checking all elements/fields
  - Consistency with other item types
- Metadata creation/harvest
  - Expect mixed model
  - Low entry barrier
  - Training/advocacy on benefits of rich metadata
- Integration with CUD (Core User Directory)
- FAST subject headings
- Work in parallel to service set-up
- Funder compliance as main driver for University

