

# UNIVERSITY OF OSLO

## Module III: Data organization, metadata, and documentation

Edina Pózer and Ivana Malovic  
The University of Oslo Library

Materials developed as part of the *Skills development project for research data*:  
<https://www.ub.uio.no/english/about/projects/rdm-skills/>



# Today's course

- One hour lecture with a short break
- Q&A session

Please feel free to write comments and questions in the chat!

# Agenda

- Data file structure
- What are metadata and why we need them?
- Documentation and how to create it?
- File naming conventions
- Version control
- Sustainable file formats



# Data file structure

# Design a data file structure

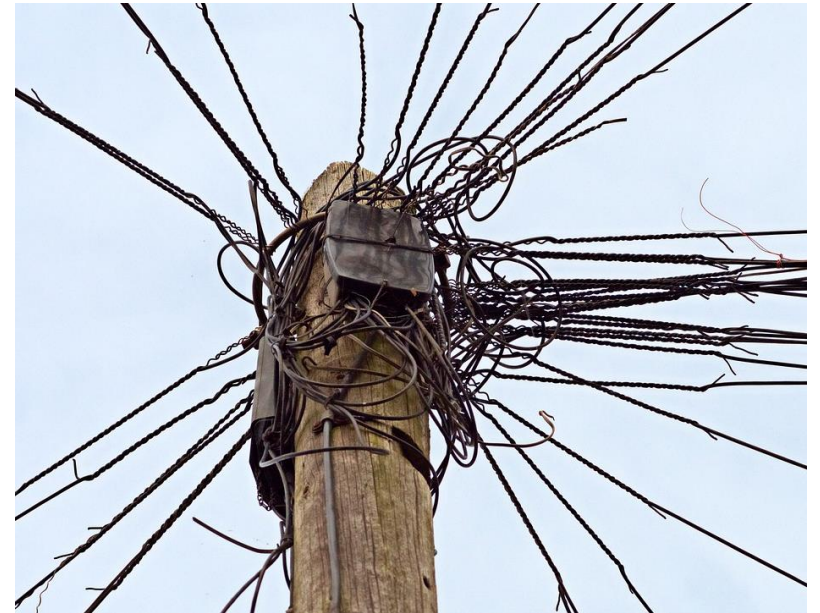
- In an early stage of your research, you are faced with the question of **what form your data files should take**. Your initial decision about the structure of your data files should be considered thoroughly.
- The **data file structure** has a huge impact on the possible ways your files can be processed and analysed and once your structure has been filled with data, any changes to it are usually laborious and time-consuming.

# Create a data organising system that...

- follows your working pattern
- is systematic and logical
- is quick and easy to navigate
- is simple enough to be used all the time
- is considered and thought through before you start (!)
- is scaleable

# Organise your data

- Do not use your computer desktop as a storage place
- Make a folder hierarchy and give your folders **descriptive** and **informative** names
- Avoid folders that become too broad or general, create more **subfolders** instead
- Keep active and finished parts of your project separate, and take the time to tidy once in a while



Metadata



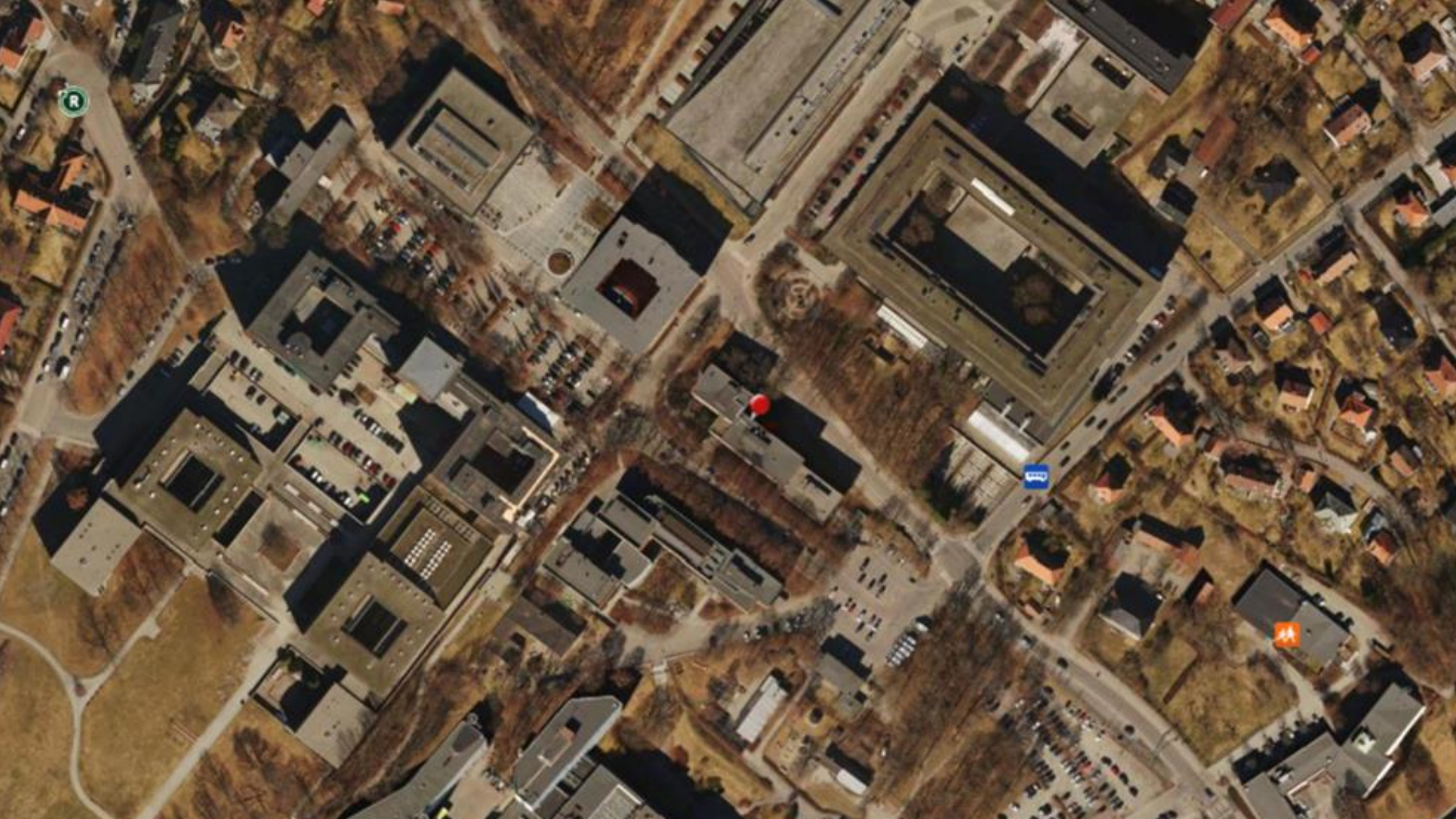


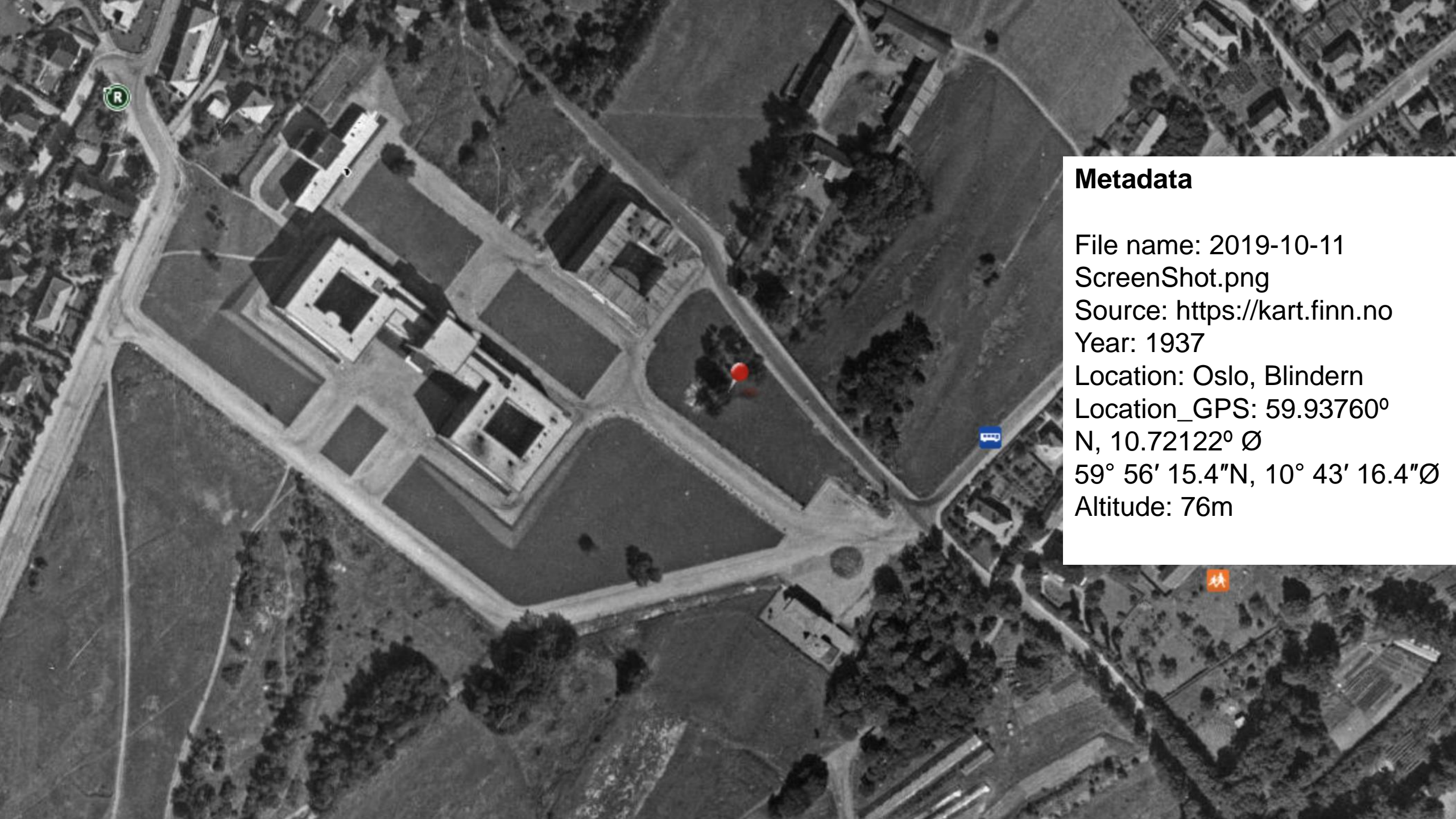
R

Red pin icon

B

A





## Metadata

File name: 2019-10-11

ScreenShot.png

Source: <https://kart.finn.no>

Year: 1937

Location: Oslo, Blindern

Location\_GPS: 59.93760°

N, 10.72122° Ø

59° 56' 15.4"N, 10° 43' 16.4"Ø

Altitude: 76m

# Metadata

- Metadata or “data about data” are **descriptions that facilitate cataloguing data and data discovery.**
- Metadata are intended for machine-reading. When data is submitted to a trusted data repository, the archive generates machine-readable metadata.
- Machine-readable metadata help to explain the purpose, origin, time, location, creator(s), term of use, and access conditions of research data
- Your discipline very likely has standards for metadata!

LEONARD  
COHEN

ALBUM

# I'm Your Man



Leonard Cohen · 1988 · 8 songs, 40 min 48 sec



# TITLE



1 **First We Take Manhattan**  
Leonard Cohen

6:00

2 **Ain't No Cure for Love**  
Leonard Cohen

4:49

# Types of metadata

Descriptive metadata	For finding or understanding a resource
Administrative metadata <ul style="list-style-type: none"><li>- Technical metadata</li><li>- Preservation metadata</li><li>- Rights metadata</li></ul>	<ul style="list-style-type: none"><li>- For decoding and rendering files</li><li>- Long-term management of files</li><li>- Intellectual property rights attached to content</li></ul>
Structural metadata	Relationships of parts of resources to one another
Markup languages	Integrates metadata and flags for other structural or semantic features within content

# Why do we need metadata?

- Discovery
- Aid in identification or understanding of a resource
- Interoperability
- Digital object management
- Preservation
- Navigation within parts of items
  
- **A huge help when making data FAIR!**

F  
indable



A  
ccessible



I  
nteroperable



R  
eusable





**Dublin Core Metadata Element Set** - a set of fifteen “core“  
elements (properties) for describing resources

# Dublin Core Metadata Element Set

- 01. Contributor** – “An entity responsible for making contributions to the resource.”
- 02. Coverage** – “The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.”
- 03. Creator** – “An entity primarily responsible for making the resource.”
- 04. Date** – “A point or period of time associated with an event in the lifecycle of the resource.”
- 05. Description** – “An account of the resource.”
- 06. Format** – “The file format, physical medium, or dimensions of the resource.”
- 07. Identifier** – “An unambiguous reference to the resource within a given context.”
- 08. Language** – “A language of the resource.”
- 09. Publisher** – “An entity responsible for making the resource available.”
- 10. Relation** – “A related resource.”
- 11. Rights** – “Information about rights held in and over the resource.”
- 12. Source** – “A related resource from which the described resource is derived.”
- 13. Subject** – “The topic of the resource.”
- 14. Title** – “A name given to the resource.”
- 15. Type** – “The nature or genre of the resource.”

**Darwin Core** - extension of Dublin Core meant to provide a stable standard reference for sharing information on biological diversity

# Darwin Core: An Evolving Community-Developed Biodiversity Data Standard

John Wieczorek, David Bloom, ... David Vieglais

Record-level Terms	Dublin Core terms, institutions, collections, nature of data record	Simple Darwin Core (flat)
Occurrence	evidence of species in nature, observers, behavior, associated media, references.	
Event	sampling protocols and methods, date, time, field notes	
Location	geography, locality descriptions, spatial data	
Identification	linkage between Taxon and Occurrence	
Taxon	scientific names, vernacular names, names usages, taxon concepts, and the relationships between them	
GeologicalContext	geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy	
ResourceRelationship	explicit relationships between identified resources (e.g., one organism to another, taxon to location, etc.)	Generic Darwin Core (relational)
MeasurementOrFact	measurements, facts, characteristics, assertions, references	

Reset zoom



< Previous

Next >

📄 All

**Figure 2. Darwin Core Categories: Simple Darwin Core is comprised of seven categories of terms (green).**

Show in Context

Download:

PPT

PNG

This subset of Darwin Core terms represents descriptive data about organisms that can be represented in one file with one row... show more →



[Contact](#) [Share](#)



# Replication data for: "Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard"

Version 1.1

Koehl, Jean-Baptiste, 2021, "Replication data for: "Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard"", <https://doi.org/10.18710/IIHGSH>, DataverseNO, V1

 Cite Dataset

[Learn about Data Citation Standards.](#)

## Dataset Metrics

0 Downloads

### Description

High-resolution versions of the figures and supplements of the Koehl (2020) manuscript entitled "Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard", which could not be attached to the manuscript itself due to size limit but that are necessary to identify the main structures and interpretation. (2021-04-07)

### Subject

Earth and Environmental Sciences

### Keyword

Faults, Tectonics, High-resolution figures, Svalbard, Eurekan, Billefjorden Group, Strain partitioning, Bedding-parallel thrust, Décollement and duplexes

### Related Publication

Koehl, J.-B. P.: Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard, Solid Earth, 12, 2021.

[Files](#)

[Metadata](#)

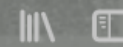
[Terms](#)

[Versions](#)



https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/IIHGSH

90%



## Citation Metadata

### Dataset Persistent ID

doi:10.18710/IIHGSH

### Publication Date

2021-04-07

### Title

Replication data for: "Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard"

### Author

Koehl, Jean-Baptiste (University of Oslo) - ORCID: 0000-0001-7189-1988

### Contact

Use email button above to contact.

Koehl, Jean-Baptiste P. (University of Oslo)

### Description

High-resolution versions of the figures and supplements of the Koehl (2020) manuscript entitled "Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard", which could not be attached to the manuscript itself due to size limit but that are necessary to identify the main structures and interpretation. (2021-04-07)

### Subject

Earth and Environmental Sciences

### Keyword

- Faults
- Tectonics
- High-resolution figures
- Svalbard
- Eurekan
- Billefjorden Group
- Strain partitioning
- Bedding-parallel thrust
- Décollement and duplexes

### Related Publication

Koehl, J.-B. P.: Early Cenozoic Eurekan strain partitioning and decoupling in central Spitsbergen, Svalbard, Solid Earth, 12, 2021.



[Files](#)

[Metadata](#)

[Terms](#)

[Versions](#)

 [Export Metadata ▾](#)


[Citation Metadata ▾](#)

[Geospatial Metadata ▲](#)

<b>Geographic Coverage</b> ?	Norway, Svalbard
<b>Geographic Unit</b> ?	central Spitsbergen
<b>Geographic Bounding Box</b> ?	15.0 20.0 79.0 77.0

# Dataverse –deposit your data

## ^ Enter metadata

 Information about the various metadata fields can be obtained by placing the cursor on the field names (a roll-over window appears). Here is some more information about some of the fields:

### → Title:

- Enter a title for your dataset.
- If your dataset is used in a publication, you may enter the title of the publication, and click on *Add "Replication Data for" to Title*.

### → Author:

- Enter your name as you use it in your publications. We recommend you to add your affiliation as well. For entering co-authors, click on the plus button. We also recommend you to add your ORCID (<https://orcid.org>).

### → Contact:

- Enter a contact email address. Also add the name of the contact person or research group/institution.

### → Description:

- Enter information about the data to be uploaded. Avoid using certain HTML tags and other special characters (e.g. [ or ]). If you need to add paragraphs, add the HTML tags `<p>` and `</p>` around each paragraph.
- If relevant, enter information about the data collection/methodology here.
- If applicable, also enter the publication abstract. The abstract should be entered into a second description field, which can be added by clicking the plus button to the right. NB! If your article is only submitted and not accepted (yet), DO NOT mention the name of the journal it has been submitted to.



# Dataverse –deposit your data

## → **Keyword:**

- Information such as the subject area(s) (e.g. morphology or zoology) and the statistical method(s) may be entered into the keyword field.
- Each keyword needs to be entered separately. Please click the plus button to enter more keywords.
- Vocabulary and Vocabulary URL are not mandatory and may be left empty.

## → **Related Publication:**

- If the files you are depositing are the background data for a publication, you should include a reference to the publication here.
- **Note!** If your manuscript has been submitted for review but has not yet been accepted, DO NOT list the name of the journal or publisher. Instead you may simply write “Submitted for review” or similar.
- **Note!** If the review of your manuscript is going to be double blind (both author and reviewer are anonymous), you must add a note about it in the *Related Publication* field. This way, the curators can assist you in anonymizing the dataset.
- (When adding more than one publication, only the first of them will be visible on the overview page of the dataset. If you don't want to highlight any of the publications in this way, you may add the following text in the first publication field: “Click Metadata tab

# Data Documentation Initiative (DDI) standard

- An international standard for describing the data produced by observational methods in the social, behavioural, economic, and health sciences
- Expressed in XML
- Supports the entire research data lifecycle:
  - For a single data set: [DDI Codebook](#) (DDI-C)
  - For a longitudinal, linked and other complex datasets: [DDI Lifecycle](#) (DDI-L)
  - Common fields include: Title, Alternate Title, Principal Investigator, Funding, Bibliographic Citation, Series Information, Summary, Subject Terms, Geographic Coverage, Time Period, Date of Collection, Unit of Observation, Universe, Data Type, Sampling, Weights, Mode of Collection, Response Rates, Extent of Processing, Restrictions, Version History
  - For interdisciplinary research soon is coming [DDI - Cross Domain Integration](#) (DDI-CDI)

# DDI standard: Example

(Borrowed from [CESSDA Data Management Expert Guide](#))

A dataset in the Finnish Social Science Data Archive (Galanakis, Michail (University of Helsinki): Intercultural Urban Public Space in Toronto 2011-2013 [dataset]. Version 1.0 (2014-02-13). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD2926> ).

[https://services.fsd.tuni.fi/catalogue/FSD2926?tab=summary&lang=en&study\\_language=en](https://services.fsd.tuni.fi/catalogue/FSD2926?tab=summary&lang=en&study_language=en)

Check: «Detailed description»

The machine-readable XML file looks like this:

<https://services.fsd.tuni.fi/catalogue/FSD2926/DDI/FSD2926e.xml>

Display study description in English

## FSD2926 Intercultural Urban Public Space in Toronto 2011-2013

Overview Detailed description Questions Publications Download data

### Authors

- Galanakis, Michail (University of Helsinki)

### Keywords

citizen participation, communities, cultural interaction, cultural pluralism, decision-making, ethnic groups, immigration, politicians, public spaces, services for young people, social inequality, urban development, urban environment, urban sociology, urban spaces

### Abstract

The dataset contains transcripts of interviews conducted mainly in Toronto, Canada, during 2011 and 2012. A few interviews were conducted in Vancouver and Guelph as well. The main themes of the interviews were multiculturalism, interculturalism, diversity and public space, and how the participants' perceptions of interculturalism and public space. The interviewees were professionally or voluntarily involved in the physical or social planning process, in providing services for youth, or in dealing with managing diversity (in policy-making, planning, arts etc). They were community activists, professional designers, managers of public spaces, social services providers, or young persons who represented the users of services aimed at communities.

The interviews were reflective, and questions asked changed according to what the interviewees talked about. The three main research questions were what the participants considered public spaces to be, how they defined interculturalism and, for expert interviews, how they planned/designed for diversity. Toronto is a very multicultural city, and one of the main aims of the study was to learn how Toronto's public space is managed and how public space could be used more creatively for the benefit of diverse groups. Other topics that came up included exclusion of youth, crime, services and facilities for youth, social and educational inequality, unemployment, public transport, street art, safety, police harassment, and privatization of public space.

In addition to 25 one- and two-person interviews, there was one focus group interview of 13 young men and women. Interviewee age ranged from adolescents to senior citizens. Background variables included the interviewee's occupation, gender and age.

The dataset is (B) available for research, teaching and study.

[Download the data](#)

Study description in other languages

- [in Finnish](#)

Related files

- [Data example \(PDF file, in English\)](#)
- [Dataset guide \(PDF file, in English\)](#)

Permanent link to this dataset:

<http://urn.fi/urn:nbn:fi:fsd:T-FSD2926>

Direct link to this tab:

[https://services.fsd.tuni.fi/catalogue/FSD2926?tab=summary&lang=en&study\\_language=en](https://services.fsd.tuni.fi/catalogue/FSD2926?tab=summary&lang=en&study_language=en)

[Study description in machine readable DDI 2.0 format](#)



Metadata record is licensed under a [Creative Commons Attribution 4.0 International license](#).

# Metadata standards overviews

- The Digital Curation Center:

<https://www.dcc.ac.uk/guidance/standards/metadata>

- Research Data Alliance Metadata Directory:

<https://rd-alliance.github.io/metadata-directory/standards/>

**Metadata is still work-in-progress!**

# Design a data file structure

- An important part of the **metadata** is often **embedded into the data file**, e.g. variable names and variable or value labels. Therefore, the **structure** of your data also **contributes to the clarity** of your **data documentation**.

Documentation

# Documentation vs. metadata?

"While data documentation is meant to be read and understood by humans, metadata (which are sometimes a part of the documentation) are primarily meant to be processed by machines."



# Documentation: what, why and how?

- **What** documentation?
- **Why** documentation?
- **How** to create documentation?

# What documentation?

- Systematically documented research data is key to making the data publishable, discoverable, citable, and reusable (and FAIR)
- Clear and detailed documentation **improve the overall data quality**
- It is vital to document **both the study** for which the data has been collected **and the data itself**. These two levels of documentation are called **project-level** and **data-level** documentation

# What documentation?

- The **project-level documentation** explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc.
- **Data-level** or **object-level documentation** provides information at the level of individual objects such as images or variables in a database/table or transcripts, etc.
- It's become a convention to create multiple **README-files**, both for project-level documentation and for data-level documentation

# Document everything your data has been through

- Field journal
- Lab journals and experimental protocols
- Scripts for analysis
- Questionnaires, codebooks, data dictionaries
- Software syntax and output files
- Methodology reports
- Geolocation, orientation (e.g. when collecting a sample)
- Instrument settings and calibration

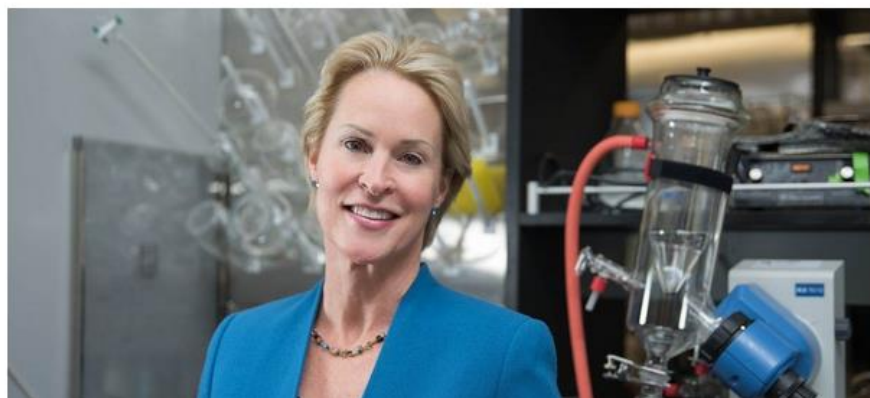


# Why documentation?

- Helps others understand your project and reuse your data



The image is a screenshot of a BBC News article. At the top, there is a navigation bar with the BBC logo, a 'Sign in' button, and links for Home, News, Sport, Reel, Worklife, and Trav. Below this is a red banner with the word 'NEWS' in white. Underneath the banner is a secondary navigation bar with links for Home, Coronavirus, Video, World, UK, Business, Tech, Science, Stories, Entertainment & Arts, and Health. Below that is a third navigation bar with links for World, Africa, Asia, Australia, Europe, Latin America, Middle East, and US & Canada. The main headline reads 'Nobel Prize-winning scientist Frances Arnold retracts paper'. Below the headline is the date '© 3 January 2020' and a small red box with a white arrow pointing left and the text 'Nobel Prize'.



“It has been retracted because the results were not reproducible, and the authors found data missing from a lab notebook.”

<https://www.bbc.com/news/world-us-canada-50989423>

# Why documentation?

- Helps others understand your project and reuse your data
  - all you need to know about your data
  - all your collaborators need to know about your data
  - all you need to know about your collaborators data
  - all anyone else who wants to work with your data needs to know about it
  - more information is better than less
  - structured information is better than unstructured information
  - no documentation is the worst!
  - start early, save time & have less stress
  - re-use templates for good documentation, READMEs, etc.
  - you must have documentation when you archive your data

# How to create documentation?

- README.txt-files:
  - Announce that they are the first file to open when looking through your old data
  - Provide a map for exploring your files
  - Create one README.txt file per folder in as many folders as you can
  - They do not need to be large, but their content should help navigation through digital files and folders
  - a project-level README.txt should give the general project information and a very coarse overview of file contents and locations
  - A data-(object-) level README.txt would be more specific as to what each file contains

# How to create documentation?

- Cornell University template and guide README.txt-files:

<https://data.research.cornell.edu/content/readme>

<https://cornell.app.box.com/v/ReadmeTemplate>

- README.txt-files: some examples

<https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/LNCK8O>

<https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/Q3FZAN>

<https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/TYPJXU>



Break

# File naming & Folder structure

# File naming & Folder structure

- Structuring your data files in folders is important for making it easier to **locate** and **organise** files and versions. A proper folder structure is especially needed when collaborating with others.
- The decision on how to organise your data files depends on the **plan and organisation of the study**. All material relevant to the data should be entered into the data folders, including detailed information on the data collection and data processing procedures.

SvalbardEx271020

File Home Share View

SvalbardEx271020

Name	Date modified	Type	Size
P-10-01	26.10.2020 16:29	File folder	
SEM 2012	26.10.2020 16:25	File folder	
P-10-03-01.tif	11.06.2012 11:17	TIF File	1 880 KB
P-10-03-02b.tif	11.06.2012 11:25	TIF File	1 880 KB
P-10-03-02c.tif	11.06.2012 11:26	TIF File	1 880 KB
P-10-03-05.tif	11.06.2012 11:38	TIF File	1 880 KB
P-10-03-79b.tif	11.06.2012 12:59	TIF File	1 880 KB
P-10-03-92b.tif	11.06.2012 13:26	TIF File	1 880 KB
P-10-03-109.tif	11.06.2012 13:49	TIF File	1 880 KB
P-10-03-119b.tif	11.06.2012 14:02	TIF File	1 880 KB
P-10-03-193b.tif	11.06.2012 15:45	TIF File	1 880 KB
P-10-03-222b.tif	11.06.2012 16:29	TIF File	1 880 KB
P-10-03-222c.tif	11.06.2012 16:30	TIF File	1 880 KB
P-10-03-226b.tif	12.06.2012 09:33	TIF File	1 880 KB
P-10-03-226c.tif	12.06.2012 09:33	TIF File	1 880 KB
Gasser_2014.pdf	27.01.2015 13:11	Adobe Acrobat D...	5 127 KB
Gernigon_and_Bronner_2012.pdf	30.01.2015 15:07	Adobe Acrobat D...	3 445 KB
Gernigon_et_al_2014.pdf	26.01.2015 16:02	Adobe Acrobat D...	54 447 KB
Glorstad_Clark-2010.pdf	15.09.2014 15:06	Adobe Acrobat D...	10 269 KB
Gronlie et al. 1980 Seismic inversion of Bj...	09.10.2012 13:28	Adobe Acrobat D...	643 KB
Gudlaugsson_et_al_1998.pdf	17.10.2014 12:50	Adobe Acrobat D...	3 411 KB
Harland_and_Gayer_1972.pdf	17.10.2014 11:47	Adobe Acrobat D...	1 474 KB
Høy_and_Lundchien_2011_NBarenstSea_...	17.09.2014 14:10	Adobe Acrobat D...	13 600 KB
Isaksen 1996 Organic geocem Bjornoya.p...	09.10.2012 13:24	Adobe Acrobat D...	1 439 KB
Klausen et al 2015 Triassic Snadd in Baren...	30.03.2015 13:15	Adobe Acrobat D...	14 911 KB
Klausen_et_al_2014_Triassic_Snadd_Fm.pdf	08.09.2014 13:20	Adobe Acrobat D...	5 069 KB
Klitzke_et_al_Barents_Sea_Region_2014-pr...	21.01.2015 15:22	Adobe Acrobat D...	2 951 KB
Gabrielsen et al. 1990_Structural_E_BS_NP...	11.03.2013 15:45	Adobe Acrobat D...	9 099 KB
Gac_et_al_2013_ultra_deep_EBB.pdf	20.01.2015 12:15	Adobe Acrobat D...	1 008 KB
Abstract.doc	15.09.2014 16:04	Microsoft Word 9...	234 KB
Copy of NuAge_090214-120115-run1.xls	22.01.2015 13:06	Microsoft Excel 97...	1 390 KB
CPP_Svalbard_Bjornoya_Copy of CUMUL...	17.12.2014 15:35	Microsoft Excel 97...	2 429 KB
2013 Wintermeeting poster-1_Edina_Final...	07.01.2013 15:12	Adobe Acrobat D...	130 KB
ICPMS data	26.10.2020 16:31	File folder	
Franz Josef presentation.pptx	09.01.2013 12:06	Microsoft PowerP...	11 334 KB

Example of a research project **without** data file structure – Edina Pózer ©

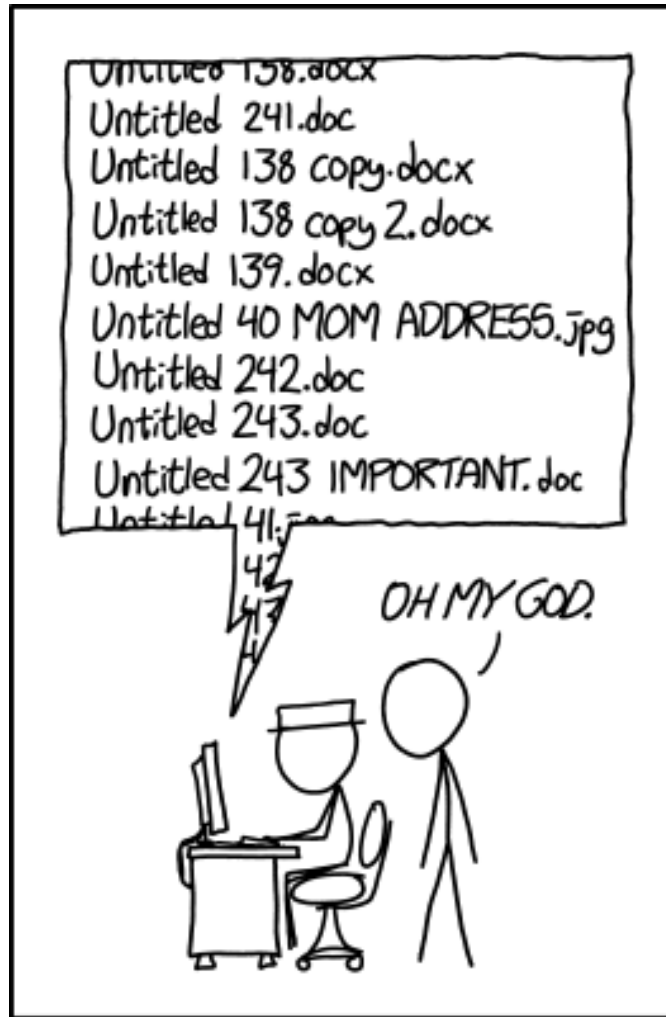
# How it could look like:

```
project_name/
├── README.md           # overview of the project
├── data/               # data files used in the project
│   ├── README.md      # describes where data came from
│   └── sub-folder/     # may contain subdirectories
├── processed_data/    # intermediate files from the analysis
├── manuscript/        # manuscript describing the results
├── results/           # results of the analysis (data, tables, figures)
├── src/               # contains all code in the project
│   ├── LICENSE         # license for your code
│   ├── requirements.txt # software requirements and dependencies
│   └── ...
└── doc/               # documentation for your project
    ├── index.rst
    └── ...
```

*Research project with a proper data file structure. Image taken from CodeRefinery, Lesson on Reproducible Research. Shared under CC-BY 4.0.*

# File naming conventions (FNC)

- Short names (but long enough that they still make sense)
- The most general information first, then add details to the name
- Underscore to separate words, DO NOT use space in file names!
- Dates backwards (YYYYMMDD)
- Numbers (e.g. version number) should have the same number of digits, use e.g. 01, not just 1.
- Version number at the end



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

# Avoid using the following characters in Folder and File names:

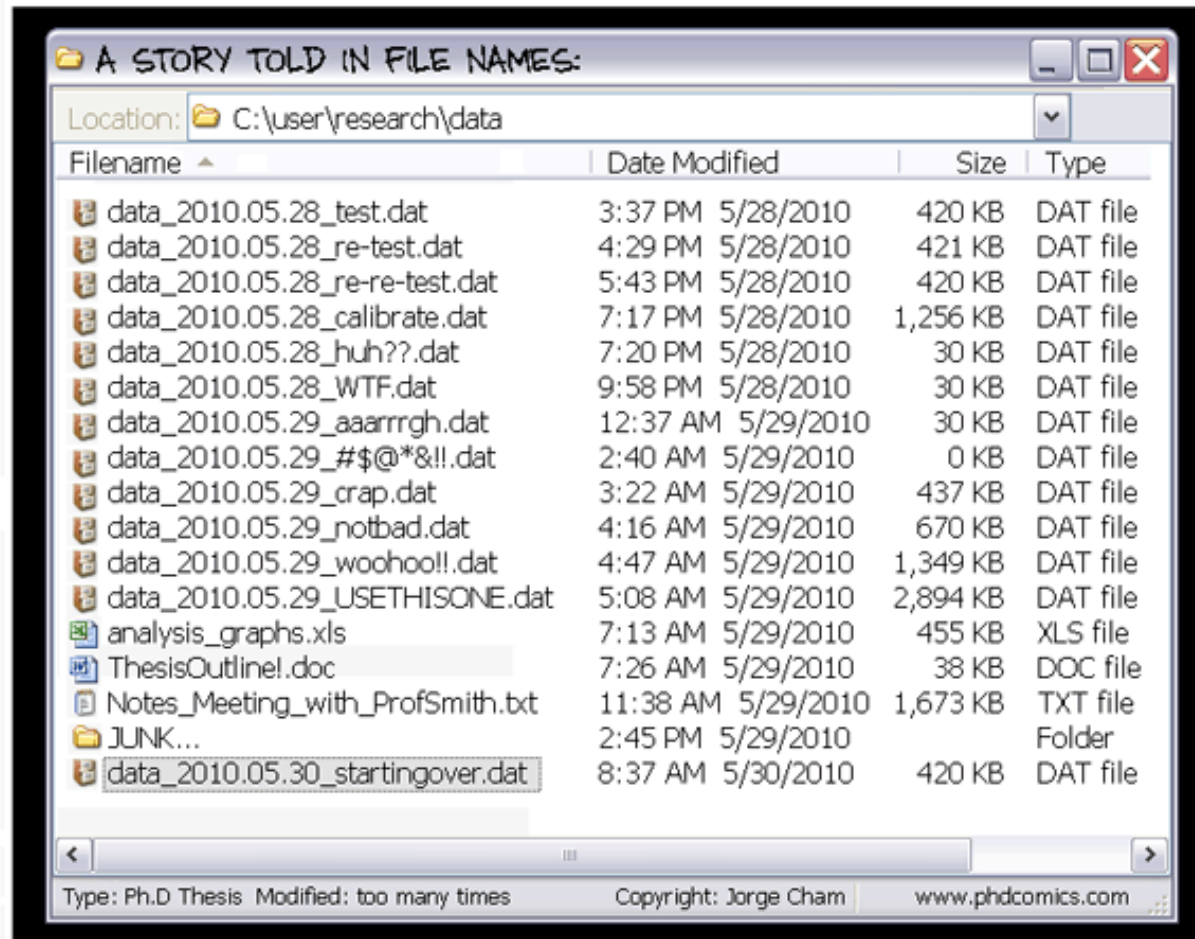
#	pound	<	left angle bracket	{	left curly bracket
%	percent	>	right angle bracket	}	right curly bracket
&	ampersand	/	forward slash	*	asterisk
\	back slash		blank spaces	?	question mark
\$	dollar sign	'	single quotes	=	equal sign
!	exclamation point	"	double quotes		



# Also, keep these rules in mind:

- Don't start or end your filename with a space, period, hyphen, or underline
- Keep your filenames to a reasonable length
- Most operating systems are case sensitive; always use lowercase

# Tidy every once in a while



# Versioning

- Refers to saving new copies of your files when you make changes so that later you can go back and retrieve specific versions of your files
  - DataFileName\_1.0 = original document
  - DataFileName\_1.1 = original document with minor revisions
  - DataFileName\_2.0 = document with substantial revisions

# Versioning

- Refers to saving new copies of your files when you make changes so that later you can go back and retrieve specific versions of your files

- image1\_v1.jpg
- image1\_v2.jpg
- image2\_v1.jpg
- image2\_v2.jpg
- ...

- image1\_v1.jpg
- image1\_v10.jpg
- image1\_v2.jpg
- ...

# Versioning

- Refers to saving new copies of your files when you make changes so that later you can go back and retrieve specific versions of your files

- image1\_20151021
- image1\_20151214
- image1\_20160123
- ...

- dataset1\_20160402\_KES
- dataset1\_20160301\_WTC
- dataset1\_20160814\_GSC
- ...

# Version control

- Perfect for collaboration
- Tracking and managing changes to a file or set of files over time
- You will be able to recall any version at any time
- It can be used for documents, software development, large websites etc.

# Version control system

- Git: Free and open source version control system



# Version control system

- Git: Free and open source version control system
- GitHub: is an internet hosting service for software development and version control using Git





# Sustainable file formats

- When your project is finished and you plan on archiving your data, it is widely recommended to transfer data and accompanying files into a **more sustainable format**.
- Trusted data archives often have guidelines for this, e.g. the recommendations of Data Archiving and Networked Services (DANS) and DataverseNO:
  - <https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>
  - <https://site.uit.no/dataverseno/deposit/prepare/#preferred-file-formats>

# Services and tools for RDM support:

- Research support and IT staff at your faculty/institute
- Subject librarian for your discipline
- [Research data management website](#)
- Digital Scholarship Center (DSC)  
[research-data@uio.no](mailto:research-data@uio.no)
- Norwegian Centre for Research Data (NSD, now Sikt)



Questions?  
Comments?

# Upcoming sessions

## Data classification and storage selection

Mar. 11, 9:00 AM, Zoom

## Copyright and licensing

Mar. 15, 9:00 AM, Zoom

## Sharing and archiving research data

Mar. 17, 9:00 AM, Zoom

## Finding and reusing research data

Mar. 18, 9:00 AM, Zoom

# Digital Scholarship Center

[Norwegian  
version of this  
page](#)

UiO's Digital Scholarship Center (DSC) can help you take advantage of digital tools and methods in your research. We offer support with data management, visualization, digital research methods, and digital skill development.



## Research Data Management at UiO

Having proper Research Data Management (RDM) routines will help you save time, enables sharing and reuse of data, and promotes open science. However, RDM in a digital context can be challenging and complex. We offer guidance and consultation on RDM topics including:

- data management plans (DMP)
- classification and storage
- organization, documentation, and metadata
- legal questions (i.e. privacy, copyright, and licensing)
- archiving and sharing

## Need advice?

Visit us at the room 230, in the 2nd floor of Niels Henrik Abels hus.

Book an appointment



Or send us an email to [research-data@uio.no](mailto:research-data@uio.no)

## Digital research methods

On these pages you will find overviews and user guides on digital research methods for researchers and students. There will be information about relevant digital research methods in many subject areas.

Currently, we have information about text mining - or digital research methods for those working with text data.

## Digital scholarship communities

Interested in teaching or helping at workshops? Join us in:

→ [Carpentries@UiO](mailto:Carpentries@UiO)

→ [CodeRefinery - Get Involved](#)

Do you have any suggestions for workshops or seminars or questions about the workshops? Send an email to [digitalscholarship@ub.uio.no](mailto:digitalscholarship@ub.uio.no)

# Carpentry: Training on foundational coding and data science skills

At University of Oslo (UiO), Carpentries workshops based on Software Carpentry, Data Carpentry, and Library Carpentry lessons are offered to facilitate sharing and re-using of code and data among graduates and researchers from all faculties and units.

# Open and reproducible research

More and more researchers across disciplines are preregistering their hypotheses, methods, and analysis plans before they collect or analyze the data to increase research transparency and visibility. This course will give an introduction to preregistration and a step-by-step guide on how to do it yourself.

## Introduction to open and reproducible research

May 11, 9:00 AM, Zoom

## Preregistration of research studies

May 13, 9:00 AM, Zoom

## Preregistration on Open Science Framework (OSF)

May 19, 12:00 PM, Zoom

## R Markdown: Writing Reproducible Research Papers with R

May 24, 9:00 AM, Undervisningsrom 1, 3rd floor Georg Sverdrups hus



# OA-publishing and research visibility

Keep copyrights to your own work by publishing in Open Access journals. Find out your research impact. The course gives you practical support in publishing process and making your research visible.

## OA-publishing and research visibility

Mar. 23, 9:00 AM, Zoom

## OA-publishing and research visibility

Apr. 26, 9:00 AM, Zoom

# Open Science Lunch

Each last Thursday of the month at 12:00 we invite you to join us virtually for an online open lunch to hear about how to make your research more open. We will discuss research transparency and visibility, open publishing, data sharing, and more!

## Upcoming



Time and place: Mar. 31, 2022 12:00 PM–1:00 PM, Zoom, [Add to calendar](#)

### Open Science Lunch: Building and running a service for sharing curated research data

Learn about how data sharing services are built and run based on an example from EBRAINS Research Infrastructure.



Time and place: Apr. 28, 2022 12:00 PM–1:00 PM, Zoom, [Add to calendar](#)

### Open Science Lunch: Open your research to the public

Learn about how you can conduct open science projects in the wild based on an example from MusicLab!



# Journal Clubs

## Our mailing list:

<https://sympa.uio.no/uio.no/subscribe/open-science-oslo>

# University of Oslo

## Welcome

Our journal club is open to both staff and students at UiO across all departments. Everyone is welcome to join us - whether you are an enthusiast, a skeptic, or a cautious explorer. Feel free to get in touch if you are interested to participate or contribute!

## Our mission

- provide an informal and friendly platform for discussions about meta scientific topics
- help each other get familiarized with open science practices (e.g., pre-registrations, sharing data, sharing preprints, etc.)
- connect students and researchers from various disciplines who are interested in meta science.

## Format


Before each meeting, we read an article on meta scientific topics, which we then discuss during the meeting. At each meeting, a different discussion leader will begin by providing a short overview of the paper and facilitate discussion throughout the meeting. Anyone can propose a possible paper or topic to present if we have not covered it already. Grab a cup of tea (coffee?) and join us! Due to the current pandemic, all sessions will be held online (Zoom) for the time being.


 Universitetet i Oslo

 [osf.io/mvx54](https://osf.io/mvx54)

Z collection

 [timo.b.roettger@gmail.com](mailto:timo.b.roettger@gmail.com)

 Timo B. Roettger

 Agata Bochynska



Materials developed as a part of the *Skills development for research data* project:  
<https://www.ub.uio.no/english/about/projects/rdm-skills/>

# Thank you!

Contact us at:

[research-data@uio.no](mailto:research-data@uio.no)

# Sources

- Deutz DB, Buss MCH, Hansen JS, Hansen KK, Kjellmann KG, Larsen AV, Vlachos E, Holmstrand KF (2020). *How to FAIR: a Danish website to guide researchers on making research data more FAIR*. <https://howtofair.dk/how-to-fair/>
- CESSDA Training Team (2017 - 2020). *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. Retrieved from <https://www.cessda.eu/DMGuide>
- *Research Data MANTRA* [online course, <https://mantra.edina.ac.uk/> ] by the Research Data Service, University of Edinburgh
- Rockenberger A (2020) *Shut Up And Write Documentation: README.txt*. Retrieved from: <https://zenodo.org/record/3778273>
- K. Briney (2014) *Wrapping Up A Project*. Retrieved from: <http://dataabinitio.com/?p=344>
- *Guide to writing "readme" style metadata* by the Research Data management Service Group, Cornell University. Retrieved from: <https://data.research.cornell.edu/content/readme> ]
- Riley J (2017). *Understanding Metadata: What is Metadata, and What is it For?: A Primer*. National Information Standards Organization (NISO). <https://www.niso.org/publications/understanding-metadata-2017>
- DCMI Metadata Terms (2020). <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- Wiczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>
- Data Archiving and Networked Services (DANS). *File formats*. Retrieved from: <https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>