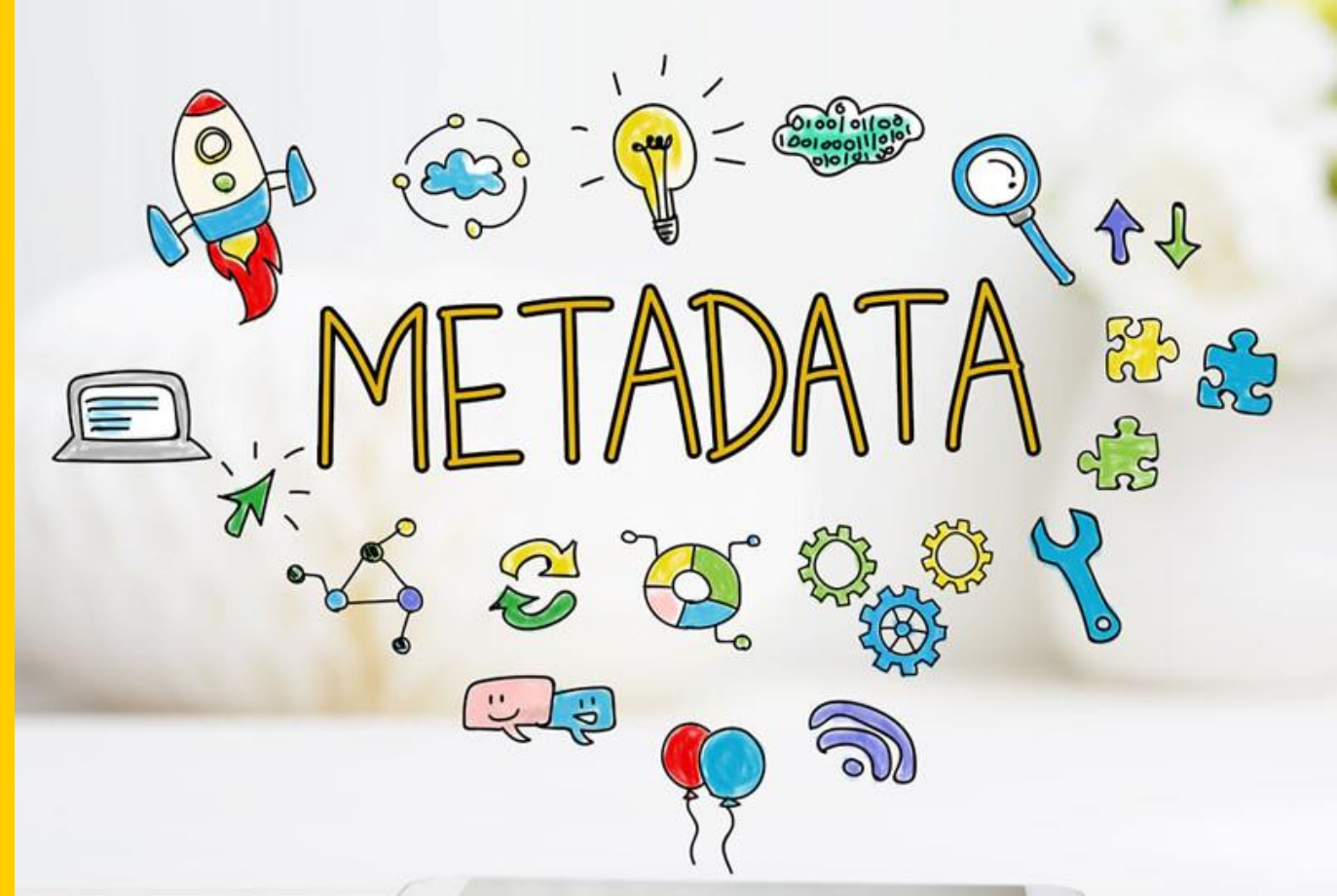


Metadata in practice



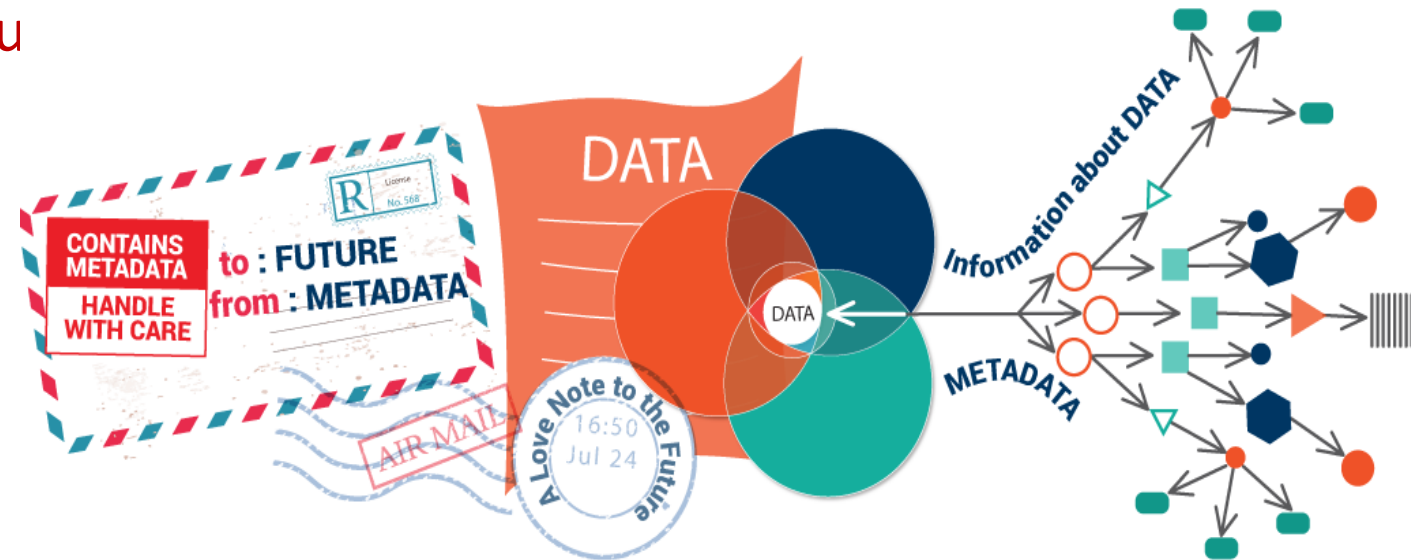
Maisam M. Dadkan

Garrett Speed

Geo Data Team (GDT)
Faculty of GeoScience

Metadata is...

Metadata: {
Data reporting
Structured information about data
A love note to the future about



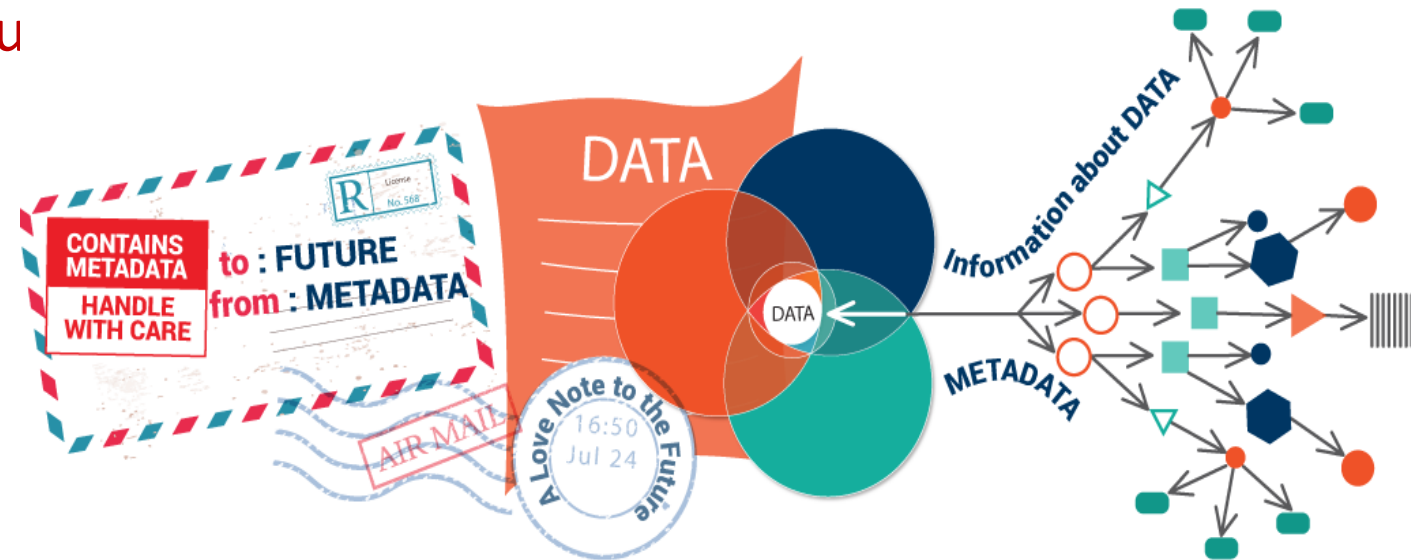
<https://www.ontotext.com/>

Metadata is...

Metadata: {

- Data reporting
- Structured information about data
- A love note to the future about

- When you provide data to someone else, what types of information would you want to include with the data to make them useful?
- When you receive a dataset from an external project/researcher, what types of details do you want to know about the data to use it?



<https://www.ontotext.com/>

Metadata is valuable

Your data with(out) metadata

Bad

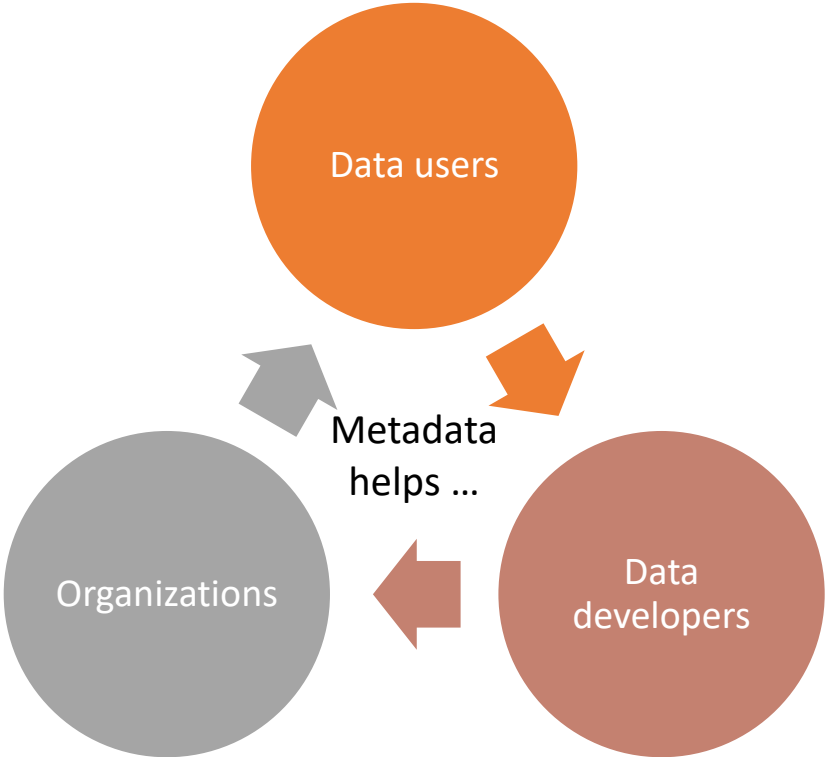


Better



The best (Rich & structured)

Nutrition Facts	
8 servings per container	
Serving size	2/3 cup (55g)
Amount per serving	
Calories	230
% Daily Value*	
Total Fat 8g	10%
Saturated Fat 1g	5%
Trans Fat 0g	
Cholesterol 0mg	0%
Sodium 160g	7%
Total Carbohydrate 37g	13%
Dietary Fiber 4g	14%
Total Sugars 12g	
Includes 10g Added Sugars	20%
Protein 3g	
Vitamin D 2mcg	10%
Calcium 260mg	20%
Iron 8mg	45%
Potassium 240mg	6%
*The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.	



Metadata is valuable

Your data with(out) metadata

Bad

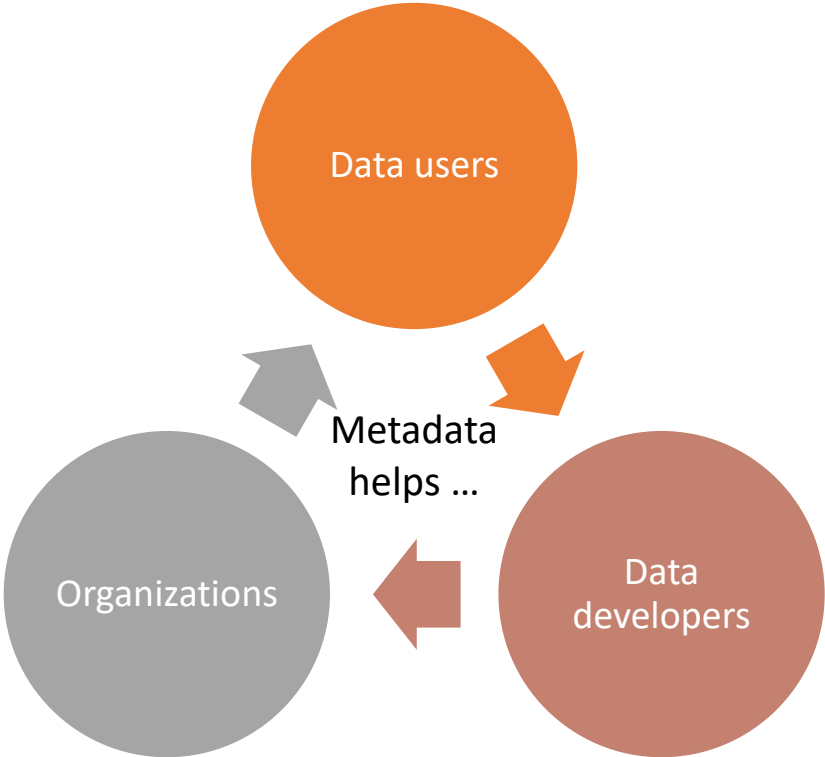


Better



The best (Rich & structured)

Nutrition Facts	
8 servings per container	
Serving size	2/3 cup (55g)
Amount per serving	
Calories	230
% Daily Value*	
Total Fat 8g	10%
Saturated Fat 1g	5%
Trans Fat 0g	
Cholesterol 0mg	0%
Sodium 160g	7%
Total Carbohydrate 37g	13%
Dietary Fiber 4g	14%
Total Sugars 12g	
Includes 10g Added Sugars	20%
Protein 3g	
Vitamin D 2mcg	10%
Calcium 260mg	20%
Iron 8mg	45%
Potassium 240mg	6%
*The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.	



Metadata for Machine (M4M)



Metadata by example



Metadata by example

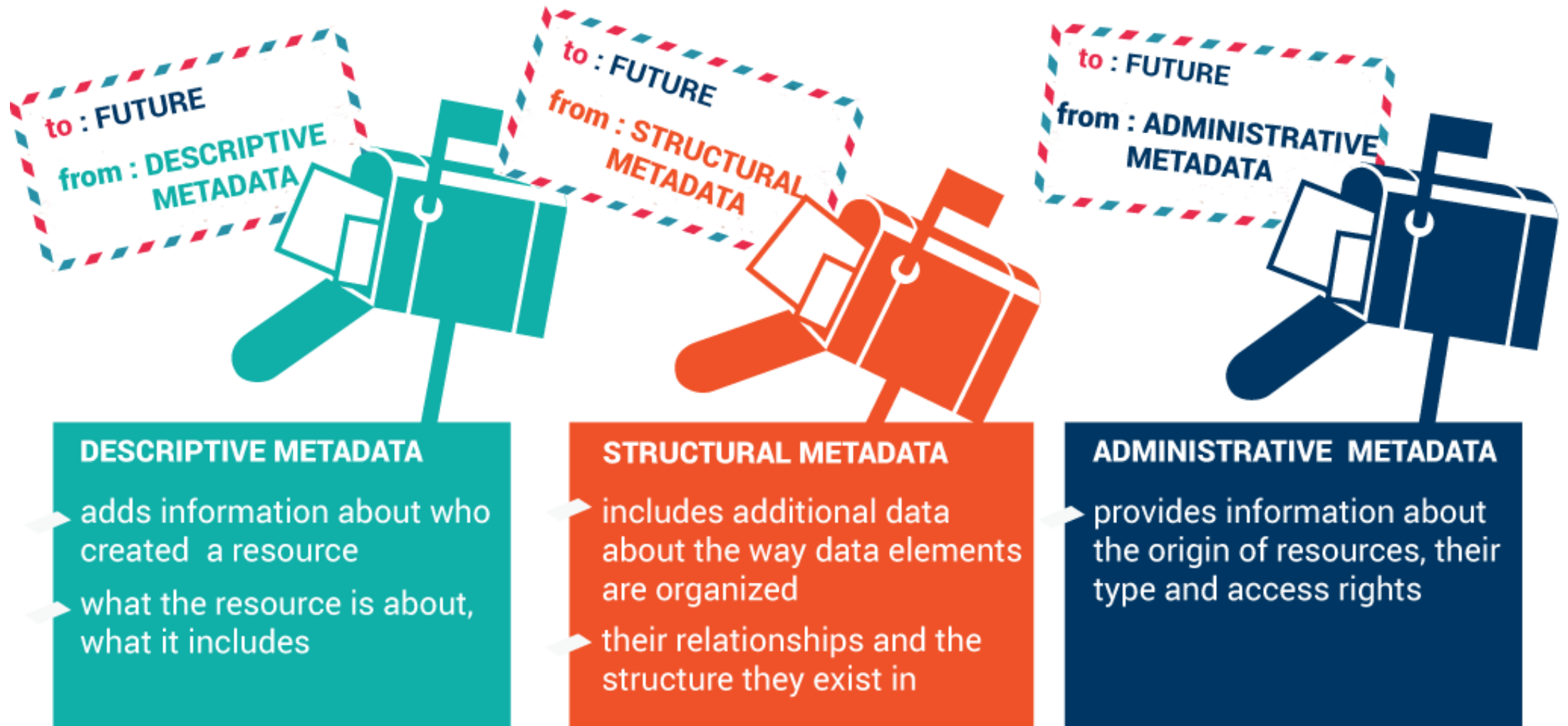
- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** were the data created?
- **WHERE** is it geographically?
- **HOW** were the data developed?
- **WHY** were the data developed?

For example:

- Title, author, subject, date, type, coordinates, ...



Metadata types



<https://www.ontotext.com/>

Metadata & FAIR

Findable

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specifies the data identifier.

Accessible

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2. metadata are accessible, even when the data are no longer available.

Interoperable

- I1. (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

Reusable

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

**All research data must be
FAIR**

Metadata & FAIR

Findable

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specifies the data identifier.

Accessible

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2. metadata are accessible, even when the data are no longer available.

Interoperable

- I1. (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

Reusable

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

**All research data must be
FAIR**

**FAIR applies equally to
data and metadata!**

Metadata standards



Metadata standards

...Everyone thinks that it is a good idea,
but nobody wants to use someone else's.



Metadata and Vocabularies

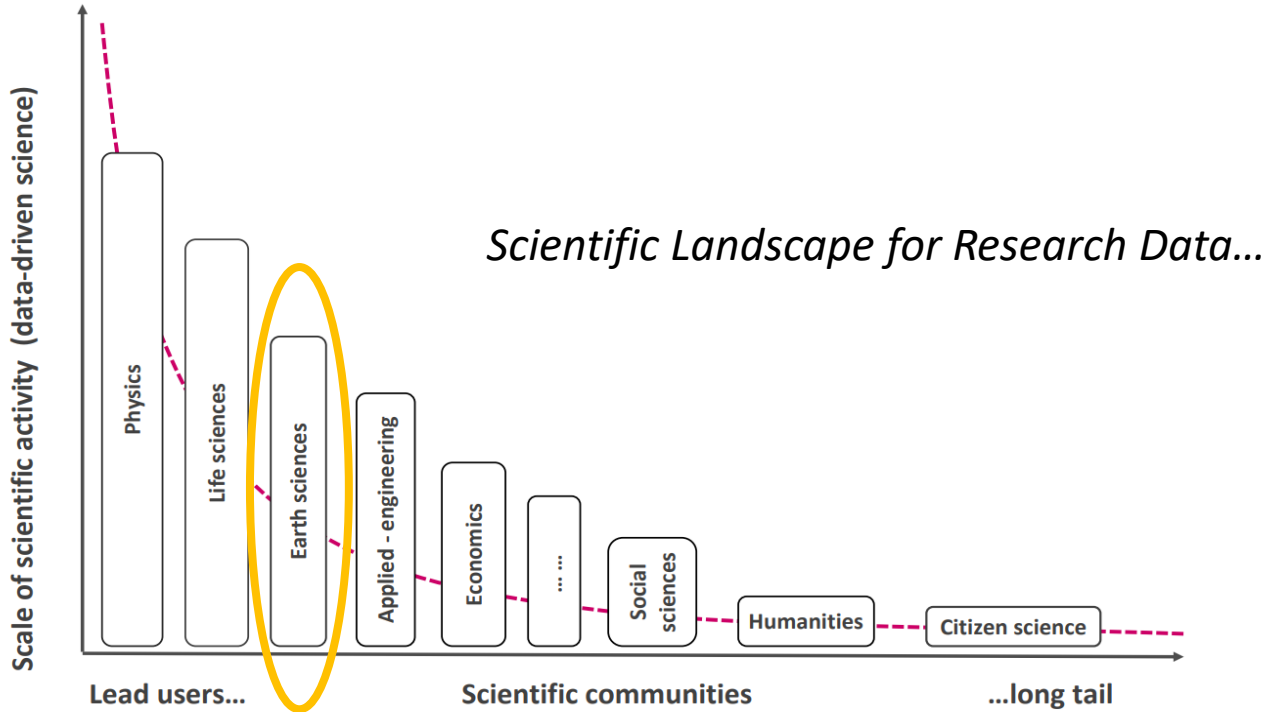
One size does not fit all!!!

Metadata standards

...Everyone thinks that it is a good idea, but nobody wants to use someone else's.



Metadata and Vocabularies
One size does not fit all!!!



Metadata standard (schema)

A metadata standard or schema is a:

Labeling,

Tagging,

or coding system,

used to record cataloging information or
structure descriptive records.

Metadata standard (schema)

A metadata standard or schema is a:

Labeling,

Tagging,

or coding system,

used to record cataloging information or
structure descriptive records.

General

- Dublin Core
- MODS (Metadata Object Description Schema)
- DataCite
- ...

Science schema

- | | |
|--------------------------------------|---------------|
| • Darwin Core; | Biology |
| • Ecological Metadata Language (EML) | Ecology |
| • NASA's Standards | Earth Science |
| • ... | |

Social Science schema

- DDI (Data Documentation Initiative Alliance)
- OLAC (Open Language Archives Community)
- ...

Humanities Schema

- VRA Core (Visual Resources Association)
- Cataloging Cultural Objects (CCO)
- ...

Find a full list at <https://www.dcc.ac.uk/guidance/standards/metadata/list>

Metadata standard (schema)

Sample Dublin Core metadata for an image

Below is an example of Dublin Core metadata for this photograph of Howard Walter Florey. [Generic examples](#) for the Dublin Core metadata schema are available from the [Dublin Core website](#).



- **Title** = Howard Walter Florey
- **Author/Creator** = Not known
- **Subject and Keywords** = The Australian National University photographs
- **Subject and Keywords** = Nobel Laureates
- **Description** = This photograph was taken at the 1948 Easter Conference, which considered the early planning of the Australian National University, at the Institute of Anatomy building. A keen amateur photographer, Florey carries his camera over his shoulder.
- **Publisher** = The Australian National University Digital Collections
- **Date** = 1948
- **Type** = image
- **Format** = image/tif
- **Resource identifier** = <http://hdl.handle.net/1885/7491>
- **Resource identifier** = ANUA225-405-3
- **Source** = ANU Archives
- **Rights** = This image is provided for research purposes only and must not be reproduced without the prior permission of the Archives Program, Australian National University

General

- Dublin Core
- MODS (Metadata Object Description Schema)
- DataCite
- ...

Science schema

- Darwin Core;
 - Ecological Metadata Language (EML)
 - NASA's Standards
 - ...
- Biology
Ecology
Earth Science

Social Science schema

- DDI (Data Documentation Initiative Alliance)
- OLAC (Open Language Archives Community)
- ...

Humanities Schema

- VRA Core (Visual Resources Association)
- Cataloging Cultural Objects (CCO)
- ...

Metadata standard choosing

Your choice of repository may determine your standard

- Repositories often require use of a particular standard.

There are many types of metadata standards/schemas

- Some are generic, while others are domain-specific. Generic ones often need to be expanded in order to cover more specific information. **Domain-specific schemas have a much richer vocabulary** and structure but tend to be highly specialized and only understandable by researchers in that area.

Always keep the user's perspective in mind

- Pick a scheme that is going to make sense for those who are most likely to access and use your data, as well as those managing and preserving your data.

Adopt or Adapt?

- Generally, you should be able to find a metadata standard to suit your needs. When you find one, use it. If you find one that is close to your needs, but not quite, you can extend or shorten it to suit your needs. This may require communicating with the organization that created the standard.

Metadata formats

Humane-readable



- Text file (.txt)
- Markdown (.md)
- Comma-separated Values (.csv)
- PDF
- ...

Machine-readable



- JSON
`{key:value}`
- RDF/XML
`<tag>element<tag>`
- html5
- ...

Metadata best practices

**Think about metadata
at the start of data
collection**

**Decide at the
beginning of the
project where you
publish data**

**Use domain-specific
metadata schema
where possible**

**Use standard or
controlled vocabulary
(See Linked Open
Vocabularies)**

**Provide both
machine- & human-
readable formats**

**Include all types of
metadata (descriptive,
structural,
administrative)**

**Document it, if you
make your own or
extend a standard one**



Demo & Q&A

Contact us via:

datateam.geo@uu.nl