

Good Practice and Guidance – Metadata for documents

1. What is metadata?

“Metadata are a subset of core data documentation, which provides standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection” (Ref UK Data Archive).

Metadata is data *about* data.

It is part of broader contextual information that accompanies data to ensure it can be found and understood over time. The information that can be recorded can range from a detailed description of the data to explanatory material about why the data was created and how it has been used.

Therefore, alongside the term *metadata* the term documentation may be used, referring to all the information necessary to interpret, understand and use a given dataset, a set of files or a single document - sometimes the words are used interchangeably.

Within research metadata is a significant and developing area with specific descriptive standards such as CERIF, the Dublin Core Metadata Initiative (DCMI) and the Data Documentation Initiative used to enable sharing, access, interpretation and re-use of research data. Research funder requirements now demand researchers create and make metadata openly available, notably to describe complex datasets, and thus facilitate access and re-use.

This document is restricted to basic appreciation of metadata, and typically application through MS Office documents/files.

2. Why create metadata – what are the Benefits?

“A crucial part of making data user-friendly, shareable and with long-lasting usability is to ensure they can be understood and interpreted by any user. This requires clear data description, annotation, contextual information and documentation

Data documentation explains how data were created or digitised, what data mean, what their content and structure are, and any manipulations that may have taken place. It ensures that data can be understood during research projects, that researchers continue to understand data in the longer term and that re-users of data are able to interpret the data. Good documentation is also vital for successful data preservation.” (Ref UK Data Archive).

Good documentation ensures your data can be:

- Searched for and retrieved
- Understood now and in the future
- Properly interpreted, as relevant context is available.

3. When should metadata be created?

It is good practice to begin to document your data at the earliest point of your work and continue to add information as it progresses. It is easier to capture it then, rather than trying to remember things at a later date.

4. What metadata should be created?

Ask yourself, *“What information would I need to understand and use this data in twenty years?”*

Potentially useful information includes basic description:

- Title
- Date
- Author(s)
- Format
- File name/path
- Storage location/URL
- Subject
- Rights
- Access information
- Keywords

5. How to create metadata in MS Office files – “Properties”

There are a number of ways you can add documentation to your data: a) Embedded documentation, b) Supporting documentation (in a separate file), and c) Catalogue metadata (usually structured according to an international standard, used to identify and locate the data that meet the user's requirements via a web browser or web based catalogue).

You can ensure that digital files are well-structured internally by adding file names, creation date, author(s), version information, full referencing etc. In addition MS Office products use **“Properties”** to record common pieces of metadata such as title, author, organisation, subjects and keywords, and additional comments.

Not only will this help you keep your files organised and possible to interpret, it will also allow you to sort folders by properties that you have added and search for documents with particular properties.

Some of these Properties fields are user editable whilst some are completed automatically by the software program.

MS Word Properties Fields (other MS Office programs have similar fields)

User Editable Fields	Automated Fields
Title	File Size
Tags	Pages
Comments	Words
Status	Edit Time
Categories	Template
Subject	Last Modified (time and date)
Hyperlink Base	Created (time and date)
Company	Last Printed (time and date)
Manager	Last Modified By
Author	

Version Control Table

Version Number	Author	Purpose/Change	Date
0.1	Andrew Burnham, ITS	Initial draft – to IAS, Sandy Pearson	19/06/2012

Acknowledgements:

This document was written with reference to guidance provided by the UK Data Archive, and the University of Cambridge, the University of Glasgow and Monash University.

DRAFT