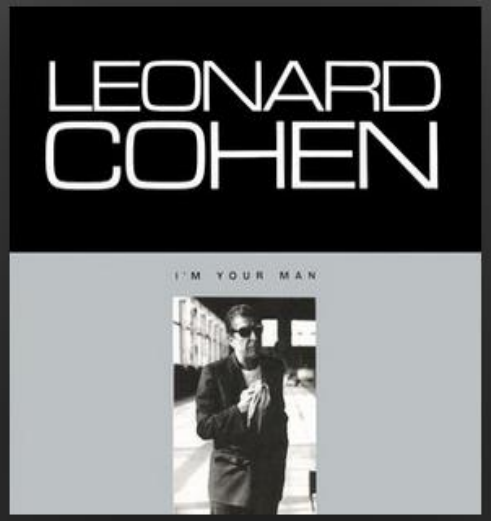
UiO : University of Oslo

Introduction to metadata

Why is it relevant for open and reproducible science?

Agata Bochynska and Elin Stangeland
Open Research group, University of Oslo Library

Metadata is data about data



ALBUM

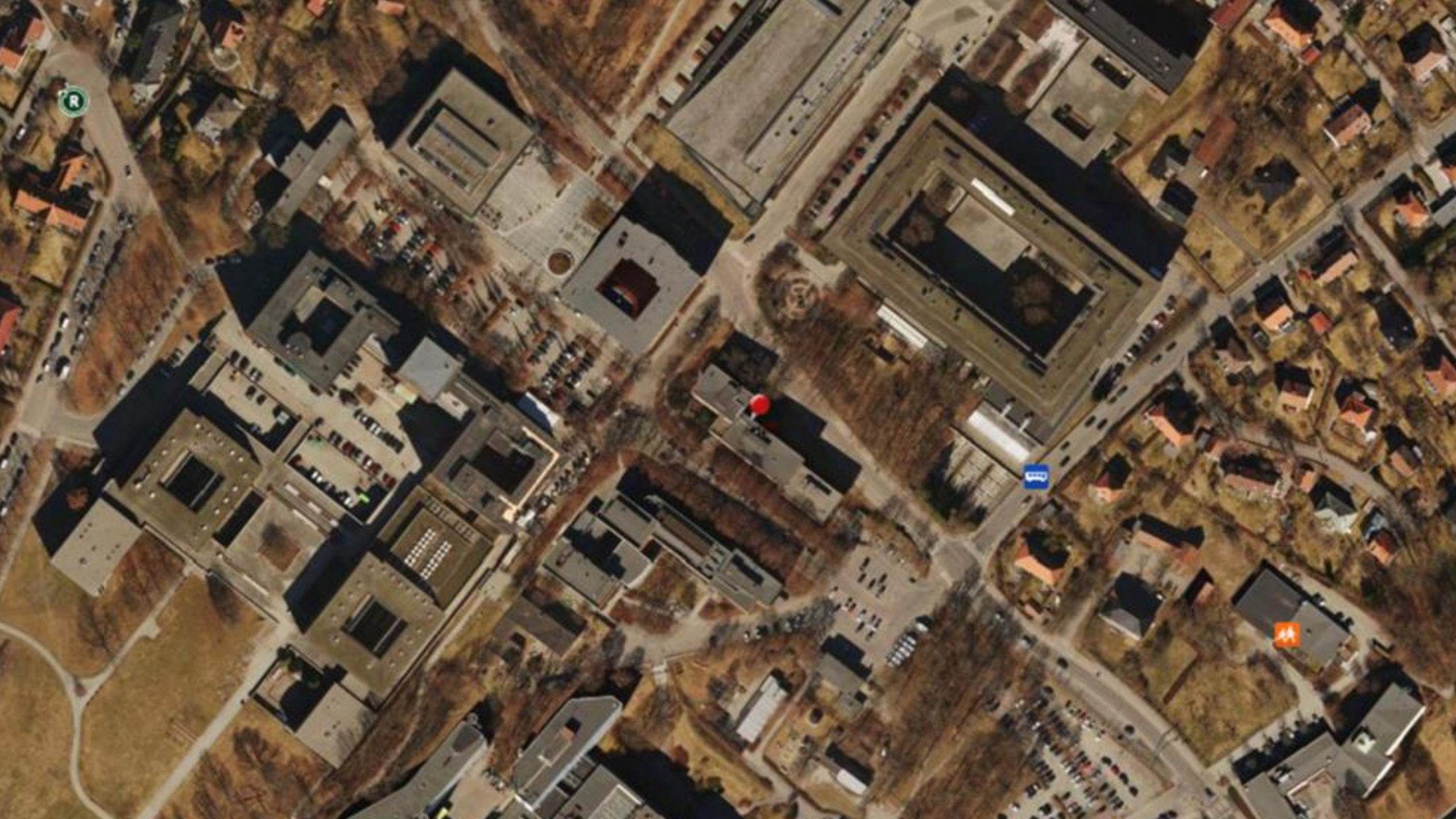
I'm Your Man

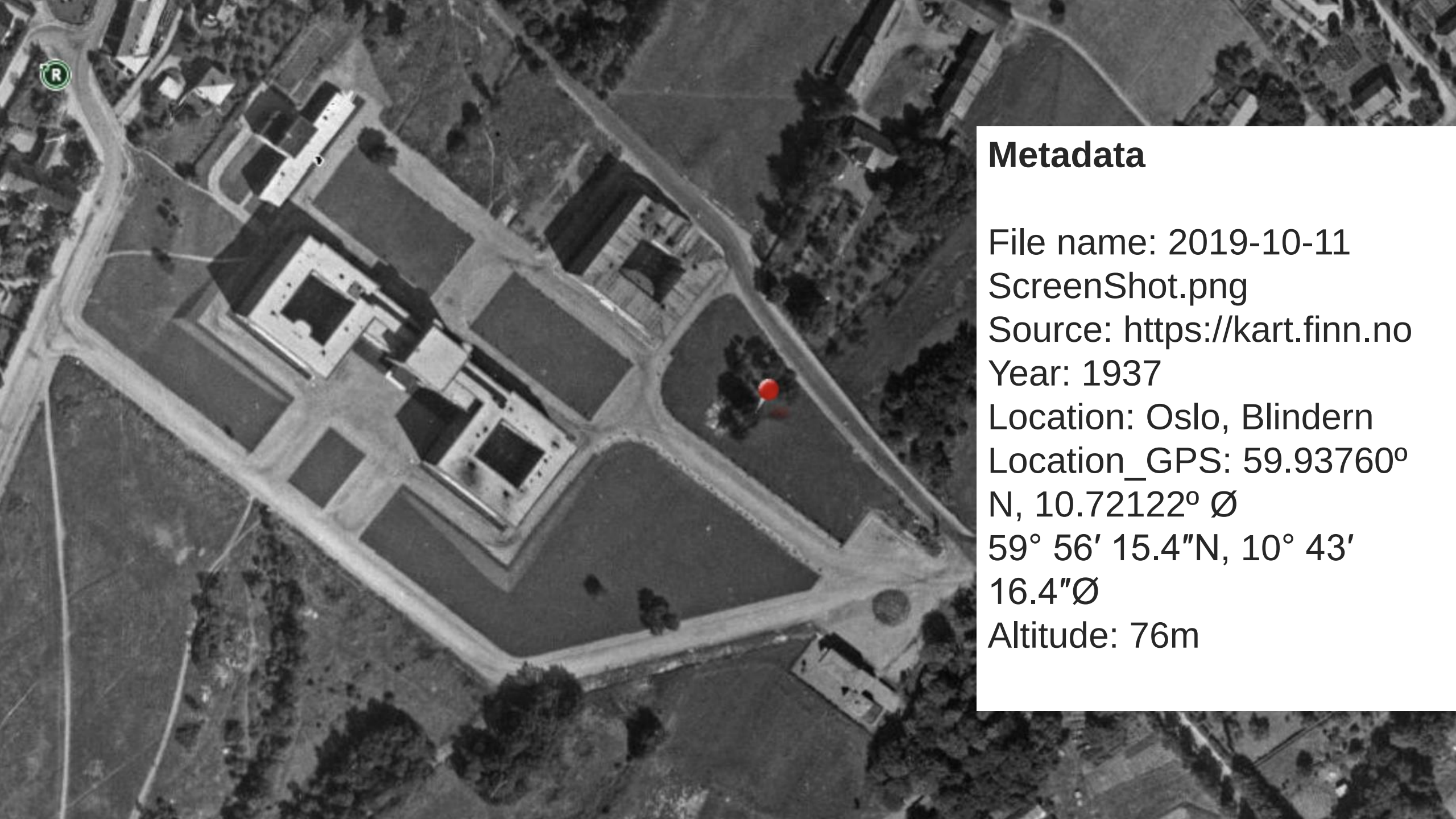
Leonard Cohen • 1988 • 8 songs, 40 min 48 sec

#	TITLE	
1	First We Take Manhattan Leonard Cohen	6:00
2	Ain't No Cure for Love Leonard Cohen	4:49

Metadata can give us critical information about a **data point**, without even seeing the data point itself.

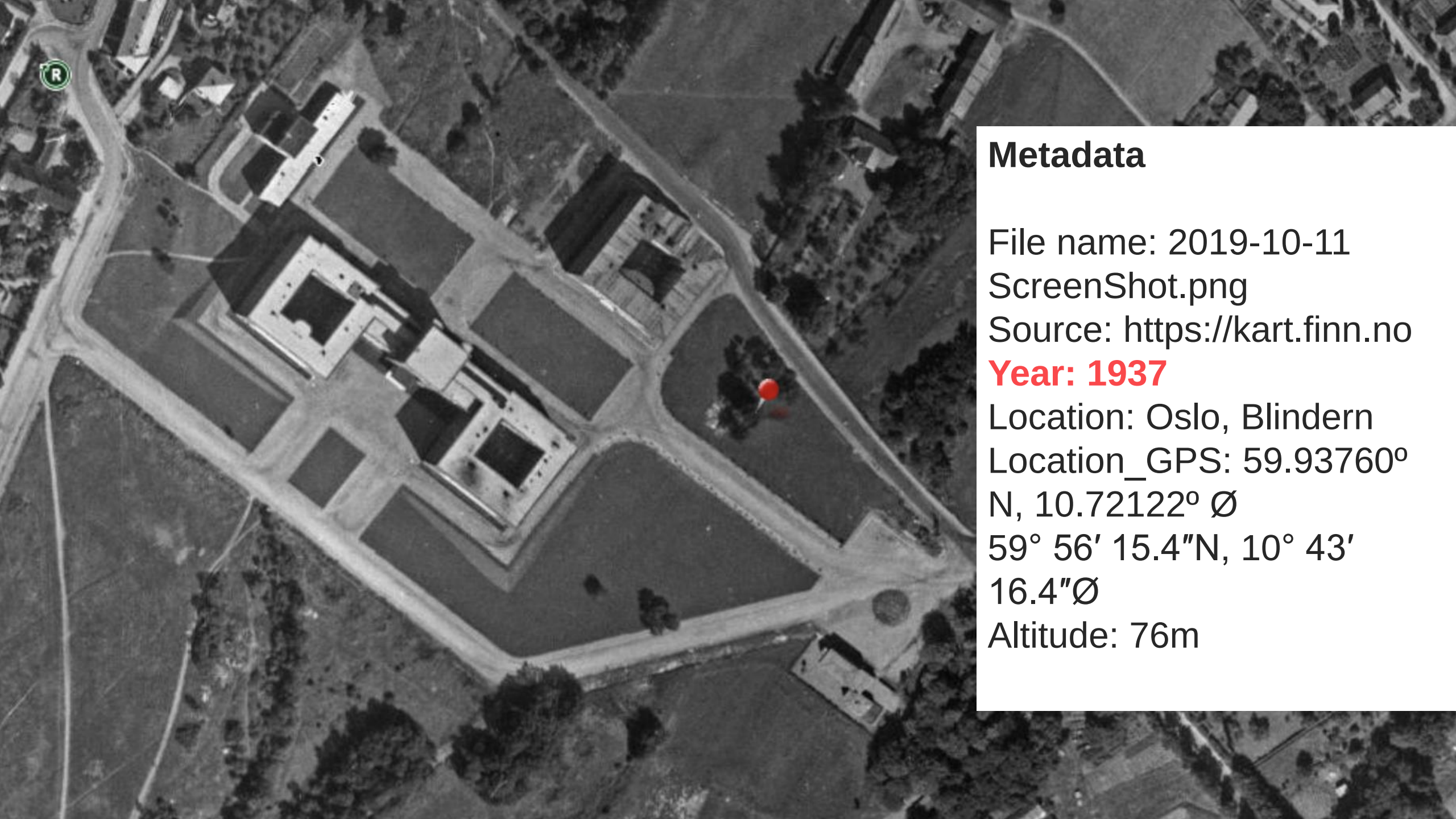
But also, access to a data point **without metadata** is often not enough.





Metadata

File name: 2019-10-11 ScreenShot.png
Source: https://kart.finn.no
Year: 1937
Location: Oslo, Blindern
Location_GPS: 59.93760º N, 10.72122º Ø
59° 56′ 15.4″N, 10° 43′ 16.4″Ø
Altitude: 76m



Metadata

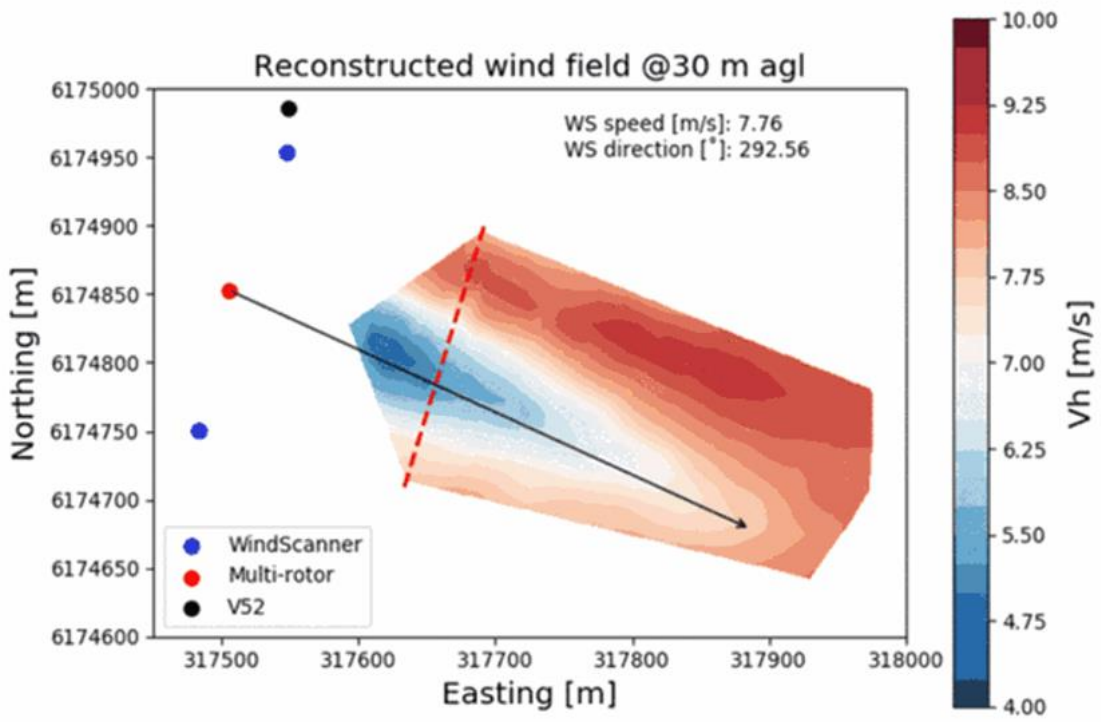
File name: 2019-10-11 ScreenShot.png
Source: https://kart.finn.no
Year: 1937
Location: Oslo, Blindern
Location_GPS: 59.93760º N, 10.72122º Ø
59° 56′ 15.4″N, 10° 43′ 16.4″Ø
Altitude: 76m

Administrative metadata are data about a project or resource that are relevant for managing it.

Descriptive or citation metadata are data about a dataset or resource that allow people to discover and identify it.

Structural metadata are data about how a dataset or resource came about, but also how it is internally structured.

https://www.howtofair.dk/how-to-fair/metadata/

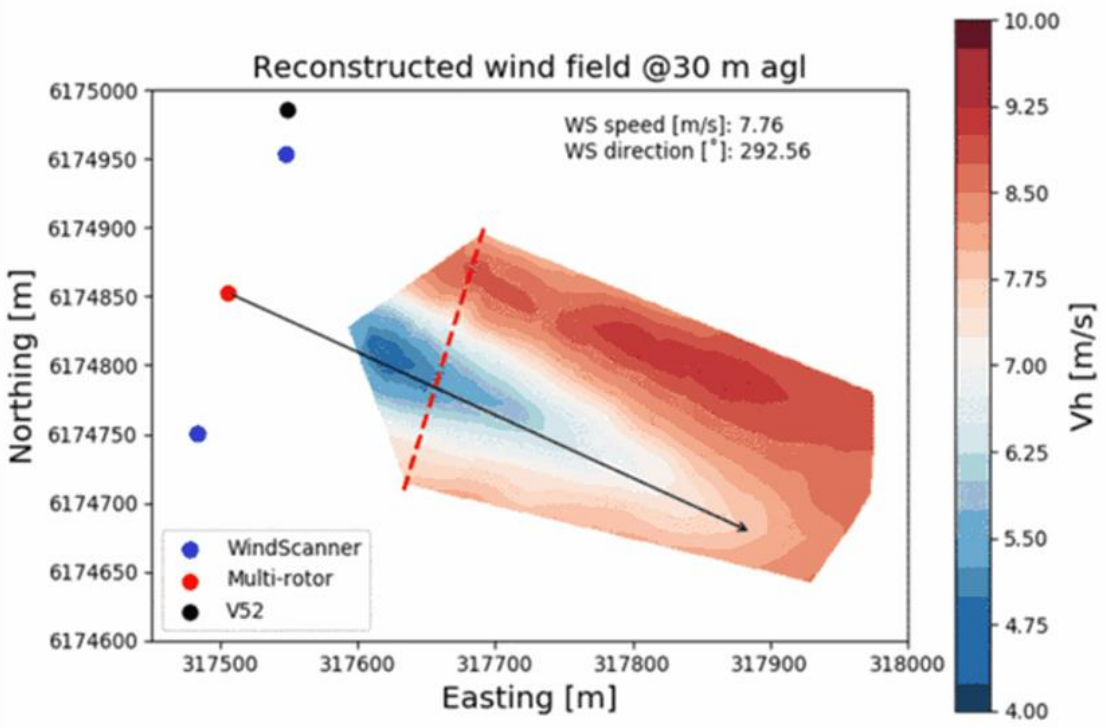


Metadata type:

Descriptive

Title:	Wake plot of wind turbine
Keyword:	wake plot, multi-rotor, wind
Creator:	Nikola Vasiljević
Location:	Risø, Denmark
Date:	14 August 2018
Wind turbine:	Vestas Multi-rotor
Instrument:	Long-range windScanner
Usage rights:	Creative Commons
Project ID:	Vestas_DTU_multi_2018
Project period:	01 November 2017 – 31 July 2019

https://www.howtofair.dk/how-to-fair/metadata/

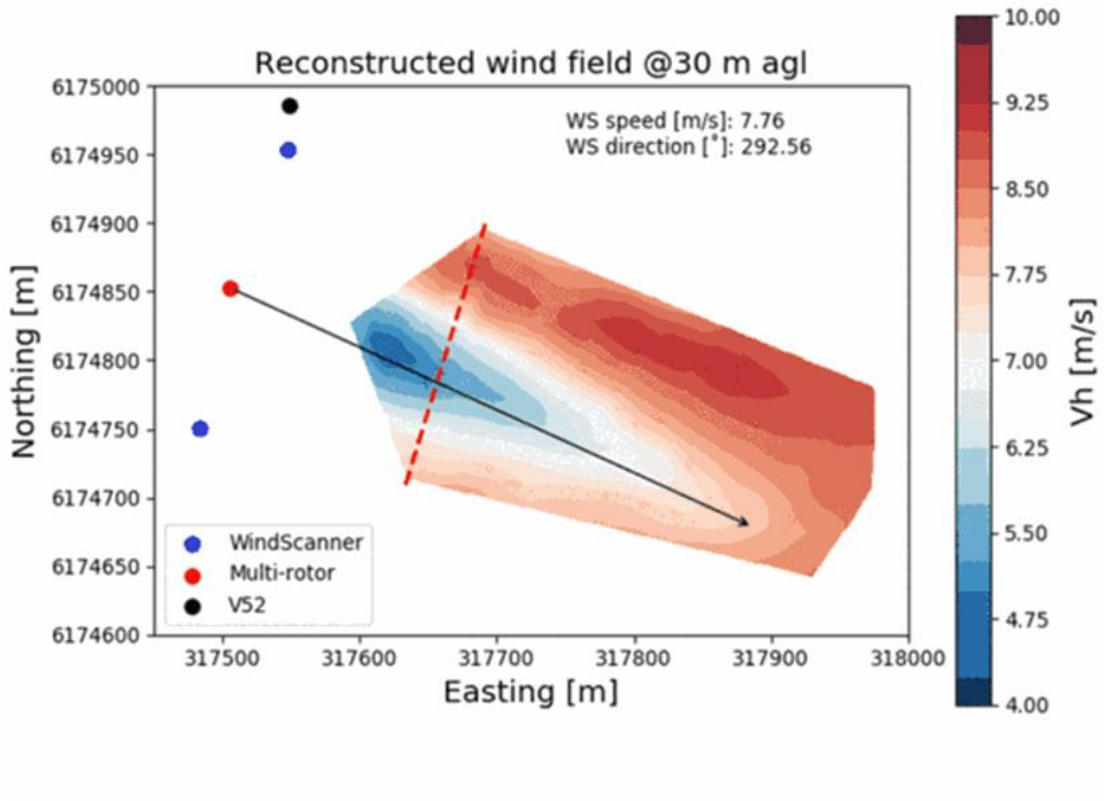


Metadata type:

Administrative

Title:	Wake plot of wind turbine
Keyword:	wake plot, multi-rotor, wind
Creator:	Nikola Vasiljević
Location:	Risø, Denmark
Date:	14 August 2018
Wind turbine:	Vestas Multi-rotor
Instrument:	Long-range windScanner
Usage rights:	Creative Commons
Project ID:	Vestas_DTU_multi_2018
Project period:	01 November 2017 – 31 July 2019

https://www.howtofair.dk/how-to-fair/metadata/



Metadata type:

Structural

Title:	Wake plot of wind turbine
Keyword:	wake plot, multi-rotor, wind
Creator:	Nikola Vasiljević
Location:	Risø, Denmark
Date:	14 August 2018
Wind turbine:	Vestas Multi-rotor
Instrument:	Long-range windScanner
Usage rights:	Creative Commons
Project ID:	Vestas_DTU_multi_2018
Project period:	01 November 2017 – 31 July 2019

https://www.howtofair.dk/how-to-fair/metadata/

innovation in metadata design, implementation & best practice

Dublin Core™ Metadata Initiative

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.

https://www.dublincore.org/specifications/dublin-core/dcmi-terms

Your discipline very likely has its own **metadata standards**

Engineering

CIF (Crystallographic Information Framework) Edit

A well-established standard file structure for the archiving and distribution of crystallographic information, CIF is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals.

Sponsored by the International Union of Crystallography, the current standard dates from 1997. As of July 2011, a new version of the CIF standard is under consideration.

CSMD (Core Scientific Metadata Model) Edit

A study-data oriented model, primarily in support of the ICAT data managment infrastructure software. The CSMD is designed to support data collected within a large-scale facility's scientific workflow; however the model is also designed to be generic across scientific disciplines.

Sponsored by the Science and Technologies Facilities Council, the latest full specification available is v 4.0, from 2013.

ISA-Tab Edit

The Investigation/Study/Assay (ISA) tab-delimited (TAB) format is a general purpose framework with which to collect and communicate complex metadata (i.e. sample characteristics, technologies used, type of measurements made) from 'omics-based' experiments employing a combination of technologies.

Created by core developers from the University of Oxford, ISA-TAB v1.0 was released in November 2008.

MIBBI (Minimum Information for Biological and Biomedical Investigations) Edit

A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an intercompatible, extensible community of standards.

The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.

NeXus Edit

NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations. NeXus is built on top of the container format HDF5, and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined domain-specific field names.

Life Sciences

ABCD (Access to Biological Collection Data) Edit

The Access to Biological Collections Data (ABCD) Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.

Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.

Darwin Core Edit

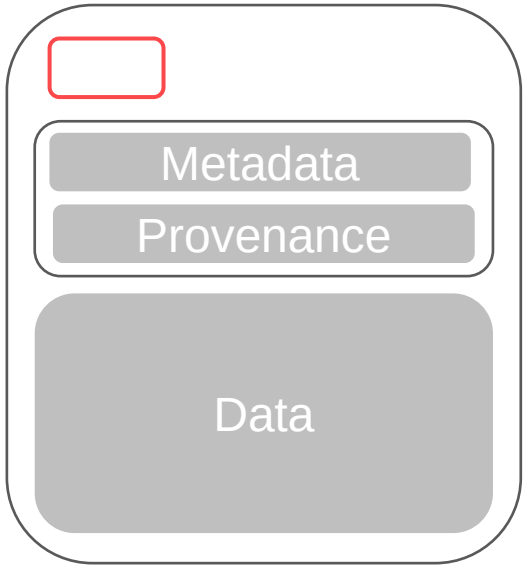
A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to

Findable Accessible Interoperable Reusable

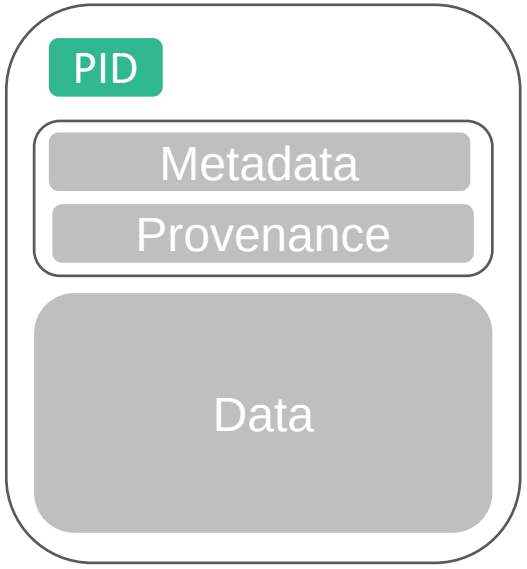
As open as possible,
as closed as necessary

Levels of FAIR

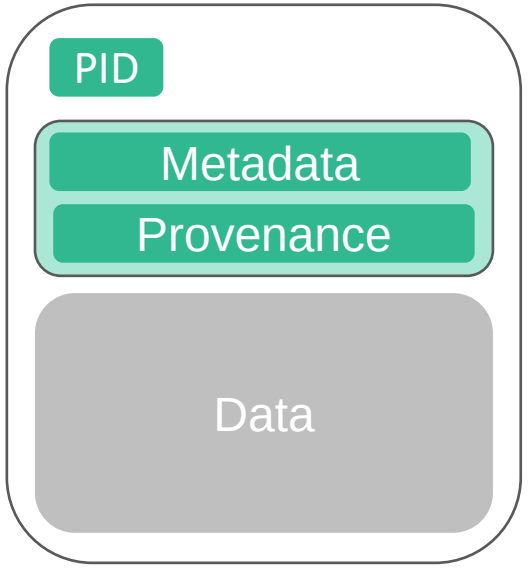
Totally UNFAIR



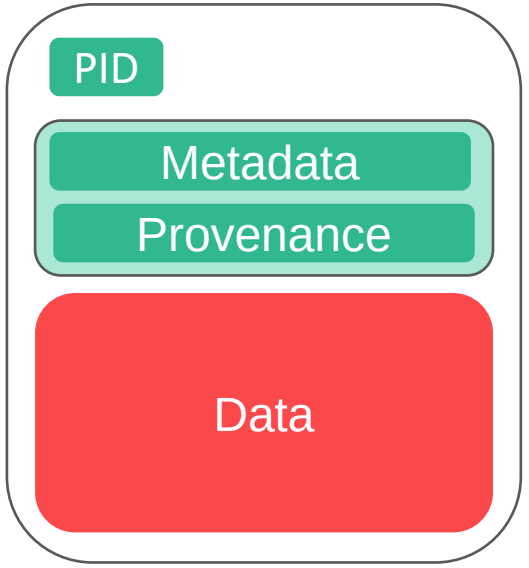
Findable
Usable for humans



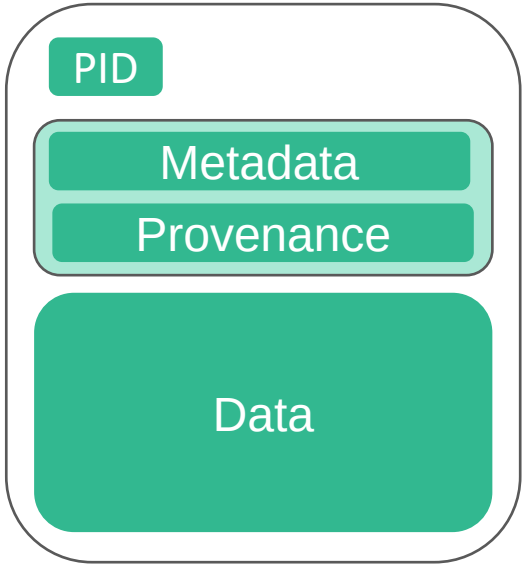
FAIR metadata



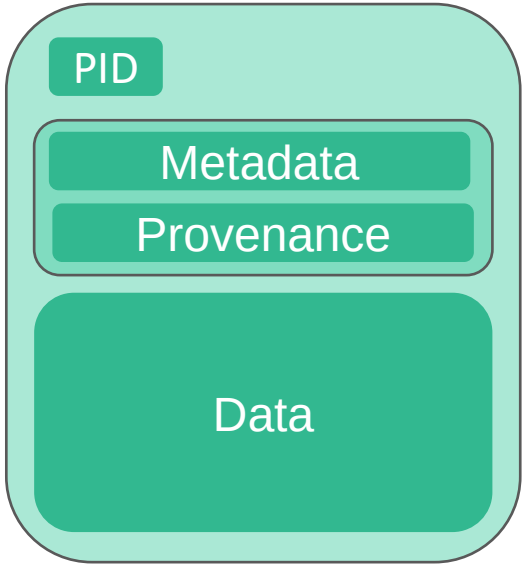
FAIR data
Restricted access



FAIR data
Open access



FAIR data
Open access and functionally linked



Adapted from:
https://www.force11.org/

Open science

Open science means transparency and knowledge-sharing in research processes to make knowledge accessible across academic groups, sectors and national boundaries. The concept of open science encompasses the entire research process […].

Source: The Research Council of Norway (2020). *Policy for open science.*

Image by: opensourceway

https://rdmkit.elixir-europe.org/

Metadata supports **data discovery and reuse**, contributing to scientific reproducibility.

https://rdmkit.elixir-europe.org/

…but it also supports **your data workflows** and **your data reuse**.

UiO : University of Oslo

Thank you!

Any questions?

Agata Bochynska and Elin Stangeland
Open Research group, University of Oslo Library