

# 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.

<b>Fields</b>		<b>Data</b>
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.

<b>Fields</b>	<b>Data</b>
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.

<b>Fields</b>	<b>Unit</b>	<b>Data</b>
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.

The image shows a Google Scholar search for "life course approach women's health" with approximately 2,270,000 results. The top result is a book titled "A life course approach to women's health" by Diana Kuh and Rebecca Hardy, published in 2002. A snippet of the book's description is visible: "This first volume in a new series of books entitled A Life Course Approach to Adult Health contains 18 chapters and reviews the latest scientific evidence as it applies to women's health. The biological and social factors at each stage of life that have ...".

The detailed view of the book shows the following information:

- Title:** A life course approach to women's health
- Authors:** Diana Kuh and Rebecca Hardy
- Series:** A Life Course Approach to Adult Health
- Table of Contents:**
  - Part I: Introduction
    - 1.1: A life course approach to women's health: does the past predict the present?, Diana Kuh & Rebecca Hardy
  - Part II: Health, Ageing and Disease
    - 2.1: A life course approach to women's reproductive health, Janet Rich-Edwards
    - 2.2: Commentary, Susan Morton
    - 3.1: Breast cancer aetiology: where do we go from here?, Isabel dos Santos Silva & Blanca L De Stavola
    - 3.2: Commentary, Nancy Potschman
    - 4.1: Menopause and gynaecological disorders: a life course perspective, Rebecca Hardy & Diana Kuh
    - 4.2: Commentary, Sybil Crawford & Catherine Johannes
    - 5.1: A life course approach to coronary heart disease and stroke, Debbie Lawlor, Shah Ebrahim & George Davey Smith
    - 5.2: Commentary, Catherine Law
    - 6.1: A life course approach to Diabetes, Helen M Colhoun & Nish Chaturvedi
    - 6.2: Commentary, Janet Rich-Edwards
    - 7.1: A life course approach to musculoskeletal ageing: muscle strength, osteoporosis and osteoarthritis, Joan Bassey, Avon Aihie Sayer & Cyrus Cooper
    - 7.2: Commentary, Jane Cauley

The image shows a search on the UCL Library Services platform for "a life course approach", yielding 4,061,425 results. The search results list a book titled "A life course approach to healthy ageing" edited by Diana Kuh, Rachel Cooper, Rebecca Hardy, Marcus Richards, and Yoav Ben-Shlomo, published by Oxford University Press in 2014.

The detailed metadata view for this book is as follows:

Field	Value
<b>Title</b>	A life course approach to healthy ageing / edited by Diana Kuh, Rachel Cooper, Rebecca Hardy, Marcus Richards, Yoav Ben-Shlomo.
<b>Author</b>	Diana Kuh, editor; Rachel Cooper, (Epidemiologist), editor; Rebecca Hardy, Dr., editor; Marcus Richards, editor; Yoav Ben-Shlomo, editor. >
<b>Varying title</b>	Healthy ageing
<b>Subjects</b>	Life cycle, Human -- Health aspects > Aging -- Social aspects > Aging -- Economic aspects > Aging -- Physiological aspects > Aged > Health Behavior > Aging -- physiology > Health Services for the Aged > Life Cycle Stages >
<b>Edition</b>	First edition.
<b>Publisher</b>	Oxford : Oxford University Press
<b>Related Titles</b>	Series: Life course approach to adult health series
<b>Creation Date</b>	2014
<b>Identifier</b>	ISBN: 9780199656516 ISBN: 0199656517
<b>Format</b>	xv, 282 pages : illustrations ; 25 cm.

**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.



In the beginning was the word and the word was with God and the word was God the same was in the beginning with God all things were made by him and without him was not anything made that was made by him was life and the life was the light of men and the light shined in darkness and the darkness comprehended it not there was a man sent from God whose name was John the same came for a witness to bear witness of the light that all men through him might believe he was not that light but was sent to bear witness of that light that was the true light which lighteth every man that cometh into the world he was in the world and the world was made by him and the world did not know him when he came unto this world and his own received him not but as many as received him to them gave he power to become the sons of God even to them that believe on his name which were born not of blood nor of the will of the flesh nor of the will of man but of God and the word was made flesh and dwelt among us and we beheld his glory the glory as of the only begotten of the father full of grace and of truth

Source: Alan Liu (2014). Digital Humanities. Class notes. [accessed online:

<http://english197w2014.pbworks.com/w/page/74506796/Class%202012%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);

Authors: Neil Kaye, Hayley Mills and Jon Johnson

- 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.