

Metadata

The Spirit of Research Data Management?

Alex Ball

2017-06-29

University of Bath

Data is the new oil

It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.

. . . we must understand that for data to mean something, it must be placed in context to the marketplace.

— Clive Humby, 2006

Data is the new water

Applications draw up data from ground like water

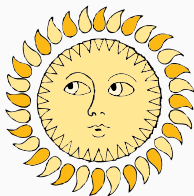


RDA is the trunk that brings the two together

Technologies make data discoverable and accessible

— Juan Bicarregui, 2016

Data is the new light



- Data is not consumed: it can be reused endlessly.
- Data yields more value as it diffuses and spreads around.

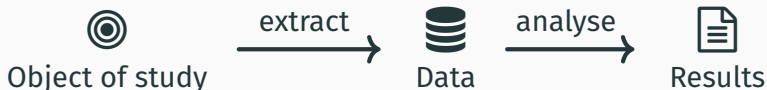
It is like opening the curtains – the more you open, the more is unveiled.

— Axelle Lemaire, 2015

Metadata

What is metadata?

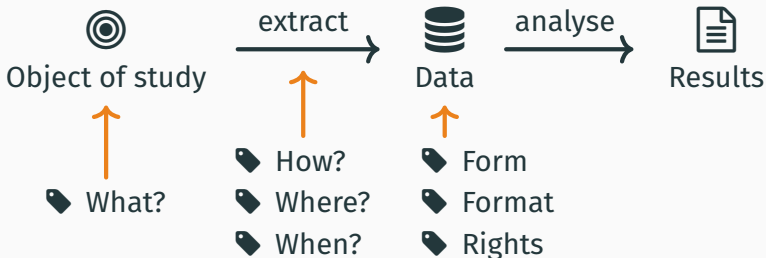
- Literally 'data about data'
- Information that helps you work with other information



- Context determines whether something is **data** or **metadata**

What is metadata?

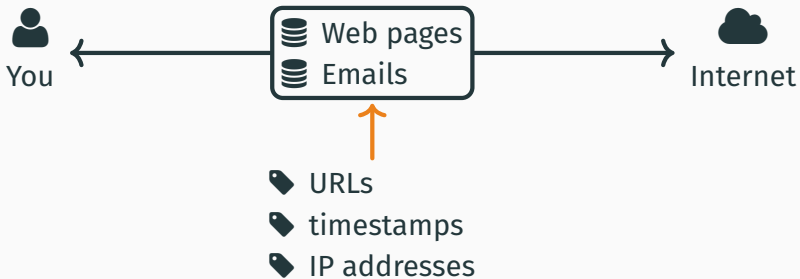
- Literally 'data about data'
- Information that helps you work with other information



- Context determines whether something is **data** or **metadata**

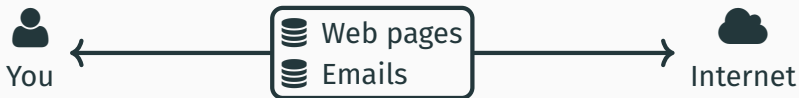
Example: Internet traffic

Your perspective:



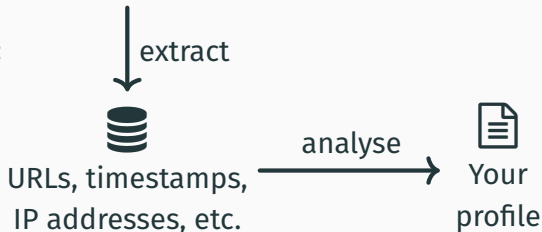
Example: Internet traffic

Your perspective:



- ↑
- 🏷️ URLs
- 🏷️ timestamps
- 🏷️ IP addresses

Spy's perspective:



Types of metadata

Metadata is defined by what you are using it to achieve:

Reference Identifying, citing, searching for a known resource

Discovery Speculative searching

Provenance Assessing authenticity or trustworthiness

Contextual Relating data to other resources, agents, activities

Rights Securing data against unauthorized/illegal actions

Packaging Arranging components of a resource

Fixity Checking integrity

Structural Loading/opening a file

Semantic Unlocking the meaning of a resource

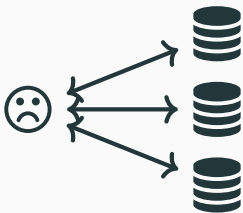
Types of metadata

In the research context, we are mostly concerned with

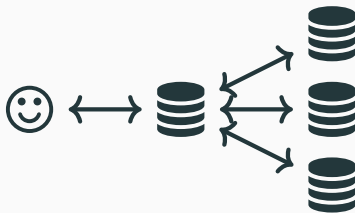
- **Discovery metadata** – help other researchers find the data, and give credit for them → impact
- **Contextual metadata** – keeping the institution and funder happy, conveying quality and relevance
- **Structural & semantic metadata** – ensure that researchers can understand and use/reuse the data

Why should I use a metadata standard?

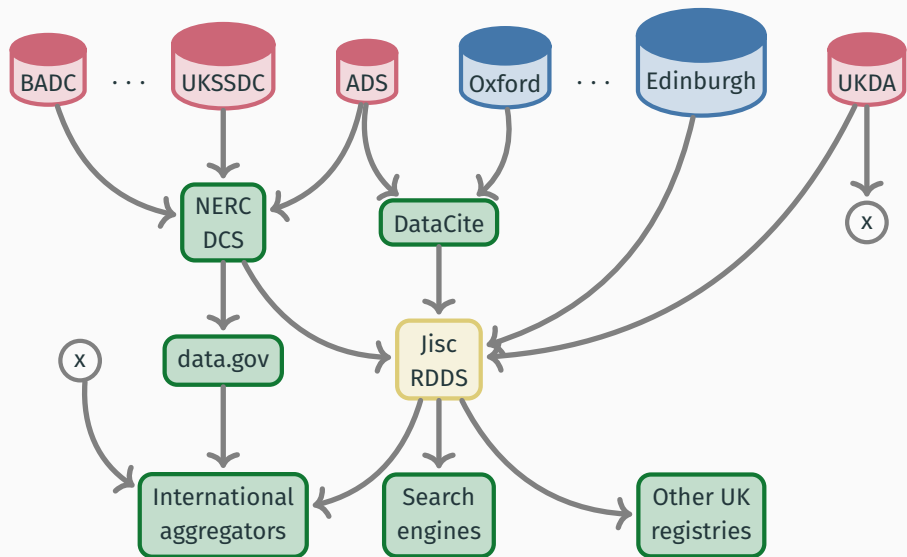
Better discovery



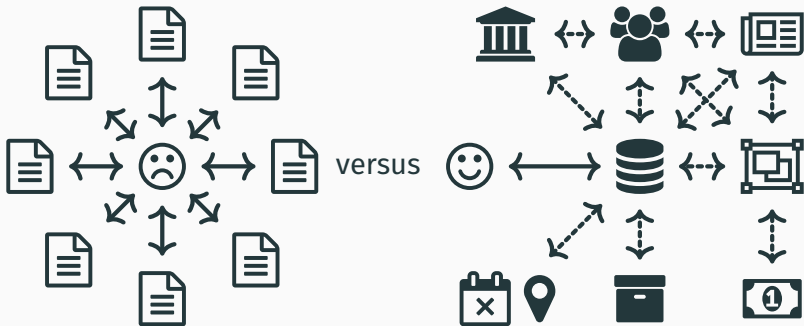
versus



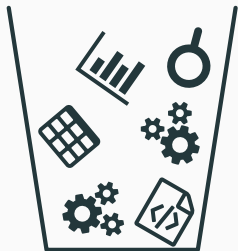
Metadata flows for data discovery



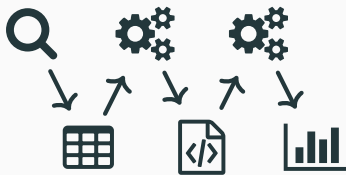
Better context



Better reuse



versus

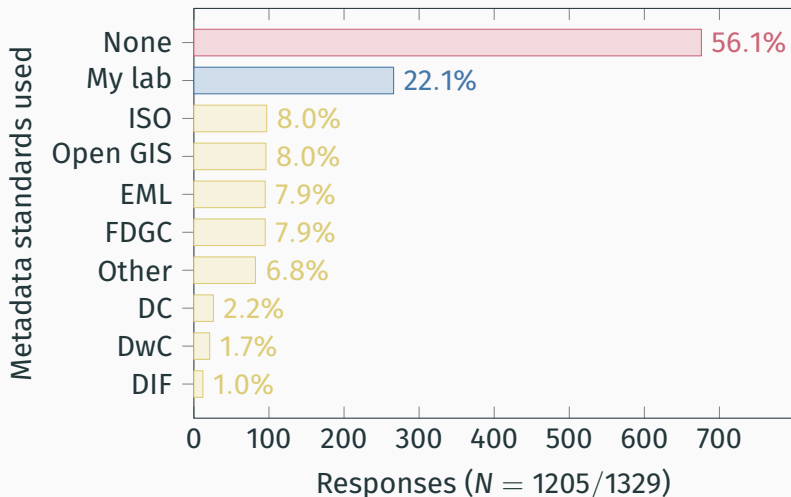


Better ecosystem

- ☰ Less working things out from scratch
- 🔑 More complete metadata
- 📈 Benefits of practising
- 📄 Better documentation of the standards
- 🐛 Concentration of development attention and effort
- 🚀 Better time-saving tools
- » etc., etc.

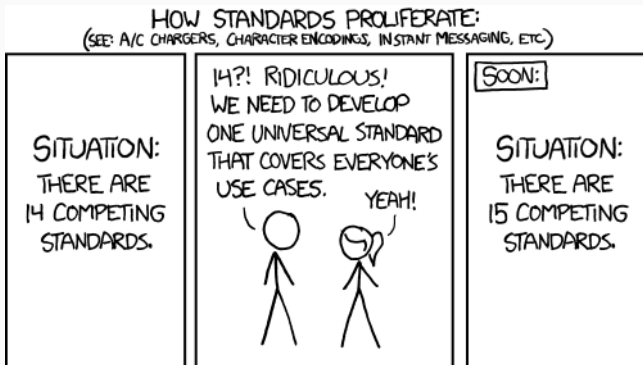
**So why doesn't everyone use a
metadata standard?**

No suitable standard?



Source: Tenopir, C. et al. (2011), 'Data Sharing by Scientists: Practices and Perceptions', *PLoS ONE* 6/6: e21101. doi: 10.1371/journal.pone.0021101

Too many standards?



Source: © ⓘ \$ Randall Munroe

‘The nice thing about standards is that you have so many to choose from’ — Tanenbaum (1988)

Isn't that, like, really *hard*?

Just fill out this simple form . . .

```
<mods xmlns="http://www.loc.gov/mods/v3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-4.xsd"> <titleInfo> <title> Title goes
here </title> </titleInfo> <name type="personal"> <namePart>Author name goes
here</namePart> <role> <roleTerm type="text">Author</roleTerm> </role> </name>
<typeOfResource>dataset</typeOfResource> <genre>Dataset</genre> <originInfo>
<publisher>Publisher name goes here</publisher> </originInfo> <language>
<languageTerm type="text">Language name</languageTerm> <languageTerm type="code"
authority="iso639-2b">ISO 639-2b code</languageTerm> </language>
<physicalDescription> <internetMediaType>MIME type goes here, repeat as
necessary</internetMediaType> <digitalOrigin>born digital</digitalOrigin>
<extent>Number of records in your database, or size of file in bytes</extent>
</physicalDescription> <abstract> Abstract goes here </abstract> <subject
authority="scheme name goes here"> <topic>Keyword goes here, repeat as
necessary</topic> <cartographics>Spatial coordinates</cartographics/>
<temporal>Temporal extent</temporal> <geographic>Spatial extent in
words</geographic> </subject> <identifier>ID goes here</identifier> <location> <url
usage="primary display" access="object in context">Location of record</url> <url
access="raw object">Location for download</url> </location> <accessCondition
type="useAndReproduction"> Usage restrictions or permissions </accessCondition>
<relatedItem ID="relatedMaterials"> <location> <url usage="primary display"
access="object in context">Record of related item</url> </location> </relatedItem>
<note type="citation"> Sample citation goes here </note> <note
type="software">Required software goes here</note> <subject ID="location"
displayLabel="Description of spatial extent again"> <cartographics> <coordinates>
List of coordinates, comma separated </coordinates> </cartographics> <topic>Type of
coordinates goes here</topic> </subject> </mods>
```

Metadata Standards Catalog

Key facts

- Ran 1 August 2013 – 1 February 2015
- 150 members from many countries and disciplines

Goals

1. Develop an **RDA Metadata Standards Directory** listing standards relevant for research data
 - Comprehensive
 - Easy for anyone to contribute or update
2. Define and develop **use cases** for research metadata
3. Develop plan for long-term growth and maintenance of the directory

The Metadata Standards Directory

| D | C | C Disciplinary Metadata

Search by Discipline



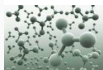
Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

Search by Resource Type

Metadata Standards

Specifications for the minimum information that should be collected about research data in order for it to be re-used.

Profiles and Extensions

Standards that have been adapted for use in particular types of repositories, or for particular types of data.

Use cases

Institutional repositories and data portals using standards to determine which metadata should be collected upon data deposit.

Tools

Software that has been developed to capture or store metadata conforming to a specific standard.

RDA Metadata Standards Directory

Metadata

RDA | Metadata Directory

[View the standards](#)

[View the extensions](#)

[View the tools](#)

[View the use cases](#)

[Browse by subject areas](#)

[Contribute](#)

[Add standards](#)

[Add extensions](#)

[Add tools](#)

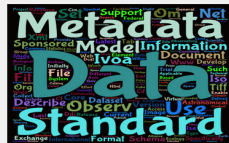
[Add use cases](#)

 [github](#)

 [@twitter](#)

 [linkedin](#)

 [facebook](#)



[Metadata Standards Directory Working Group](#)

The RDA Metadata Standards Directory Working Group is supported by individuals and organizations involved in the development, implementation, and use of metadata for scientific data. The overriding goal is to develop a collaborative, open directory of metadata standards applicable to scientific data can help address infrastructure challenges.

The RDA Metadata Standards Directory is maintained by [Sean Chen](#), [Kate Anne Alderete](#), and [Alex Bas](#).
The theme is maintained by [Dustin Allen](#).
This page was generated by [GitHub Pages](#).

<http://www.dcc.ac.uk/resources/metadata-standards>

<http://rd-alliance.github.io/metadata-directory/>

Duration

- Running January 2016 – July 2017

Goal: develop the RDA Metadata Standards Directory into a **RDA Metadata Standards Catalog**:

- Enhance the **set of records** describing metadata standards
- Develop a new **user interface** for submitting information, searching, browsing and displaying standards information
- Provide a new **API** allowing automated tools to submit information, perform queries and retrieve information

Is this the right one
for me?

How do I use it?

How do I refer to
it/find it again?

Can I convert existing
metadata to it? Will I
be locked in?

DDI (Data Documentation Initiative)

A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel:

- DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2012.
- DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.

Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.

Used in

Demography

Economics

Health policy

History

Human geography

Land use

Law

Politics

Social policy

Sociology

Statistics

Documentation

[View specification](#)

[Visit website](#)

Responsible organizations

- Maintainer: [DDI Alliance](#)

[View website](#)

The Metadata Standards Catalog

Metadata Standards Catalog Search Sign in

Search the Catalog

Search for schemes that match any of the following criteria:

Name of scheme

Subject area

Identifier

[github](#) [@twitter](#) [linkedin](#) [facebook](#)

The Metadata Standards Catalog

Metadata Standards Catalog

Search the Catalog

Search for schemes that match any of the following criteria

Name of scheme

Subject area

Identifier

[github](#) [@twitter](#) [linkedin](#)

Add new tool

Name of tool

Description

- Please provide a short description of the intended use of the tool, and its capabilities.

If the scheme you need is not listed, you can add it using the following link.

Metadata scheme(s) supported by this tool

ABCD (Access to Biological Collection Data)
ABCDDNA
ABCDEFG (Access to Biological Collection Databases Extended for Geosciences)
AGLS Metadata Profile

- Select a scheme if the tool accepts metadata conforming to it as an input, or emits metadata conforming to it as an output.

Type of tool

1. Type

Call to action

Metadata: the Spirit of Research Data Management

Metadata

- brings data to **life**
- **powers** local systems
- gives data a **presence** beyond its present location
- **breaks down barriers** to data reuse and sharing

What you can do

Research institutions should

- **use standards** and profiles in their own systems
- **participate** in national and global research data infrastructures
- **encourage researchers** to use standards wherever possible
- support their researchers with **training, tools and guidance**

DataCite Metadata Schema v4.0

Mandatory elements

- Creator
- Title
- Publication year
- Publisher
- Identifier
- Resource type

Recommended elements

- Subject, Description
- Contributor (with type, affiliation)
- Date (with type)
- Geo-location
- Related identifiers

Optional elements

- Alternate identifier
- Format, Version, Size
- Rights, Language
- Funding reference

Metadata → better data

- Even bad documentation is better than nothing
- The more **structure**, the better
 - Clear headings and sections in documentation
 - Consistent metadata
- Look for **metadata standards** you can use
 - Metadata Standards Directory/Catalog
- Not an exact fit? Create a local **profile**
 - Avoid completely bespoke schemes
- Be **consistent**

Thank you for listening

Grazie per l'attenzione

Any questions?

References (1)

-  Bicarregui, J. (2016), *Building and Sustaining Data Infrastructures: Putting Policy into Practice*, <https://dx.doi.org/10.6084/m9.figshare.4055538>.
-  DataCite Metadata Working Group (2016), *DataCite Metadata Schema for the Publication and Citation of Research Data*, version 4.0 (DataCite e.V.). doi: 10.5438/0012.
-  Jakobsson, A. (2015), 'Data is Light: Observations from the 6th RDA Plenary', (29 Sept.), <https://www.rd-alliance.org/blogs/data-light-%E2%80%93-observations-6th-rda-plenary.html>.

References (2)

-  Palmer, M. (2006), 'Data is the new oil', ANA Marketing Maestros (3 Nov.), http://ana.blogs.com/maestros/2006/11/data_is_the_new.html.
-  Tanenbaum, A. S. (1988), *Computer Networks*, (2nd edn., Upper Saddle River, NJ: Prentice-Hall).
-  Tenopir, C. et al. (2011), 'Data Sharing by Scientists: Practices and Perceptions', *PLoS ONE* 6/6: e21101. doi: 10.1371/journal.pone.0021101.