

Metadata have been used for a long time. Libraries used indexing cards to describe books and documents, for example. Digital technology means that metadata can now be searched by machines. The standardised description of research data through its metadata can vary according to the specific requirements of different scientific fields but overall this form of description is a major open science issue.

## Metadata - data about data



### How it works

Metadata gives a **precise description of data**. If we imagine a dataset as a tin can, the metadata are like the can's label that describes its contents – 'manufacture' or production date, producer, origin, methodology, persistent identifier, etc.

### Note

Metadata can be:

- **Embedded** - automatically produced by devices or software (date, format, etc.).
- **Enriched** - added by the author (keywords, project name, licence, etc.).



### Usefulness



Metadata help you structure, organise and manage all your data.  
You can trace data processing operations.  
They define the data's uses.  
They help prepare data for sharing and interoperability.



### Making data FAIR:

Findable  
Accessible  
Interoperable  
Reusable

**Did you know?** Even if data disappears, their metadata remains! Metadata accompanying a dataset should always remain accessible. They provide useful information when data cannot be shared (embargo, restricted access) or are withdrawn because they are obsolete, etc.

## Structuring metadata



### Metadata standards

A standard is a blueprint for metadata that a group of users have adopted as a model. It is recognised, standardised and used on a large scale.  
A standard may be **specialised**, depending on the discipline or type of data.

It is recommended that you use **dedicated vocabularies** like thesauri, lexicons, taxonomies, etc.



### Exchange formats

This kind of format is a **digital representation of a metadata standard**. It enables machines to read and exchange content. Common open formats include XML, TXT and CSV.

### Note

Exchange formats should **not be confused** with file formats:

- **File formats** are the result of the work of **technical committees** that produce **specifications** that are either freely accessible (free or open formats) or protected (proprietary formats).
- **Exchange formats** are used to share content between machines.

# Focus on examples of metadata standards

**DataCite Metadata Schema:**  
DOI attribution

**Encoded Archival Description (EAD):** Description of archives

**Darwin Core (DwC):**  
Biodiversity field

**Data Documentation Initiative (DDI):** Social, behavioural and economic sciences

**International Press Telecommunications Council (IPTC):** An author's description of an image

**Exchangeable image file format (EXIF):** Automatic description of a photograph

## An example of an interdisciplinary standard - Dublin Core

**Simple Dublin Core** was initially made up of 15 elements and has now been enriched by around forty more specific elements. This is known as **qualified Dublin Core** and is part of web architecture.

The example below illustrates specifications added to the 'Date' and 'Relationship' elements.

Single element	Definitions	Example of a specific element
<b>Title</b>	<b>Name of the resource</b>	
<b>Subject</b>	<b>Theme of the resource's content</b>	
<b>Description</b>	<b>Abstract, table of contents, etc.</b>	
<b>Creator</b>	<b>Main author of the resource</b>	
<b>Publisher</b>	<b>Entity responsible for disseminating the resource</b>	
<b>Contributor</b>	<b>Co-authors involved in developing the resource</b>	
<b>Date</b>	<b>Date created or made available</b>	<b>Date Created, Date Copyrighted, Date Valid, Date Available, Date Modified, Date Accepted, Date Submitted, Date Issued</b>
<b>Type</b>	<b>Type of content - image, sound, text, etc.</b>	
<b>Format</b>	<b>Format or size of the resource</b>	
<b>Identifier</b>	<b>Unique reference, DOI, URL, ISSN, etc.</b>	
<b>Source</b>	<b>Reference to a resource which the current resource was derived or created from</b>	<b>Has Format</b> (format transformation relationships involve one resource being derived from another) <b>Is Version Of</b> (version relationships involve one resource being a state or historical release of another resource by the same creator)
<b>Language</b>	<b>Original language of the resource</b>	
<b>Relation</b>	<b>Reference to a related resource</b>	
<b>Coverage</b>	<b>Spatial and temporal scope</b>	
<b>Rights</b>	<b>Information on rights associated with the resource</b>	