closer
The home of longitudinal research

What is metadata?

Hayley Mills

12 September 2019

CLOSER, UCL Institute of Education

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

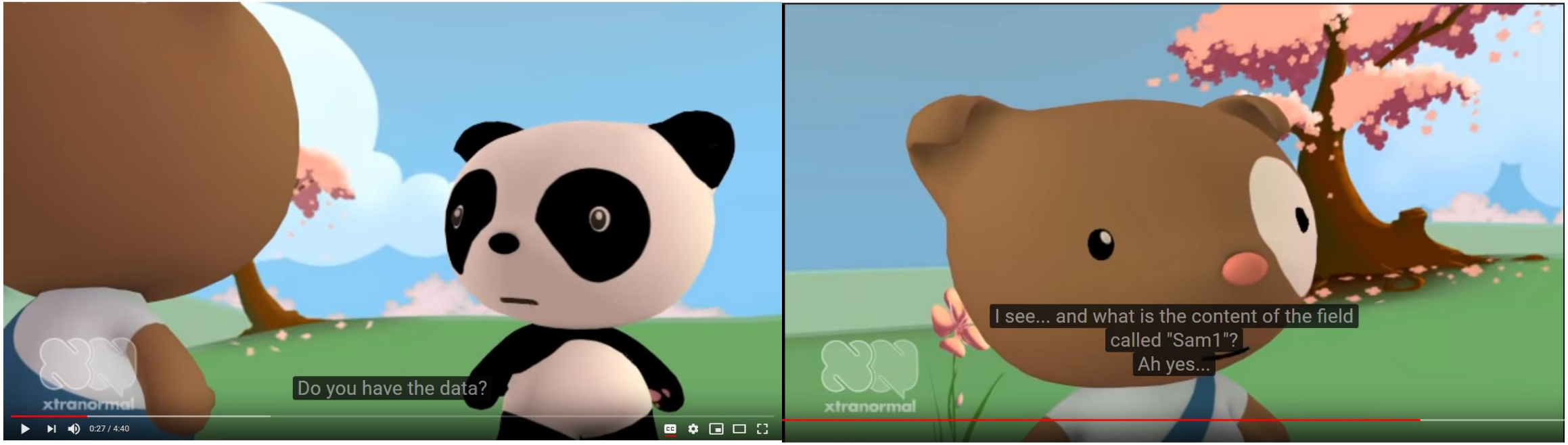
UCL

MRC Medical Research Council

This is some data

Person	Sam1	Sam2	Sam3	Sam4	Sam5
1	6	170	165		140
2	7	200		250	180
3	18	250	270		370
4	5	125		190	100
5	6	115	140		275
6	5	180	170	190	300

… and what is the content of the field called “Sam1”



https://www.youtube.com/watch?v=66oNv_DJuPc

Your average dataset is however a bit better

person	apple_est	apple_pb	apple_pb2	orange_pb	person_est
1	6	170	165		140
2	7	200		250	180
3	18	250	270		370
4	5	125		190	100
5	6	115	140		275
6	5	180	170	190	300

Often accompanied by documentation

Measurement Exercise

Today we will be doing measurements of various things and making a dataset and doing some analysis

Your student number — person

Estimating weights

Please take an apple from the box and write down how much you think it weights in ounces — apple_est

Weighing with a precision balance

Using the **same** apple write down how much it weighs in grams — apple_pb

Weighing with a precision balance (2)

If the apple is a Granny Smiths variety

Use the **same** apple write down how much it weighs in grams — apple_pb2

Weighing with a precision balance (3)

Pick a Seville Orange from the box of oranges and write down how much it weighs in grams — orange_pb

Estimating height

Write down how much you estimate your height in inches — person_est

Please enter this into the shared spreadsheet using the column names

From which we can derive this metadata

Label	Weight of apple (onz)	Weight of apple (g)	Weight of Granny Smith (g)	Weight of Seville (g)	Height of Person (inches)
Name	apple_est	apple_pb	apple_pb2	orange_pb	person_est
Concept	weight	weight	weight	weight	height
Unit Type	apples	apples	apples	oranges	person
Method	Estimated	Precision balance	Precision balance	Precision balance	Estimated
Unit	Ounces	Grams	Grams	Grams	Inches
Population	All apples	All apples	Granny Smith apples	Seville oranges	All students

Definition of 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." **

Metadata provides information enabling us to make sense of **data** (e.g. documents, images, datasets), **concepts** (e.g. controlled vocabularies, classification schemes), **real-world entities** (e.g. people, organisations, places) and **processes** (e.g. data collections, archiving, computation)

**- National Information Standards Organization*

http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Structured metadata

- Structured metadata defines the relationships between metadata.

“data defining the logical components of complex or compound objects and how to access those components”

- The primary purpose of structured metadata is for computers to deliver the semantic content to audiences, for instance on a website.

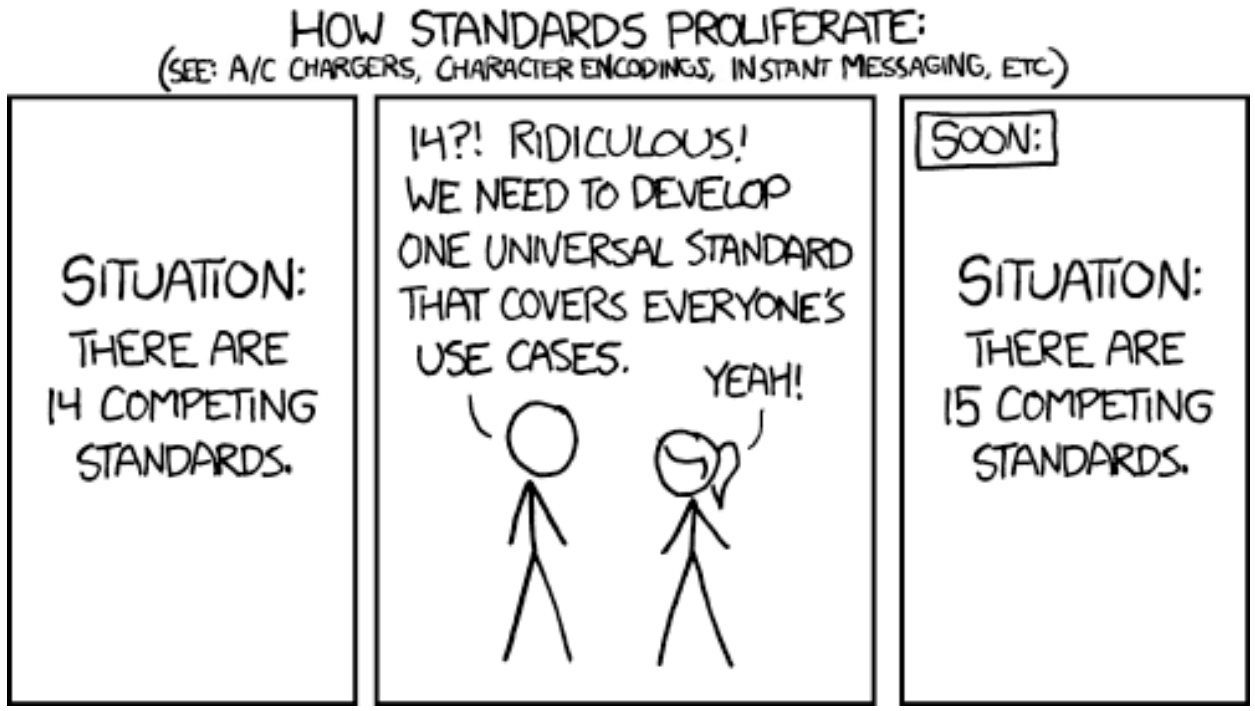
- Structural metadata is (pretty much) not for humans!

https://storyneedle.com/structural-metadata-key-to-structured-content

Open metadata standards

Open standards provide requirements, specifications or characteristics that meet a specific purpose.

- Open standards normally have a mechanism to amend, update or add to
- Local extensions are often a feature, so you can add features if you have a requirement that is not widely needed
- It is unlikely that your problems have not mostly been solved elsewhere



Open metadata standards

- Getting standards right, is not always easy!
  - For example it took many years before the “browser wars” gave us websites that look the same.
- Durable standards are those that have developed out of a shared problem and meet the majority needs of a community
- Tools are more efficiently written where there are larger user communities

- Standards allow:
  - re-use
  - discovery
  - access
  - interoperability internally and externally
  - sharing of metadata between communities