# 1 What is metadata and why is it important?

## 1.1 What is metadata

Metadata is data about data.

However, this definition is not particularly useful if you are not already clear what 'data' are and how to interpret them.

Instead, it is more useful to think of the difference between data and information: for data to become information, you need to understand the context in which the data are situated. Metadata is what provides this essential context.

**Metadata + Data = Information**

Example: this is some data but without any context, or metadata, it is very hard to know what any of it means.

| **Fields** | | **Data** |
|---|---|---|
| Eye colour | | Brown |
| Shoe size | Paris points | 42 |
| Height | Centimetres | 175 |
| Favourite colour | | Green |
| Self-description | | Very handsome |

| | |
|---|---|
| | Brown |
| | 42 |
| | 175 |
| | Green |
| | Very handsome |

Now, we have added some metadata in the form of 'fields' which describe each piece of data. We can see that these data refer to someone with brown eyes and size 42 feet. However, we are still missing some important pieces of contextual information – in particular, whilst we know the numbers 42 and 175 refer to shoe size and height respectively, we are not told what units these data have been measured in.

| **Fields** | **Data** |
|---|---|
| Eye colour | Brown |
| Shoe size | 42 |
| Height | 175 |
| Favourite colour | Green |
| Self-description | Very handsome |

Applying additional metadata provides further information about the units in which numerical data is presented and makes the data even easier to interpret.

| Fields | Unit | Data |
|---|---|---|
| Eye colour | | Brown |
| Shoe size | Paris points | 42 |
| Height | Centimetres | 175 |
| Favourite colour | | Green |
| Self-description | | Very handsome |

In this way, metadata provides the 'what', 'how', 'where', 'when', 'why' and 'who' of the data. It allows us to make sense of the data by providing the contextual information of where it comes from, the circumstances in which it was collected.

## 1.2 What does metadata look like?

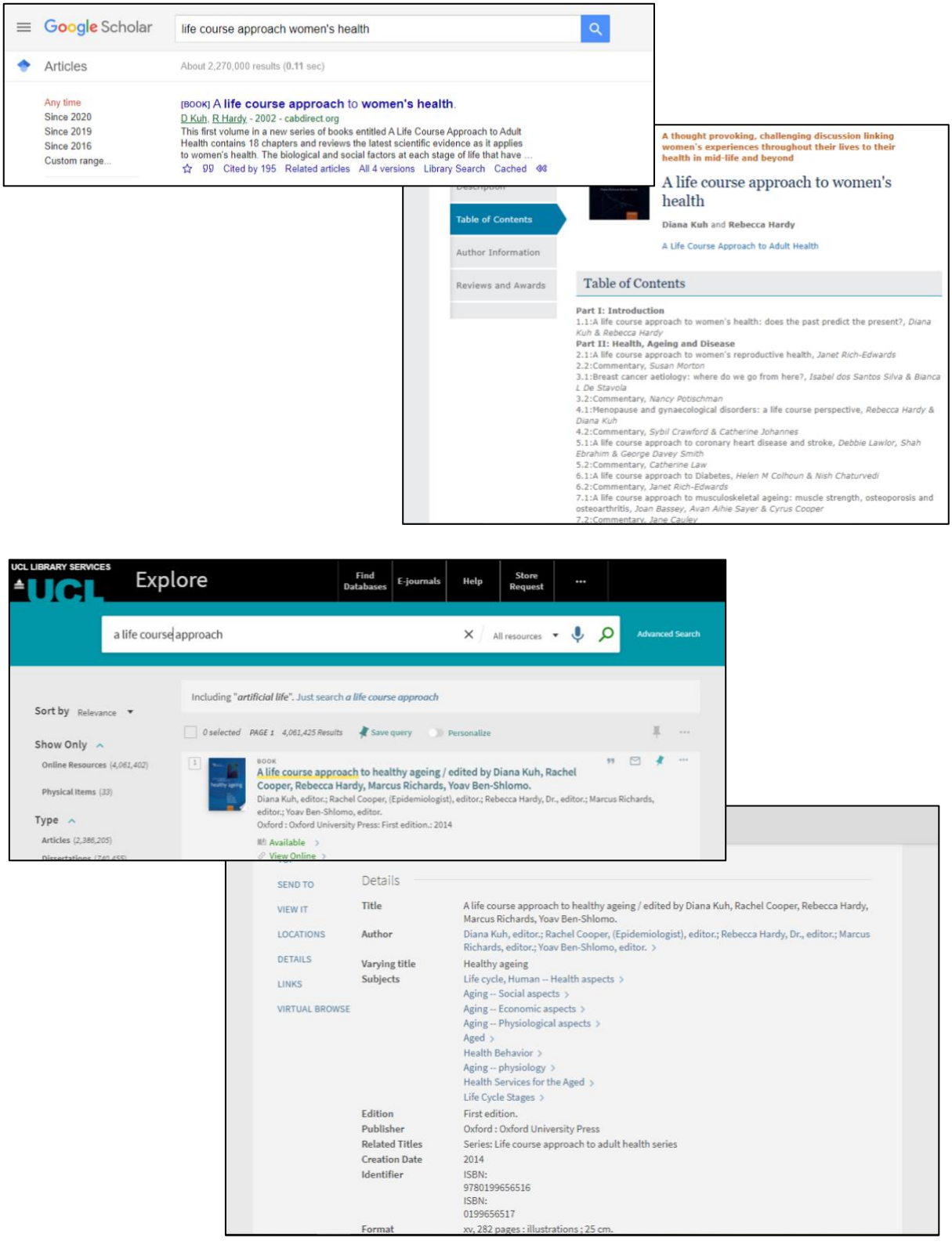
Metadata is a way of describing, synthesising and classifying complex sets of data. It includes information such as the title of a document, the name of a dataset, the study it relates to, the year it was collected, etc.

You will have come across several different ways in which metadata is presented, including:

- Tables of contents
- Indexes
- Glossaries
- References
- Citations
- Keywords

Authors: Neil Kaye, Hayley Mills and Jon Johnson

This is information about a work or dataset that is presented alongside it when you search, for example, via Google Scholar, University library catalogues or publishers' websites.

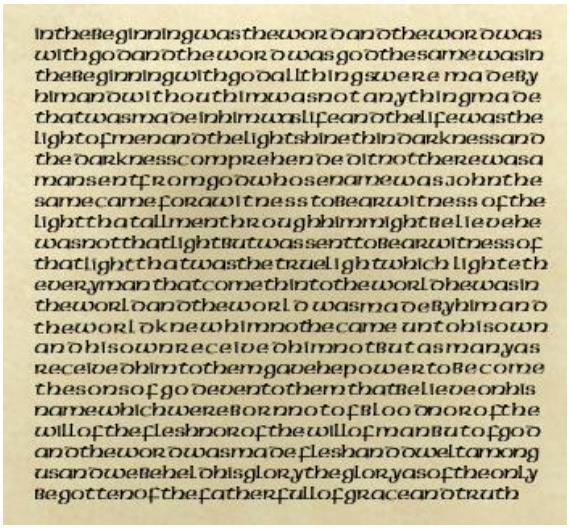



**Suggested citation:** Kaye, N., Mills, H. & Johnson, J. (2020). *Understanding metadata*. CLOSER Learning Hub, London, UK: CLOSER

## 1.3 Why is metadata important?

Why do researchers need metadata?

Using data without its metadata is like reading a long, complicated book without any punctuation. You have most of the information you need, but not all of it. It's harder to navigate. You have the terminology, but without the definitions. Metadata is required in order to make sense of this information. Adding metadata is like adding punctuation to the words and having structured metadata is like organising the words in the book into chapters with a content page.




Source: Alan Liu (2014). Digital Humanities. Class notes. [*accessed online: http://english197w2014.pbworks.com/w/page/74506796/Class%2012%20Notes*] English 149, University of California Santa Barbara

Metadata is hugely important for effective data use as it contains crucial information necessary to exploit the full potential of datasets for research.

Metadata helps you to:

- use the data you collect (e.g. it identifies variable names, labels and response codes);

- find other data (e.g. knowing the data collection year and life stage of the participant can help locate the studies you are interested in); and, in turn
- use the data you find (e.g. understanding whether data come from a self-reported estimate or from a scientific measurement can help you to interpret its accuracy).

Moreover, researchers can use metadata to gain a more thorough understanding of the data available to them. It provides a way for the original study team, who collected the data, to communicate with researchers using their data at a later stage.

Metadata represent a common mechanism for communication between researchers. It enhances and facilitates the reuse of data. The metadata associated with a dataset allows researchers not involved in the initial collection and interpretation of data to understand what was intended in the original study, what and where the data came from and how the survey was administered.

More technical details are also included in the metadata, including how survey logic (e.g. which participants were required to answer certain questions, and who was required to skip particular questions) was employed and under what conditions, which enables anyone looking at the data to understand the connection between the survey implementation through to the data received.

Ideally, metadata is produced and presented using standardised specifications. This again facilitates the re-use of data, data discovery, data access and the ability to share metadata between communities of researchers.