(Pixabay, CC0)

Metadata

Good documentation ensures your data can be properly interpreted as relevant context is available. Besides supplementary documents like lab journals or notebooks, documentation should include metadata, "data on the data itself". Metadata contains generally descriptive information like the creator, date and subject of the dataset, information that tells how the dataset relates to other documents and information on the format, access rights or preservation aspects.

It is important to capture this information right from the start of the project. Sometimes metadata is embedded in the dataset as it is automatically created by the software or tool that generates the data. If not, metadata can be registered in a separate file, like a readme.txt, spreadsheet or xml document.

If you want to share your data with others, or reuse it later on, you would do well to record the metadata in a standardized way. When you use an established metadata scheme, your data will be discoverable and intelligible both for humans and machines, now and in the future.

The standard that is most suited for your datasets depends on your discipline. You'll find a list disciplinary meta data standards on the website of the Digital Curation Center[1]. Data archives and repositories often require a specific metadata scheme. Check this in an early stage.

Please turn over for an example of a metadata standard (Dublin Core).

---

[1] http://www.dcc.ac.uk/resources/metadata-standards

Dublin Core

A well known standard, used by two national data archives (DANS and 3.TU.Datacentrum) is Dublin Core. The Dublin Core Metadata Element Set includes :

Title - the title of the data object
Creator - the person or entity responsible for creating the data object
Subject - subject terms or keywords that describe the data object
Description - a brief description, or abstract, of the data object
Publisher - the entity responsible for making the data object available
Contributor - a person or entity who contributed to the creation of the data object
Date - data of creation, publication, or revision of the data object
Type - the type of object. For data this would typically be "dataset"
Format - a description of the format or file type(s) of the data object
Identifier - a permanent identifier used to locate and identify the data object
Language - the language(s) used within the data object (if applicable)
Source - a relational element describing the lineage of the data object
Relation - a relational element describing the relationship of this data object to other objects, collections, or entities
Coverage - describes the spatial and temporal context of the data object
Rights - describes any rights, restrictions, or terms of use

Example of Dublin Core metadata for a dataset in the 3TU.Datacentrum:

Meerman, J.H.N. (John) (2014) 3D Cell Phenotyping Data. Leiden University. Dataset.
http://dx.doi.org/10.4121/uuid:d5b91e46-07e7-4077-bd63-3fa2b82c847f

Dataset | **3D Cell Phenotyping Data**

▶ ▶ ▶ ▶ ▶ ▶ ▶ ▶ ▶ Link/cite as doi:10.4121/uuid:d5b91e46-07e7-4077-bd63-3fa2b82c847f | **show link code** | **full citation**

▼go to DATA section▼

title	?	3D Cell Phenotyping Data
creator	?	*Meerman, J.H.N. (John) (Associate Professor)*
contributor	?	Price, L.S. (Leo) (Senior Staff Member)
contributor	?	Zi, Di (Research Associate)
date created	?	2014-07-22
date published	?	2014
description	?	The publication describes how, in many situations, 3D cell cultures mimic the natural organization of tissues more closely than 2D cultures. Conventional methods for phenotyping such 3D cultures use either single or multiple simple parameters based on morphology and fluorescence staining intensity. However, due to their simplicity many details are not taken into account which limits system-level study of phenotype characteristics. The authors have developed a new image analysis platform to automatically profile 3D cell phenotypes with 598 parameters including morphology, topology, and texture parameters such as wavelet and image moments. As proof of concept, they analyzed mouse breast cancer cells (4T1 cells) in a 384-well plate format following exposure to a diverse set of compounds at different concentrations. The result showed concentration dependent phenotypic trajectories for different biologically active compounds that could be used to classify compounds based on their biological target. To demonstrate the wider applicability of our method, they analyzed the phenotypes of a collection of 44 human breast cancer cell lines cultured in 3D and showed that their method correctly distinguished basal-A, basal-B, luminal and ERBB2+ cell lines in a supervised nearest neighbor classification method.
language	?	en
publisher	?	Leiden University
subject	?	3D cultured micro-tissues
subject	?	breast cancer
subject	?	cell cultures phenotyping
subject	?	nearest neighbor classification
▲ in collection	?	General collection of datasets
related publication	?	⤢ Ultra High Content Image Analysis and Phenotype Profiling of 3D Cultured Micro-Tissues (article, 2014)

DATA

Data-Zi-Di-et-al-29-compounds.rar (3.9 GB) (application/x-rar-compressed)