Getting the most out of your research
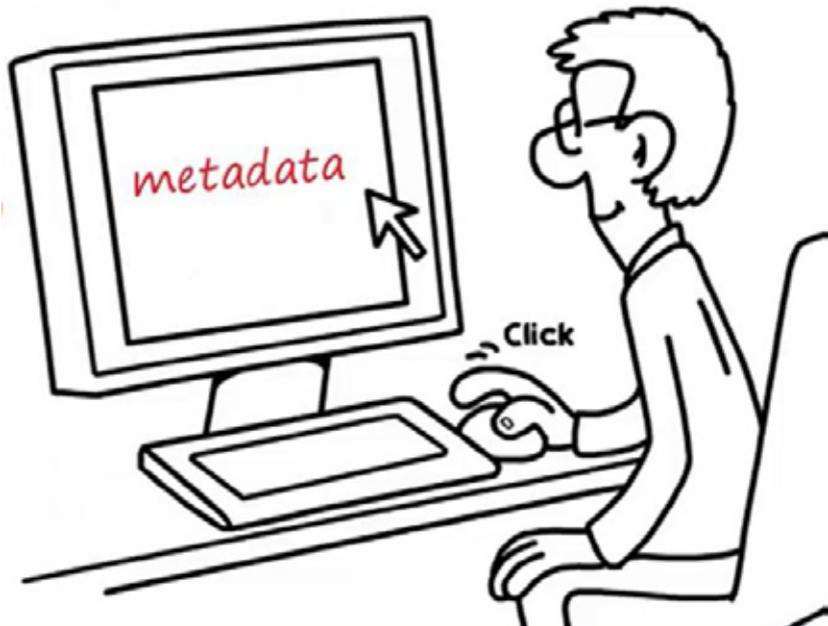
Metadata, spatial data, and collaboration

Authors

Joanne Stevenson
Resilient Organisations, Resilience to Nature's Challenges National Science Challenge

Mirjam Schindler
University of Canterbury, Building Better Homes, Towns, and Cities National Science Challenge

National SCIENCE Challenges | RESILIENCE TO NATURE'S CHALLENGES | Kia manawaroa – Ngā Ākina o Te Ao Tūroa

National SCIENCE Challenges | BUILDING BETTER HOMES, TOWNS AND CITIES | Ko Ngā wā Kainga hei whakamāhorahora

Resilient ORGANISATIONS

www.resorgs.org.nz

Introduction

If you plan to share your data with anyone, adding metadata can improve others' ability to use and interpret your work. When it comes to large-scale research collaborations, metadata is key to producing optimal results.

Access to consistent, high-quality metadata is critical to finding, understanding, and reusing scientific data. This short guide describes the basic elements you should add to each dataset, and provides guidance on how to create metadata for optimal outcomes.

The Basics

What data should I include with my dataset?

In the table on the next page we have compiled a list of items you should include when you record metadata.

The element descriptions and guidelines are synthesized from sources advocating best practice data management: Dublin Core Metadata Initiative (dublincore.org), National Information Standards Organization (niso.org), and the New Zealand Government's data management and discovery service (data.govt.nz).

It looks like a big list, but it's faster and easier to capture this information than you might think.

A LITTLE HIGH-QUALITY METADATA CAN GO A LONG WAY!

Element	Element description	Guidelines
Title	The name given to the resource. Typically, a title will be a name by which the resource is formally known.	The title should be descriptive enough that a member of the public can have a reasonable expectation of its subject. It should be long enough to be informative but short enough to be readily understood in isolation by someone unfamiliar with your work.
Subject	A topic of the content of the resource.	The subject should allow a user to understand the broad content of the data and classify it in a disciplinary background.
Type	The distinguishing data storage components and structures.	Describe the type of resource, i.e. for instance a spatial dataset, a text, a report, a set of files, a database, a survey.
Description	An account of the content of the resource.	Text that describes the contents, purpose, source and structure of the dataset, and (where necessary) includes any comments about its quality and reliability. Use full sentences. Wherever possible include information about the "spatial extent (relating to a geospatial or geographic area) or temporal extent (relating to a period of time), note this in the description. Include any notes or cautions that may help future users assess the data's suitability for their purposes. It may also be appropriate to note the version. Although the source of the description may be a web page or other structured text with presentation tags, it is generally not good practice to include HTML or other structural tags within the description element.
Tags/ keywords	Keywords used to describe the data topic.	Provide keywords that define what your resource is about. Draw terms that are a suitable domain vocabulary. You want your keywords to be relevant to what people are searching for so they have a bet ter chance of reusing your resource.
File formats	A file format is the structure of how information is stored in a computer file.	Describe in which file format the resource is available, e.g. .doc (Word), .xls (Excel), .csv (comma-separated values), .pdf, .shp (geospatial vector data), .tif (bitmap). This is important for potential users to know how data can be reused, edited or visualised.

Element	Element description	Guidelines
Resources	Description of which resources the dataset contains.	Provide data-specific information, such as how many observations the dataset contains, a list of variables, definitions of codes or symbols and abbreviations, the unit of measurement if applicable.
Methodology	A description of the steps taken to produce the resource.	Use full sentences. Describe how the data were generated from the raw or collected data. Describe how the data has been obtained, which steps have been applied to modify the data, which algorithms and software have been used and which assumptions made.
		Describe any quality-assurance procedures performed on the data. Note which issues might be of concern to users of the data. Include as much information as possible since it helps potential users to evaluate how fit the data is for their purpose. A potential user should ideally be able to replicate the resource from its original state given the information provided.
Purpose of dataset creation	A description of the objective and context of data creation.	Describe the context in which the dataset has been created and the objective of its creation to put things into perspective. Describe why you created the dataset.
		This information can help users to evaluate whether the data is also fit for their purpose. Without context, interpretation of the data and an understanding of its limitation is difficult.
Contacts	Contact details for data sources and questions.	Provide contact details of the creator of the resource, original sources of the data and information on who to contact in case of questions.
Dates	A date of an event in the lifecycle of the resource.	Two components are important: when the data was created and which time frame the data describes.
		Include information on when the resource was created, updated, modified.
		Describe the temporal extent of the data. Dates should be given as accurately as possible.
Frequency of Update	Information about when data has been modified.	Provide information on how frequently the dataset has been modified or updated since its original creation.

Element	Element description	Guidelines
Licensing	The licence under which the data is published.	Licence information tells others what they can do with your resource. It encourages the reuse of your resource and creates visibility of your efforts downstream if you ask for attribution. Describe whether your resource is openly available, available for reuse but not for modification, subject to restrictions or available if attributed. See https://creativecommons.org/licenses/ for all Creative Commons licences and their definitions.
Rights	Information about rights held in and over the resource.	Include a rights management statement for the source or a reference to a service providing such information. Note any information on data confidentiality.
Note	Any comments users should take into consideration when using the data.	Include any comments users should take into consideration when using the data; major issues known with the dataset that might constrain the use of the data; known data quality issues.

Research Collaboration Standard Metadata

Enhancing your Description Element for Optimal Outcomes

In the table on the previous pages, the 'Description' element is defined as: Text that describes the contents, purpose, source and structure of the dataset, and (where necessary) includes any comments about its quality and reliability. This is where you can capture some of the most interesting and useful information about your data. Here are some tips for maximising your metadata Descriptions.

How do I describe spatial elements of my data?

Spatial data, also known as geospatial data, is information about the location of an object, its shape or size, or how it relates to other objects in space. Adding spatial information to your data can increase the power of your research. Spatial data and analysis allow people to link information and insights to a common location-based reference. So, if you have captured spatial elements in your research – even something as simple as the town in which survey respondents live – be sure to capture that information in your metadata.

Start by describing the method used to spatially represent geographic information. The spatial extent of your dataset as a whole should be described as well as the spatial extent of each spatial location within the dataset. Make sure to include information on the reference system used and the spatial resolution of the data. If the data is gridded data, describe the size of the grids and how the grid was defined. If the data is in vector format, describe its features. If the datasets contain data of different spatial resolutions, describe on which spatial scale each has been collected, modified or which aggregation has been applied. What is the lowest and what the largest spatial scale in the dataset? If sample spatial locations or elements are included, describe how they have been selected and how they relate to another.

You might simply be using postcodes or place names from the GeoNames database in a spreadsheet. You can write this in your Description element.

How do I describe temporal elements of my data?

Temporal elements relate to data that has an instance in time. Data might include valid time information reflecting the time period for which the data is 'true' in the real world. For example, the latest New Zealand Census data was collected on 5 March 2013. Alternately, transaction times can capture information about the time when data was captured or altered.

Temporal information to include in the Description elements of your metadata are:

i) the time of data collection/creation,

ii) the date of publication

iii) the date of last revision, and

iv) the temporal extent of the data.

Information about the currency of the dataset and in which time intervals the data is reported is essential.

How do I describe data quality?

Internal data quality describes the completeness of the dataset, its accuracy, errors in the dataset, temporal and spatial resolution, consistency and reliability. Describe how missing data has been treated, whether the quality is consistent across the dataset, potential factors that could have impacted the validity of the data, which measures have been taken to assure quality. Basically include any issues that you might think a potential user should be aware of.

How do I describe the context in which the data was created/collected?

The objective of the project for which the data has been created or collected and who was involved gives important perspective to a dataset. Provide information on the timeframe of data creation, the methodology that was applied, external limitations, funding sources that supported the collection of the data etc. Useful information to provide can answer questions like who/how/what/when/where and most importantly why the dataset was created.

On an ongoing basis, ensure that the metadata is adequately maintained and kept up to date.

Attaching metadata to your file

You can add metadata to any file you create. This can be as simple as providing a "readme.txt" file for each file containing information on the above issues (example: https://cornell.app.box.com/v/ReadmeTemplate). You can also add metadata directly to your files within the programme in which they were created.

Microsoft Word / Excel / PowerPoint

This guidance applies the process of adding Metadata to files in the PC version of Microsoft Office 10 products.

1. Open the file you wish to add metadata to.

2. Select the File menu and then the Info option. The right-hand side of the screen displays the document properties.

3. Enter the Author, Title and Tags in the respective fields.

PDFs

1. Open the file you wish to add metadata to.

2. Select the File menu and then the Properties option. The Description tab displays the document properties.

3. Enter the Title, Author, Subject and Keywords in the respective fields.

Videos

1. Right click on the file you wish to add metadata to.

2. Select the File menu and then the Properties option. The Details tab displays the document properties.

3. Enter the Contributing Artists, Title, Tags, etc. in the respective fields. Include Description element information in the Comments section.

ArcGiS

You can view and edit metadata for spatial data using ArcCatalog or ArcMap. The Description tab lets you view and edit metadata for ArcGIS items. In ArcMap,

1. Right-click on a data layer in ArcGIS to which you wish to add data

2. Select Data and then View Item Description

3. Click Edit on the top of the panel to edit the Title, Tags, Purpose, License etc.

How to credit the dataset

If applicable, you can state how the dataset should be cited if reused under the mentioned liHere is an example of metadata collected for survey data with geospatial components.

Use Case

Here is an example of metadata collected for survey data with geospatial components.

Title	Survey response data – residents' use of local amenities.
Subject	Multiple-choice survey on the use of public local amenities among households living in Christchurch.
Type	Spreadsheet with survey responses.
Description	The datatset includes survey questions and responses by 1500 residents of Christchurch on their use and attitudes towards public local amenities.
Tags/ keywords	Survey, public local amenities, residential choice, economic benefits.
File formats	.xls spreadsheet.
Resources	Survey questions and responses by 1500 residents of Christchurch.
	Interviewees were asked to state their frequency of use [daily/once per week/ monthly], age, income [EUR], household size [1/2/3-4/5+], choice criteria [A,B,C], X [], Y [], Z [] and point their home location, work location and the location of their preferred local amenity on a map [UTM].Responses were given on household level.
Methodology	The survey questions have been designed following guidelines by XY in order to retrieve information on YZ. The survey was conducted face-to-face using tables by 5 trained interviews in public spaces in the suburbs A, B, C with residents present at that time. The survey locations where chosen based on criteria 1 and criteria 2. The interview took about 15 minutes to conduct. It includes multiple-choice questions and questions where respondents had to locate places on a map. Ex-post, the data was cleaned by removing incomplete or erroneous entries.
	The software used for conducting the survey was XY.
	Note: locations freely pointed at the map are rough indications only; accuracy up to meshblock level; the profile of the sample survey respondents is in general representative of the population of Christchurch; yet, the group of retired residents is slightly over-represented (x % instead of y %).

Purpose of dataset creation	Survey on the use of public local amenities among residents; research project funded by XY and lead by XY in order to quantify the benefits residents retrieve from public local amenities in Christchurch.
Contacts	Contact for survey design: XY, University YZ; address, phone, email. Contact for data collection & preparation: XY, University YZ; address, phone, email.
Dates	The survey has been conducted 1-20 January 2017. The data has been published 2/12/2017
Frequency of Update	
Licensing	CC BY-NC-SA
	Use for non-commercial purposes as long as authors are credited and new creations are licensed under identical terms
	Please cite the dataset as YX (2017). Survey response data on the use of public local amenities in Christchurch. YZ.
Rights	All data owned by the project XY.
Note	This survey has been conducted for the particular purpose X and the choice of income classification was driven by the reference to tax levels on local amenities rather than standard practice.
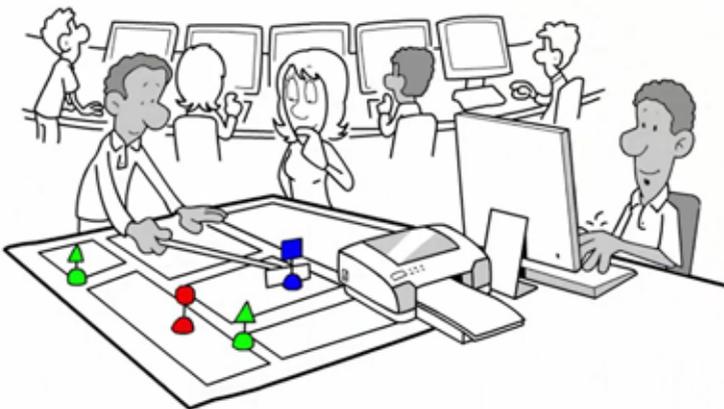
Other Resources

This short guide is a basic starter resource for people getting started on their collaborative data management journey.

Please also see our explainer videos and additional detailed resources created organisations advocating best-practice data management below.

Explainer Videos

Metadata Overview
https://www.youtube.com/watch?v=zs1t2fbOQ5c

Metadata: Where do I start?
https://www.youtube.com/watch?v=R8g4XweYOz8&t=13s

Additional Resources

www.open.ac.uk/about/web-standards/editorial-standards/metadata/adding-metadata-documents#office_docs

www.open.ac.uk/about/web-standards/standards/editorial-standards/metadata

data.govt.nz/toolkit/preparing-your-data-checklist/

data.govt.nz/toolkit/what-metadata-should-i-include-with-my-dataset/

Spatial Data Management

toolkit.data.gov.au/index.php?title=Publishing_your_data#Intro_to_spatial_data

www.nss.gov.au/nss/home.nsf/pages/Statistical%20Spatial%20Framework%20Guidance%20Material

www.nss.gov.au/nss/home.NSF/pages/Statistical+Spatial+Framework+Guidance+Material/$File/Geocoding%20Unit%20Record%20Data.pdf

Data Quality

www.nss.gov.au/nss/home.nsf/pages/NSS+Data+Quality+resources+landing+page