

Research Data Management: Using Metadata to Find, Interpret & Share Your Data

Today's session:

- **What is metadata?**
- How do you document your data?

What do we mean by *data*?

General	Social Sciences	Hard Sciences
<ul style="list-style-type: none">• images• video• mapping/GIS data• numerical measurements	<ul style="list-style-type: none">• survey responses• focus group and individual interviews• economic indicators• demographics• opinion polling	<ul style="list-style-type: none">• measurements generated by sensors/laboratory instruments• computer modeling• simulations• observations and/or field studies• specimen

What is metadata?

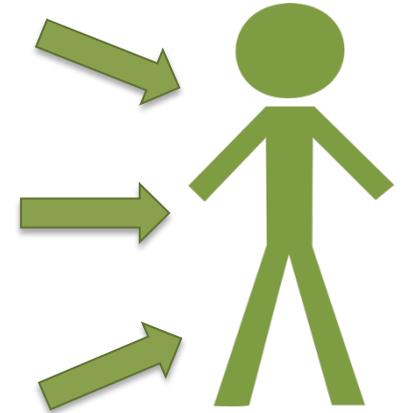
*“Metadata is structured information that describes, explains, locates, or otherwise makes it **easier to retrieve, use, or manage an information resource.** Metadata is often called data about data or information about information”*

– NISO, Understanding Metadata 2004;1

Why is metadata important?

Metadata enables you:

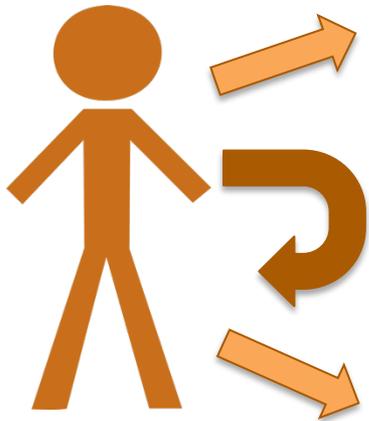
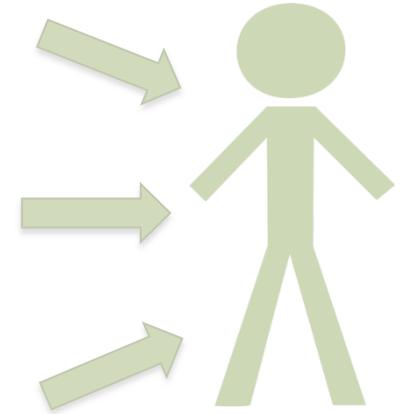
- to find data from other researchers
- to use the data that you do find



Why is metadata important?

Metadata enables you:

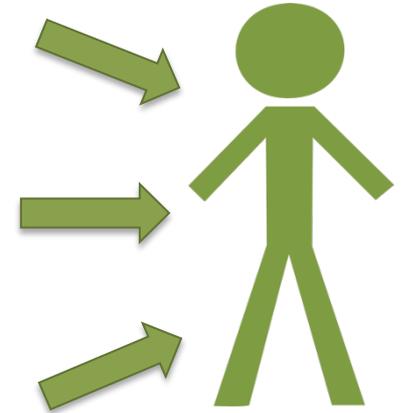
- to find data from other researchers
- to use the data that you do find



- to help others to find and use you data
- to use your own data in the future

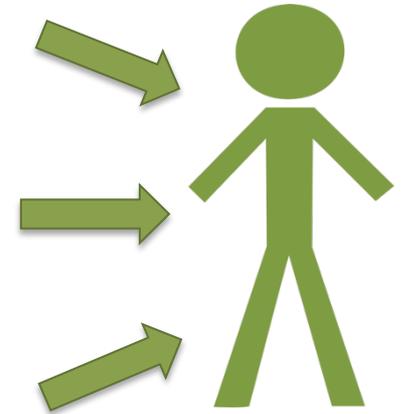
What is metadata?

When you **receive** a dataset from an external source, what types of details do you want to know about the data?

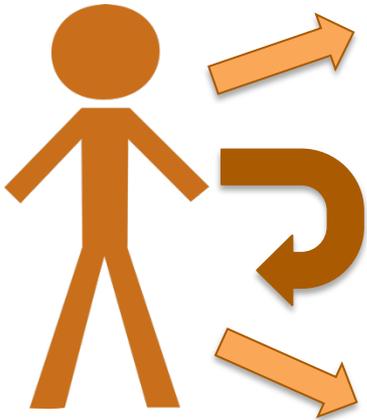


When receiving data...

- What are the data **gaps**?
- What **processes** were used for creating the data?
- Are there any **fees** associated with the data?
- In what **scale** were the data created?
- What do the values in the tables **mean**?
- What **software** do I need in order to read the data?
- Can I **give** these data to someone else?



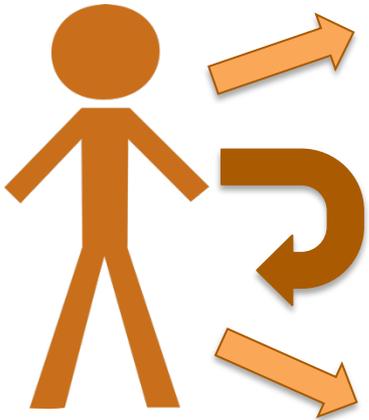
What is metadata?



When you ***provide*** data to someone else, what types of information would you want to include with the data?

When providing data...

- **Why** were the data created?
- What **limitations**, if any, do the data have?
- What does the data **mean**?
- How should the data be **cited**?



Metadata Types



Descriptive



Structural



Administrative

Documenting your data...

- In a filename
- In a readme file
- In a spreadsheet
- In an XML file
- Into a database

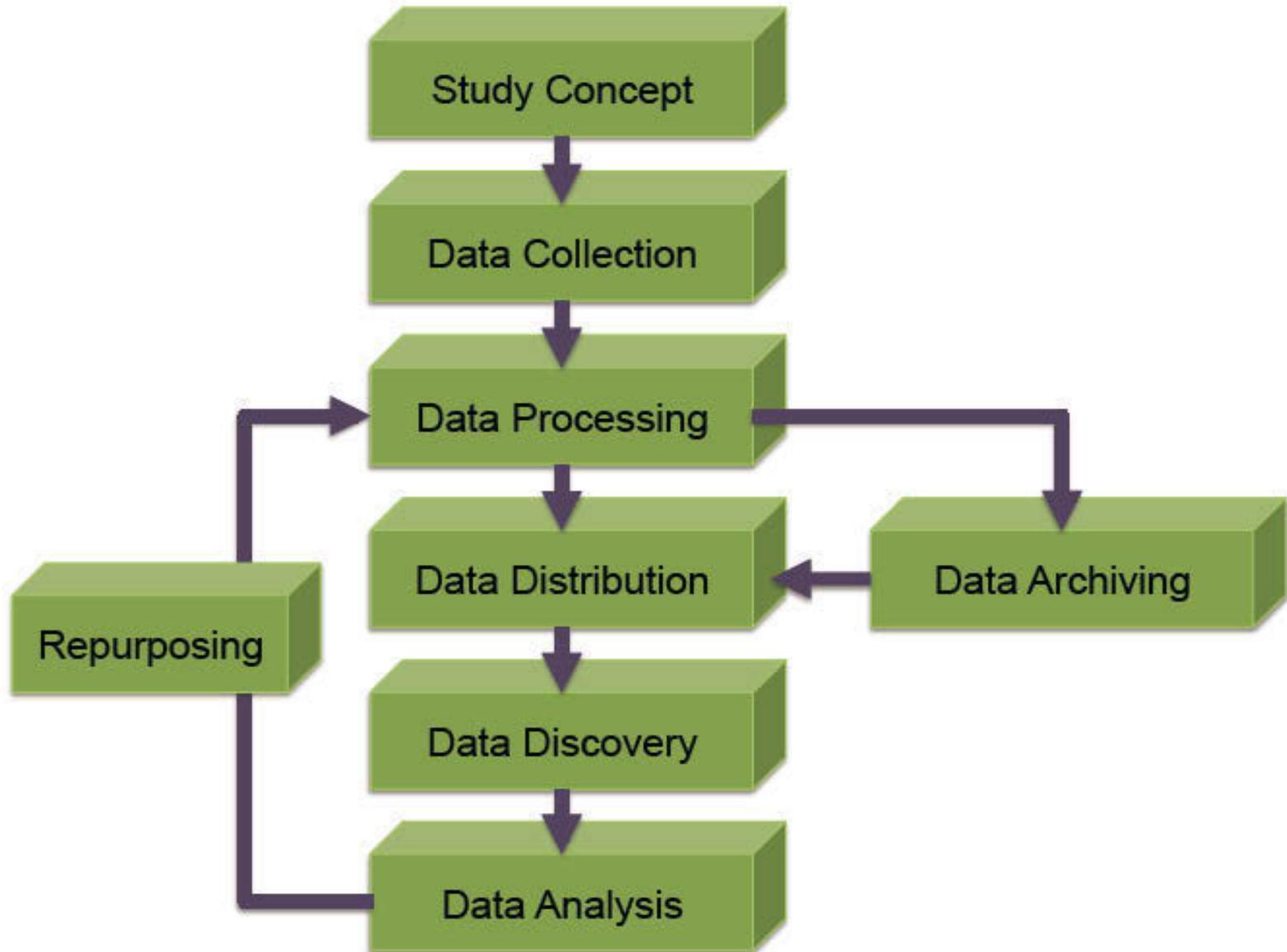
Today's session:

- What is metadata?
- **Documenting your data?**

When?

The earlier the better!

When?



Adapted from DDI version 3.0 Combined Life Cycle Model

My Research (not really)

Squirrel Avoidance Project



Photo Courtesy of [Nickmargolies](#) on Wikimedia. License CC-BY.

Metadata Considerations

- 1. What type of data are you producing?**
2. What are your metadata resources?
3. What metadata elements do you need to capture?
4. How are you going to ensure consistent data entry?

1: What data will you be producing?



Site	DateTime	Sp	PA
MIT	20140115T0930	S1	1
MIT	20140215T0930	S1	0
MIT	20140315T1130	S2	2
MIT	20140315T1130	S1	4
HUM	20140115T0130	S1	1
HUM	20140215T0130	S1	1
HUM	20140315T0130	S2	2
HUM	20140315T0130	S1	1

Metadata Considerations

1. What type of data are you producing?
- 2. What are your metadata resources?**
3. What metadata elements do you need to capture?
4. How are you going to ensure consistent data entry?

2: Metadata standards

- Darwin Core
- DDI – Data Documentation Initiative
- FGDC – Federal Geographic Data Committee
- Dublin Core
- ...

Find your metadata standard:

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

2: Metadata resources

- ~ Auto-generated data
- ✓ External metadata
- ✓ Metadata standards
 - Dublin Core
 - Darwin Core



**Squirrel
Avoidance
Project**

Metadata Considerations

1. What type of data are you producing?
2. What are your metadata resources?
- 3. What metadata elements do you need to capture?**
4. How are you going to ensure consistent data entry?

3: Metadata Elements



Descriptive



Structural



Administrative

Site	DateTime	Sp	PA
MIT	20140115T0930	S1	1
MIT	20140215T0930	S1	0

Site	DateTime	Site	DateTime	Sp	PA
MIT	20140315	MIT	20140115T0930	S1	1
MIT	20140315	MIT	20140215T0930	S1	0

Site	DateTime	Sp	PA
HUM		S2	2
HUM		S1	4
HUM		S1	1
MIT	20140115T0930	S1	1
MIT	20140215T0930	S1	0
MIT	20140315T1130	S2	2
MIT	20140315T1130	S1	4
HUM	20140115T0130	S1	1
HUM	20140215T0130	S1	1
HUM	20140315T0130	S2	2
HUM	20140315T0130	S1	1

S2	2
S1	4
S1	1
S1	1
S2	2
S1	1

Metadata Considerations

1. What type of data are you producing?
2. What are your metadata resources?
3. What metadata elements do you need to capture?
- 4. How are you going to ensure consistent data entry?**

4: Consistency

- ✓ **Use a metadata standard**
 - ✓ Use controlled vocabulary
 - ✓ Use technical standards

4: Consistency

- ✓ Use a metadata standard
 - ✓ **Use controlled vocabulary**
 - ✓ Use technical standards

**catalog vs. catalogue
squirrel vs. Sciuridae
automatons vs. robots**

4: Consistency

- ✓ Use a metadata standard
 - ✓ Use controlled vocabulary
 - ✓ **Use technical standards**

YYYY

YYYY-MM

YYYY-MM-DD

YYYY-MM-DDThh:mmTZD

Term Name: format	
URI:	http://purl.org/dc/terms/format
Label:	Format
Definition:	The file format, physical medium, or dimensions of the resource.
Comment:	Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].
References:	[MIME] http://www.iana.org/assignments/media-types/

Term Name: date	
URI:	http://purl.org/dc/terms/date
Label:	Date
Definition:	A point or period of time associated with an event in the lifecycle of the resource.
Comment:	Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
References:	[W3CDTF] http://www.w3.org/TR/NOTE-datetime

4: Consistency

- ✓ **Use a metadata standard**
 - ✓ Use controlled vocabulary
 - ✓ Use technical standards

Data dictionary



Metadata Considerations

1. What type of data are you producing?
2. What are your metadata resources?
3. What metadata elements do you need to capture?
4. How are you going to ensure consistent data entry?
- 5. How are you going to capture your metadata?**

Documenting your data...

- In a filename
- In a readme file
- In a spreadsheet
- In an XML file
- Into a database

Documenting your data...

- In a filename

Example file naming conventions:

[investigator]_[method]_[subject]_[YYYYMMDD]_[version].[ext]

[project #] _[method]_[version]_[YYYYMMDD].[ext]

[YYYYMMDD] _[version]_[subject]_[datacollector].[ext]

[type of file]_[author]_[date].[ext]

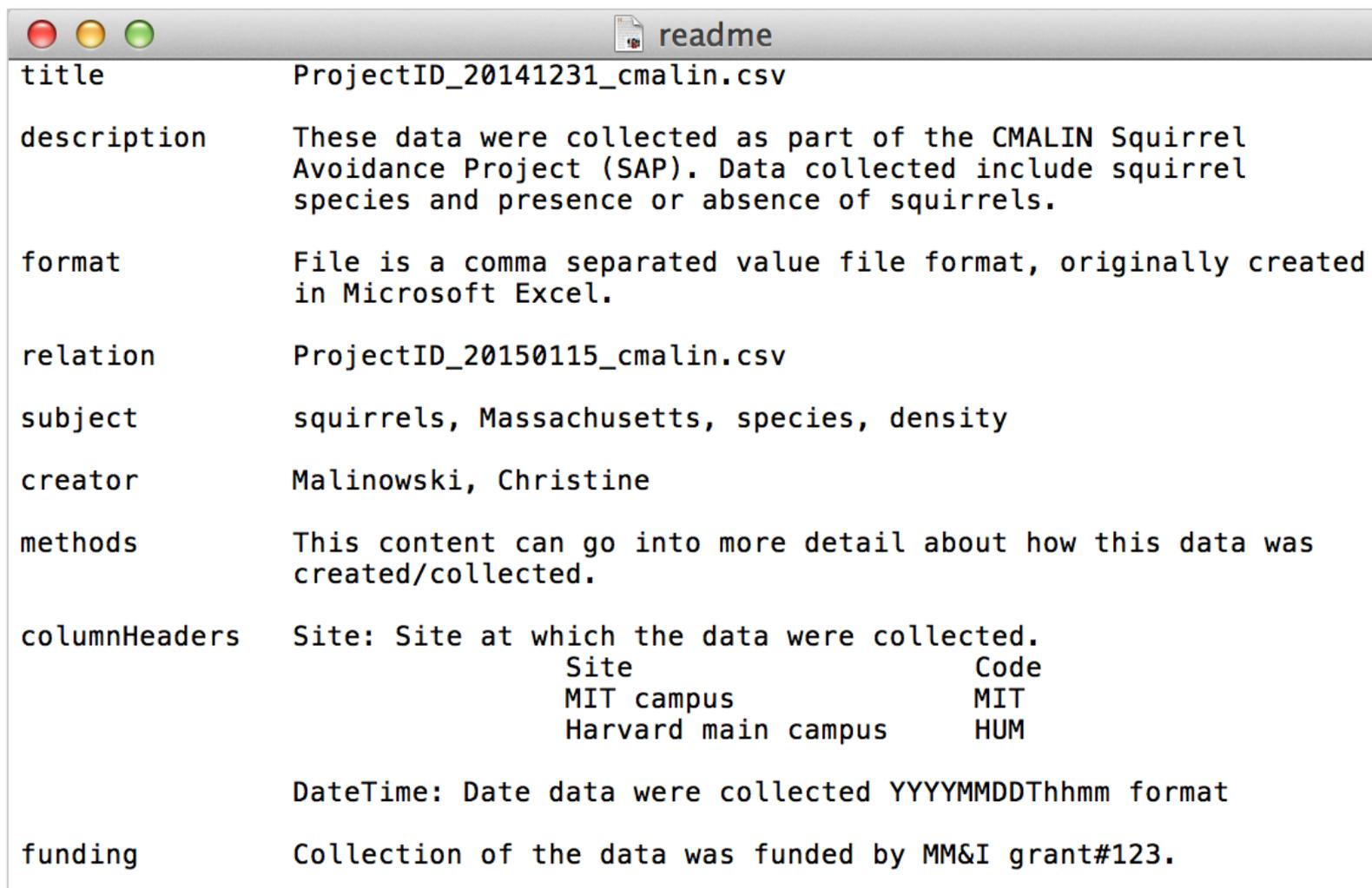
Site	DateTime	Sp	PA
MIT	20140115T0930	S1	1
MIT	20140215T0930	S1	0
MIT	20140315T1130	S2	2
MIT	20140315T1130	S1	4
HUM	20140115T0130	S1	1
HUM	20140215T0130	S1	1
HUM	20140315T0130	S2	2
HUM	20140315T0130	S1	1



[project #] _[subject]_[collector].[ext]
SAP1234_observations_cmalin.csv

Documenting your data...

- In a readme file



```
title      ProjectID_20141231_cmalin.csv
description  These data were collected as part of the CMALIN Squirrel
            Avoidance Project (SAP). Data collected include squirrel
            species and presence or absence of squirrels.
format     File is a comma separated value file format, originally created
            in Microsoft Excel.
relation   ProjectID_20150115_cmalin.csv
subject    squirrels, Massachusetts, species, density
creator    Malinowski, Christine
methods    This content can go into more detail about how this data was
            created/collected.
columnHeaders  Site: Site at which the data were collected.
            Site      Code
            MIT campus  MIT
            Harvard main campus  HUM
            DateTime: Date data were collected YYYYMMDDThhmm format
funding    Collection of the data was funded by MM&I grant#123.
```

More on readme files

Best Practices

- Create one readme file for each data file/dataset
- Name the readme so that it is easily associated with the data file(s) it describes
- Write your readme document as a plain text file
- Format multiple readme files identically
- Use standardized date formats
- Follow the conventions for your discipline (aka mimic known metadata standard)

http://data.research.cornell.edu/sites/default/files/SciMD_ReadMe_Guidelines_v4_1_0.pdf

Documenting your data...

- In a filename
- In a readme file
- In a spreadsheet
- In an XML file
- Into a database

Concerns About Creating Metadata

Concern	Solution
This seems like a lot of work!	incorporate metadata creation into data development process – distribute the effort
Time and resources to create, manage, and maintain metadata?	include in grant budget and schedule
Ensure readability / usability of metadata?	use a standardized metadata format

Conclusion

- Questions?
- Other tips for your peers?
- See our web site:

<http://libraries.mit.edu/data-management>

MIT OpenCourseWare
<http://ocw.mit.edu>

RES.STR-002 Data Management
Spring 2016

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.