CHESS Course: An Introduction to FAIR Data Management for Geoscientists, Syllabus

Day 1 - Session 2

The basics: data and metadata

NorDataNet
Norwegian Scientific Data Network

Session Topics

- What are data? What are metadata?
- Discovery, site, and use metadata.
- Which metadata do you need to record?
- What is a metadata standard?
- What is provenance?
- How to record various types of metadata.
- Metadata templates (Arven etter Nansen, EBAS)
- Gap handling for metadata (missing elements).

NorDataNet
Norwegian Scientific Data Network

Learning Objectives

- Discriminate between different types of metadata
- Identify the values of metadata
- Recognize the value of using standards
- Identify what metadata are relevant for a dataset
- How to write your own metadata record

NorDataNet
Norwegian Scientific Data Network

Data and Metadata

The concept of data and metadata can be at times unclear

- Data is a collection of information, such as observations, measurements, computations of models etc…
  - It can be used to analyze trends and patterns, to extract and visualize actual values of variables
- Metadata, on the other hands, is data about the data, i.e. they extensively provide description about the data they are attached to
  - It gives the necessary context to the user to be able to access, understand and use the data correctly
  - Several types of metadata can and should be used to describe the data

NorDataNet
Norwegian Scientific Data Network

Temperature
31.5

Has anyone checked the quality of this value?

For what purpose?

Of what?

Collected when?

Precision/accuracy?

Location?

Temperature

31.5

In what units?

According to whom?

AKA – T, Temp, degC, C, °F… lots of different names

Is this value averaged? Calculated?

When was the sensor last cleaned/calibrated?

Collected how?

Metadata help us address these issues

- Metadata
  - Data about data
  - Info for someone unfamiliar with your data and possibly your research
  - The help to FIND, EVALUATE, UNDERSTAND and REUSE data

- Delivering metadata using standards:
  - consistent
  - digital
  - written for people
  - such that computers are happy

What is in a metadata standard?

- A metadata standard is made up of defined elements, including the type of information the user should enter (e.g. text, numbers, date).
- Examples of elements include Title, Abstract, Keyword, Online Link
- Multiple standard exists and they are linked to the type of metadata they address and the communities they target
- Terminology for the same concepts may vary across standards (values of mapping)

Types of metadata

Type	Purpose	Description	Examples
Discovery metadata	Used to find relevant data	Discovery metadata are also called index metadata and are a digital version of the library index card. It describes who did what, where and when, how to access data and potential constraints on the data. It shall also link to further information on the data like site metadata. GCW is required to expose this information through WMO Information System as well. Discovery metadata are thus WIS metadata, although the GCW portal can translate to WIS for those not using WMO standards directly.	ISO19115 GCMD DIF
Use metadata	Used to understand data found	Use metadata are describing the actual content of a dataset and how it is encoded. The purpose is to enable the user to understand the data without any further communication. It describes content of variables using standardised vocabularies, units of variable, encoding of missing values, map projections etc.	Climate and Forecast Convention BUFR GRIB Darwin Core Archive
Configuration metadata	Used to tune portal services for datasets for users.	Configuration metadata are used to improve the services offered through a portal to the user community. This can be e.g. how to best visualise a product. This information is maintained by the GCW portal and is not covered by discovery or use metadata standards.	Used locally by data centres
Site metadata	Used to understand data found	Site metadata are used to describe the context of observational data. It describes the location of an observation, the instrumentation, procedures etc. To a certain extent it overlaps with discovery metadata, but more so it really extends discovery metadata. Site metadata can be used for observation network design.	WIGOS OGC O&M

NorDataNet
Norwegian Scientific Data Network

Discovery metadata

For proper search of data

- A data set description that provides information to determine if a particular data set meets the users' needs.
- Provides essential information to enable a user to find
  - if a particular dataset exists
  - the data's location
  - ownership
  - how to obtain further information.
- Includes the science discipline of the data, data location, spatial coverage, data provider, data resolution, data quality, etc.

MD_Metadata: Something somewhere that can be accessed under certain conditions and for which someone is responsible.

NorDataNet
Norwegian Scientific Data Network

Documentations

Table 2: GCMD DIF elements.

GCMD DIF elements	Description	GCMD
Entry_ID	The **<Entry_ID>** is the unique document identifier of the metadata record. The **<Entry_ID>** is determined by the metadata author or data center contact personnel and may be identical to identifiers used by the data provider's data center or organization. For example, the National Snow and Ice Data Center (NSIDC) Distributed Active Archive Center (DAAC) identifies their metadata records as *NSIDC-xxxx*, where *xxxx* is a numerical designator. Also, the identifier is case insensitive meaning nsidc-xxxx and NSIDC-xxx refer to the same metadata record.	MU
Entry_Title	The <Entry_Title> is the title of the data set described by the metadata.	MU
Parameters (Science Keywords)	The **<Parameters>** field allows for the specification of Earth science keywords that are representative of the data set being described. These keywords are important for the precise search and retrieval of information from the GCMD. The author must select these keywords from the controlled set of science keywords. The **<Parameters>** field consists of a 7-level hierarchical classification of science keywords	MC
ISO Topic Category	The **<ISO_Topic_Category>** field is used to identify the keywords in the ISO 19115 - Geographic Information Metadata (http://www.isotc211.org/) Topic Category Code List. It is a high-level geographic data thematic classification to assist in the grouping and search of available geographic data sets.	MC
Data Center	The **<Data Center>** is the data center, organization, or institution responsible for distributing the data.	M
Summary	The **<Summary>** field provides a brief description of the data set along with the purpose of the data. This allows potential users to determine if the data set is useful for their needs.	MU
Metadata Name	The ISO 19115 **<Metadata_Name>** field is used to identify the current DIF standard name.	MU
Metadata Version	The **<Metadata_Version>** field is used to identify the current DIF metadata standard.	MU
Data Set Citation	The **<Data_Set_Citation>** field allows the author to properly cite the data set producer.	R
Personnel	**<Personnel>** defines the point of contact for more information about the data set or the metadata.	R
Instrument	The Instrument or **<Sensor_Name>** is the name of the	RC

https://earthdata.nasa.gov/esdis/eso/standards-and-references/directory-interchange-format-dif-standard

Example DIF-9

In the schema for a standard you can find the specifications of the elements and their occurance

master ··· DIF Schemas / 9.x / **dif_v9.9.3.xsd**

Source view | Diff to previous | History

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <xs:schema  xmlns:xs="http://www.w3.org/2001/XMLSchema"
4              targetNamespace="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/"
5              xmlns="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/">
6
7  <xs:element name="DIF">
8     <xs:complexType>
9        <xs:sequence>
10          <xs:element ref="Entry_ID" minOccurs="1" maxOccurs="1"/>
11          <xs:element ref="Entry_Title" minOccurs="1" maxOccurs="1"/>
12          <xs:element ref="Data_Set_Citation" minOccurs="0" maxOccurs="unbounded"/>
13          <xs:element ref="Personnel" minOccurs="0" maxOccurs="unbounded"/>
14          <xs:element ref="Discipline" minOccurs="0" maxOccurs="unbounded"/>
15          <xs:element ref="Parameters" minOccurs="1" maxOccurs="unbounded"/>
16          <xs:element ref="ISO_Topic_Category" minOccurs="0" maxOccurs="unbounded"/>
17          <xs:element ref="Keyword" minOccurs="0" maxOccurs="unbounded"/>
18          <xs:element ref="Sensor_Name" minOccurs="0" maxOccurs="unbounded"/>
19          <xs:element ref="Source_Name" minOccurs="0" maxOccurs="unbounded"/>
20          <xs:element ref="Temporal_Coverage" minOccurs="0" maxOccurs="unbounded"/>
21          <xs:element ref="Paleo_Temporal_Coverage" minOccurs="0" maxOccurs="unbounded"/>
22          <xs:element ref="Data_Set_Progress" minOccurs="0" maxOccurs="1"/>
23          <xs:element ref="Spatial_Coverage" minOccurs="0" maxOccurs="unbounded"/>
24          <xs:element ref="Location" minOccurs="0" maxOccurs="unbounded"/>
25          <xs:element ref="Data_Resolution" minOccurs="0" maxOccurs="unbounded"/>
26          <xs:element ref="Project" minOccurs="0" maxOccurs="unbounded"/>
27          <xs:element ref="Quality" minOccurs="0" maxOccurs="1"/>
28          <xs:element ref="Access_Constraints" minOccurs="0" maxOccurs="1"/>
29          <xs:element ref="Use_Constraints" minOccurs="0" maxOccurs="1"/>
30          <xs:element ref="Data_Set_Language" minOccurs="0" maxOccurs="unbounded"/>
31          <xs:element ref="Originating_Center" minOccurs="0" maxOccurs="1"/>
32          <xs:element ref="Data_Center" minOccurs="1" maxOccurs="unbounded"/>
33          <xs:element ref="Distribution" minOccurs="0" maxOccurs="unbounded"/>
34          <xs:element ref="Multimedia_Sample" minOccurs="0" maxOccurs="unbounded"/>
35          <xs:element ref="Reference" minOccurs="0" maxOccurs="unbounded"/>
36          <xs:element ref="Summary" minOccurs="1" maxOccurs="1"/>
37          <xs:element ref="Related_URL" minOccurs="0" maxOccurs="unbounded"/>
38          <xs:element ref="Parent_DIF" minOccurs="0" maxOccurs="unbounded"/>
39          <xs:element ref="IDN_Node" minOccurs="0" maxOccurs="unbounded"/>
40          <xs:element ref="Originating_Metadata_Node" minOccurs="0" maxOccurs="1"/>
41          <xs:element ref="Metadata_Name" minOccurs="1" maxOccurs="1"/>
42          <xs:element ref="Metadata_Version" minOccurs="1" maxOccurs="1"/>
43          <xs:element ref="DIF_Creation_Date" minOccurs="0" maxOccurs="1"/>
44          <xs:element ref="Last_DIF_Revision_Date" minOccurs="0" maxOccurs="1"/>
45          <xs:element ref="DIF_Revision_History" minOccurs="0" maxOccurs="1"/>
46          <xs:element ref="Future_DIF_Review_Date" minOccurs="0" maxOccurs="1"/>
47          <xs:element ref="Private" minOccurs="0" maxOccurs="1"/>
48          <xs:element ref="Extended_Metadata" minOccurs="0" maxOccurs="unbounded"
49       </xs:sequence>
```

NorDataNet
Norwegian Scientific Data Network

Example DIF-9

In the schema for a standard you can find the specifications of the elements and their occurance

```
      </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="Entry_ID" type="xs:string"/>
<xs:element name="Entry_Title" type="xs:string"/>
```

master / DIF Schemas / 9.x / **dif_v9.9.3.xsd**

Source view | Diff to previous | History

```
<?xml version="1.0" encoding="UTF-8"?>

<xs:schema  xmlns:xs="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/"
            xmlns="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/">

<xs:element name="DIF">
   <xs:complexType>
      <xs:sequence>
         <xs:element ref="Entry_ID" minOccurs="1" maxOccurs="1"/>
         <xs:element ref="Entry_Title" minOccurs="1" maxOccurs="1"/>
         <xs:element ref="Data_Set_Citation" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Personnel" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Discipline" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Parameters" minOccurs="1" maxOccurs="unbounded"/>
         <xs:element ref="ISO_Topic_Category" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Keyword" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Sensor_Name" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Source_Name" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Temporal_Coverage" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Paleo_Temporal_Coverage" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Data_Set_Progress" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Spatial_Coverage" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Location" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Data_Resolution" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Project" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Quality" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Access_Constraints" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Use_Constraints" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Data_Set_Language" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Originating_Center" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Data_Center" minOccurs="1" maxOccurs="unbounded"/>
         <xs:element ref="Distribution" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Multimedia_Sample" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Reference" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Summary" minOccurs="1" maxOccurs="1"/>
         <xs:element ref="Related_URL" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Parent_DIF" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="IDN_Node" minOccurs="0" maxOccurs="unbounded"/>
         <xs:element ref="Originating_Metadata_Node" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Metadata_Name" minOccurs="1" maxOccurs="1"/>
         <xs:element ref="Metadata_Version" minOccurs="1" maxOccurs="1"/>
         <xs:element ref="DIF_Creation_Date" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Last_DIF_Revision_Date" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="DIF_Revision_History" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Future_DIF_Review_Date" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Private" minOccurs="0" maxOccurs="1"/>
         <xs:element ref="Extended_Metadata" minOccurs="0" maxOccurs="unbounded"
      </xs:sequence>
```

NorDataNet
Norwegian Scientific Data Network

Example DIF-10

In the schema for a standard you can find the specifications of the elements and their occurance

```
</xs:annotation>
<xs:sequence>
    <xs:element name="Dataset_Creator"          type="xs:string"               minOccurs="0"/>
    <xs:element name="Dataset_Editor"           type="xs:string"               minOccurs="0"/>
    <xs:element name="Dataset_Title"            type="xs:string"               minOccurs="0"/>
    <xs:element name="Dataset_Series_Name"      type="xs:string"               minOccurs="0"/>
    <xs:element name="Dataset_Release_Date"     type="xs:string"               minOccurs="0"/><!-- toc
    <xs:element name="Dataset_Release_Place"    type="xs:string"               minOccurs="0"/>
    <xs:element name="Dataset_Publisher"        type="xs:string"               minOccurs="0"/>
    <xs:element name="Version"                  type="xs:string"               minOccurs="0"/><!-- toc
    <xs:element name="Issue_Identification"     type="xs:string"               minOccurs="0"/>
    <xs:element name="Data_Presentation_Form"   type="xs:string"               minOccurs="0"/>
    <xs:element name="Other_Citation_Details"   type="xs:string"               minOccurs="0"/>
    <xs:element name="Persistent_Identifier"    type="PersistentIdentifierType" minOccurs="0"/>
    <xs:element name="Online_Resource"          type="xs:anyURI"               minOccurs="0"/><!-- cha
</xs:sequence>
</xs:complexType>
```

```
URI

<xs:complexType>
    <xs:sequence>
        <xs:element name="Entry_ID"                  type="EntryIDType"/>                                                                <!
                                                                                                                                         <!
                                                                                                                                         <!
        <xs:element name="Version_Description"       type="VersionDescriptionType"      minOccurs="0"/>                                  <!
        <xs:element name="Entry_Title"               type="EntryTitleType"/>                                                             <!
        <xs:element name="Dataset_Citation"          type="DatasetCitationType"         minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Personnel"                 type="PersonnelType"               minOccurs="0"   maxOccurs="unbounded"/> <!
                                                                                                                                         <!
        <xs:element name="Science_Keywords"          type="ScienceKeywordsType"                         maxOccurs="unbounded"/> <!
        <xs:element name="ISO_Topic_Category"        type="ISOTopicCategoryType"        minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Ancillary_Keyword"         type="AncillaryKeywordType"        minOccurs="0"   maxOccurs="unbounded"/> <!
                                                                                                                                         <!
                                                                                                                                         <!
        <xs:element name="Platform"                  type="PlatformType"                                maxOccurs="unbounded"/> <!
        <xs:element name="Temporal_Coverage"         type="TemporalCoverageType"                        maxOccurs="unbounded"/> <!
        <xs:element name="Dataset_Progress"          type="DatasetProgressType"         minOccurs="0"/>                                  <!
        <xs:element name="Spatial_Coverage"          type="SpatialCoverageType"/>                                                        <!
        <xs:element name="Location"                  type="LocationType"                minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Data_Resolution"           type="DataResolutionType"          minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Project"                   type="ProjectType"                                 maxOccurs="unbounded"/> <!
        <xs:element name="Quality"                   type="QualityType"                 minOccurs="0"/>                                  <!
        <xs:element name="Access_Constraints"        type="AccessConstraintsType"       minOccurs="0"/>                                  <!
        <xs:element name="Use_Constraints"           type="UseConstraintsType"          minOccurs="0"/>                                  <!
        <xs:element name="Dataset_Language"          type="DatasetLanguageType"         minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Originating_Center"        type="OriginatingCenterType"       minOccurs="0"/>                                  <!
        <xs:element name="Organization"              type="OrganizationType"                            maxOccurs="unbounded"/> <!
        <xs:element name="Distribution"              type="DistributionType"            minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Multimedia_Sample"         type="MultimediaSampleType"        minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Reference"                 type="ReferenceType"               minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Summary"                   type="SummaryType"/>                                                                <!
        <xs:element name="Related_URL"               type="RelatedURLType"                              maxOccurs="unbounded"/> <!
                                                                                                                                         <!
        <xs:element name="Metadata_Association"      type="MetadataAssociationType"     minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="IDN_Node"                  type="IDNNodeType"                 minOccurs="0"   maxOccurs="unbounded"/> <!
        <xs:element name="Originating_Metadata_Node" type="OriginatingMetadataNodeType" minOccurs="0"/>                                  <!
        <xs:element name="Metadata_Name"             type="MetadataNameType"/>                                                           <!
        <xs:element name="Metadata_Version"          type="MetadataVersionType"/>                                                        <!
                                                                                                                                         <!
                                                                                                                                         <!
        <xs:element name="DIF_Revision_History"      type="DIFRevisionHistoryType"      minOccurs="0"/>                                  <!
                                                                                                                                         <!
        <xs:element name="Metadata_Dates"            type="MetadataDatesType"/>                                                          <!
        <xs:element name="Private"                   type="PrivateType"                 minOccurs="0"/>                                  <!
        <xs:element name="Additional_Attributes"     type="AdditionalAttributesType"    minOccurs="0"   maxOccurs="unbounded"/> <!

        <xs:element name="Product_Level_Id"          type="ProcessingLevelIdType"                                                        !
        <xs:element name="Collection_Data_Type"      type="CollectionDataTypeEnum"                                                       !
        <xs:element name="Product_Flag"              type="ProductFlagEnum"                                                              !
```

You can validate your metadata against this schemas

NorDataNet
Norwegian Scientific Data Network

Filling in Metadata Standards

I need to fill in a metadata element about:

Dataset Production Status: Describes the production status of the data set regarding its completeness.

What should I put in there?

Metadata file:

<Title>The title of the dataset</Title>
<Abstract> This dataset collects…</Abstract>
<Dataset_Production_Status>XXX</Dataset_Production_Status>
<Start_Date>2020-01-20</Start_Date>

NorDataNet
Norwegian Scientific Data Network

Filling in Metadata Standards

I need to fill in a metadata element about:

Dataset Production Status: Describes the production status of the data set regarding its completeness.

Data provider A can use:

“Not ready yet”

“Done”

“Still acquiring data”

“Continuously updating”

“???”

Data provider B can use:

“Not finished”

“Finished and stored”

“unknown”

“Not started yet”

“See www.mydataset.com”

NorDataNet
Norwegian Scientific Data Network

Controlled *vocabulary*

controlled vocabularies are a source of authoritative terms to be entered for values of certain elements

Label	Description
Planned	Refers to data sets to be collected in the future and are thus unavailable at the present time.
In Work	Refers to data sets currently undergoing production or data that is continuously being collected or updated.
Complete	Refers to data sets in which no updates or further data collection will be made.
Obsolete	A new version of the dataset has been generated. The new version should be used, this is kept for back tracing.

NorDataNet
Norwegian Scientific Data Network

Controlled *vocabulary*

controlled vocabularies are a source of authoritative terms to be entered for values of certain elemen

<Title>The title of the dataset</Title>
<Abstract> This dataset collects…</Abstract>
<Dataset_Production_Status>Complete</Dataset_Production_Status>
<Start_Date>2020-01-20</Start_Date>

Label	
Planned	
In Work	
Complete	Refers to data sets in which no updates or further data collection will be made.
Obsolete	A new version of the dataset has been generated. The new version should be used, this is kept for back tracing.

NorDataNet
Norwegian Scientific Data Network

Profiles

Some standards can have different profiles, i.e. a generic standard is adapted to specific requirements of a community, leading to a specific profile of that standard.

A specific element of a standard can be within a profile:

- from optional to mandatory attributes
- from a string to a controlled vocabulary
- from a controlled vocabulary to a subset of it

NorDataNet
Norwegian Scientific Data Network

ACDD Convention

When encoding data as netCDF/CF is good practise to include discovery metadata in the file using the Attribute Convention for dataset Discovery (ACDD). Discovery metadata will then be directly connected to the data themselves and can be extracted for ingestion in the searchable catalogue.

Attribute	Description
id	An identifier for the data set, provided by and unique within its naming authority. The combination of the "naming authority" and the "id" should be globally unique, but the id can be globally unique by itself also. IDs can be URLs, URNs, DOIs, meaningful text strings, a local key, or any other unique string of characters. The id should not include white space characters.
naming_authority	The organization that provides the initial id (see above) for the dataset. The naming authority should be uniquely specified by this attribute. We recommend using reverse-DNS naming for the naming authority; URIs are also acceptable. Example: 'edu.ucar.unidata'.
title	A short phrase or sentence describing the dataset. In many discovery systems, the title will be displayed in the results list from a search, and therefore should be human readable and reasonable to display in a list of such names. This attribute is also recommended by the NetCDF Users Guide and the CF conventions.
summary	A paragraph describing the dataset, analogous to an abstract for a paper.
keywords	A comma-separated list of key words and/or phrases. Keywords may be common words or phrases, terms from a controlled vocabulary (GCMD is required), or URIs for terms from a controlled vocabulary (see also "keywords_vocabulary" attribute).
geospatial_lat_min	Describes a simple lower latitude limit; may be part of a 2- or 3-dimensional bounding region. Geospatial_lat_min specifies the southernmost latitude covered by the dataset. Must be decimal degrees north.
geospatial_lat_max	Describes a simple upper latitude limit; may be part of a 2- or 3-dimensional bounding region. Geospatial_lat_max specifies the northernmost latitude covered by the dataset. Must be decimal degrees north.
geospatial_lon_min	Describes a simple longitude limit; may be part of a 2- or 3-dimensional bounding region. geospatial_lon_min specifies the westernmost longitude covered by the dataset. See also geospatial_lon_max. Must be decimal degrees east.
geospatial_lon_max	Describes a simple longitude limit; may be part of a 2- or 3-dimensional bounding region. geospatial_lon_max specifies the easternmost longitude covered by the dataset. Cases where geospatial_lon_min is greater than geospatial_lon_max indicate the bounding box extends from geospatial_lon_max, through the longitude range discontinuity meridian (either the antimeridian for -180:180 values, or Prime Meridian for 0:360 values), to geospatial_lon_min; for exampl geospatial_lon_min=170 and geospatial_lon_max=-175 incorporates 15 degrees of longitude (ranges 170 to 180 and -180 to -175). Must be decimal degrees

https://www.nordatanet.no/en/node/75

NorDataNet
Norwegian Scientific Data Network

Use metadata

For proper interpretation of data

- Standardised naming of variables
- Units of variables
- Encoding of missing values
- Indication of spatial or temporal features
    - e.g. averages over a certain period or space
- Relations between variables

Example:

- Climate and forecast convention (CF) describes
- WMO BUFR tables
- WMO GRIB tables

NorDataNet
Norwegian Scientific Data Network

CF Standard Name Table

Version 29, 08 July 2015

Refer to the Guidelines for Construction of CF Standard Names for information on how the names are constructed and interpreted, and how new names could be derived.

A note about units
The canonical units associated with each standard name are usually the SI units for the quantity. Section 3.3 of the CF conventions states: "Unless it is dimensionless, a variable with a standard_name attribute must have units which are physically equivalent (not necessarily identical) to the canonical units, possibly modified by an operation specified by either the standard name modifier ... or by the cell_methods attribute." Furthermore, Section 1.3 of the CF conventions states: "The values of the units attributes are character strings that are recognized by UNIDATA's Udunits package [UDUNITS], (with exceptions allowed as discussed in Section 3.1, "Units")." For example, a variable with the standard name of "air_temperature" may have a units attribute of "degree_Celsius" because Celsius can be converted to Kelvin by Udunits. For the full range of supported units, refer to the Udunits documentation. Refer to the CF conventions for full details of the units attribute.

Search

Search Standard Names | Show All Standard Names

AND OR (separate search terms with spaces)
Also search help text

View by Category

Atmospheric Chemistry	Atmosphere Dynamics	Carbon Cycle	Cloud	Hydrology
Ocean Dynamics	Radiation	Sea Ice	Surface	

Standard Name	Canonical Units	AMIP	GRIB
aerodynamic_particle_diameter	m		
aerodynamic_resistance	m-1 s		
age_of_sea_ice	year		
age_of_stratospheric_air	s		
age_of_surface_snow	day		
air_density	kg m-3		
air_potential_temperature	K	theta	13
air_pressure	Pa	plev	1
air_pressure_anomaly	Pa		26
air_pressure_at_cloud_base	Pa		
air_pressure_at_cloud_top	Pa		
air_pressure_at_convective_cloud_base	Pa		

CF Standard Name Table

Version 29, 08 July 2015

Refer to the Guidelines for Construction of CF Standard Names for information on how the names are constructed and interpreted, and how new names could be derived.

A note about units

The canonical units associated with each standard name are usually the SI units for the quantity. Section 3.3 of the CF conventions states: "Unless it is dimensionless, a variable with a standard_name attribute must have units which are physically equivalent (not necessarily identical) to the canonical units, possibly modified by an operation specified by either the standard name modifier ... or by the cell_methods attribute." Furthermore, Section 1.3 of the CF conventions states: "The values of the units attributes are character strings that are recognized b
"degree_Celsius" because Celsius can be

Search

AND OR (separate search terms
Also search help text

View by Category

Atmospheric Chemistry	Atmosphere
Ocean Dynamics	Radiation

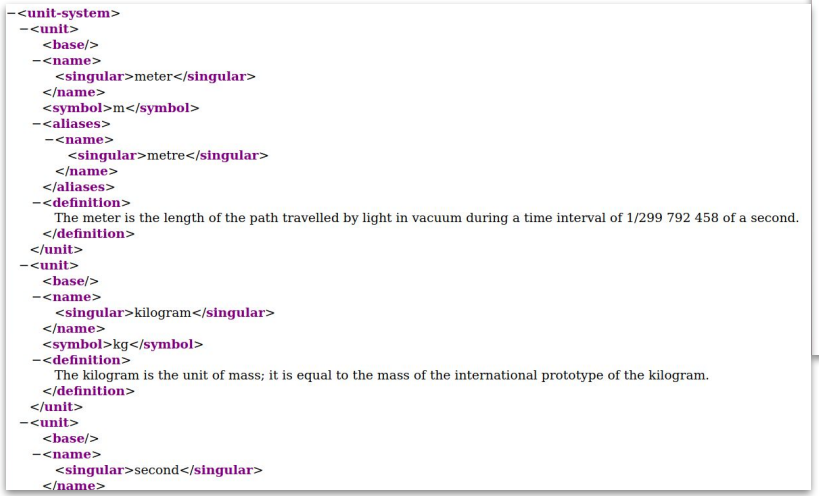
- acoustic_signal_roundtrip_travel_time_in_sea_water
- aerodynamic_particle_diameter
- aerodynamic_resistance
- age_of_sea_ice
- age_of_stratospheric_air
- age_of_surface_snow
- air_density
- air_potential_temperature
- air_pressure
  Air pressure is the force per unit area which would be exerted when the moving gas molecules of which the air is composed strike a theoretical surface of any orientation.
- air_pressure_anomaly
  The term "anomaly" means difference from climatology. Air pressure is the force per unit area which would be exerted when the moving gas molecules of which the air is composed strike a theoretical surface of any orientation.
- air_pressure_at_cloud_base
  The phrase "cloud_base" refers to the base of the lowest cloud. Air pressure is the force per unit area which would be exerted when the moving gas molecules of which the air is composed strike a theoretical surface of any orientation.
- air_pressure_at_cloud_top
- air_pressure_at_convective_cloud_base

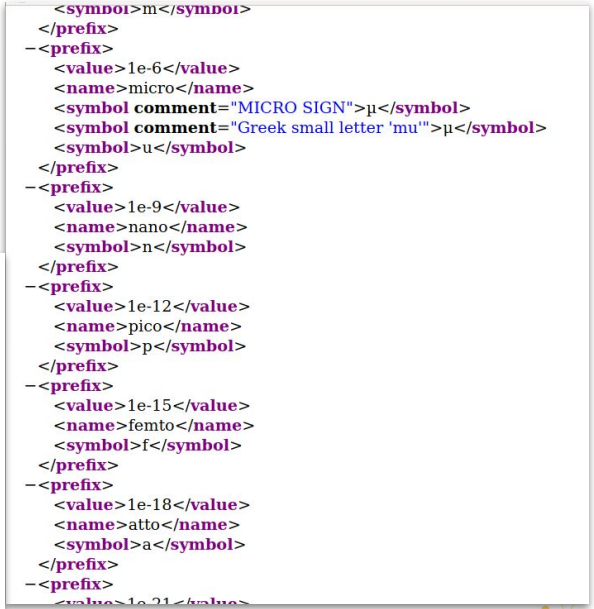
aerodynamic_particle_diameter			
aerodynamic_resistance	m-1 s		
age_of_sea_ice	year		
age_of_stratospheric_air	s		
age_of_surface_snow	day		
air_density	kg m-3		
air_potential_temperature	K	theta	13
air_pressure	Pa	plev	1
air_pressure_anomaly	Pa		26
air_pressure_at_cloud_base	Pa		
air_pressure_at_cloud_top	Pa		
air_pressure_at_convective_cloud_base	Pa		

UDUNITS

The database for the UDUNITS-2 package comprises one XML file containing unit prefixes and four XML files containing unit definitions:

- SI unit prefixes
- SI base units
- SI derived units
- Units accepted for use with the SI
- Non-SI units





NorDataNet
Norwegian Scientific Data Network

Site metadata

For proper interpretation of data

- The details and history of local conditions, instruments, operating procedures, data processing algorithms and other factors pertinent to interpreting data (i.e., metadata) should be documented and treated with the same care as the data themselves.

***OM_Observation**: an EVENT whose RESULT is an **estimate** of a value of some PROPERTY of some THING obtained using a specified PROCEDURE …*

NorDataNet
Norwegian Scientific Data Network

WIGOS Metadata Categories

#	Category	Description
1	Observed variable	Specifies the basic characteristics of the observed variable and the resulting datasets.
2	Purpose of observation	Specifies the main application area(s) of the observation and the observing programme(s) and networks the observation is affiliated to.
3	Station/platform	Specifies the environmental monitoring facility, including fixed station, moving equipment or remote sensing platform, at which the observation is made.
4	Environment	Describes the geographical environment within which the observation is made. It also provides an unstructured element for additional meta-information that is considered relevant for adequate use of the data and that is not captured anywhere else in this standard.
5	Instruments and methods of observation	Specifies the method of observation and describes characteristics of the instrument(s) used to make the observation. If multiple instruments are used to generate the observation, then this category should be repeated.
6	Sampling	Specifies how sampling and/or analysis are used to derive the reported observation or how a specimen is collected.
7	Data processing and reporting	Specifies how raw data are transferred into the observed variable and reported to the users.
8	Data quality	Specifies the data quality and traceability of the observation.
9	Ownership and data policy	Specifies who is responsible for the observation and owns it.
10	Contact	Specifies where information about the observation or dataset can b

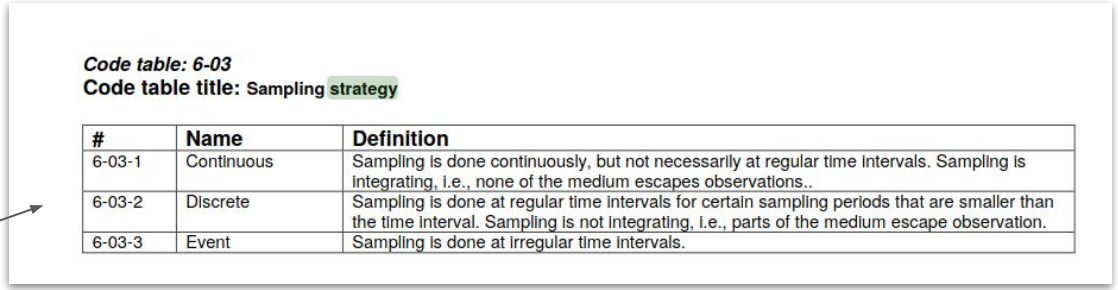
WIGOS Metadata Categories

#	Category	Description
5	Instruments and methods of observation	Specifies the method of observation and describes characteristics of the instrument(s) used to make the observation. If multiple instruments are used to generate the observation, then this category should be repeated.
6	Sampling	Specifies how sampling and/or analysis are used to derive the reported observation or how a specimen is collected.
7	Data processing and reporting	Specifies how raw data are transferred into the observed variable and reported to the users.
8	Data quality	Specifies the data quality and traceability of the observation.
9	Ownership and data policy	Specifies who is responsible for the observation and owns it.
10	Contact	Specifies where information about the observation or dataset can b

Category	ID	Name	Definition	MCO	Phase
6. Sampling	6-01	Sampling procedures	Procedures involved in obtaining a sample	O	III
	6-02	Sample treatment	Chemical or physical treatment of sample prior to analysis	O	III
	6-03	Sampling strategy	The strategy used to generate the observed variable	O*	I
	6-04	Sampling time period	The period of time over which a measurement is taken	M#	III
	6-05	Spatial sampling resolution	Spatial resolution refers to the size of the smallest observable object. The intrinsic resolution of an imaging system is determined primarily by the instantaneous field of view of the sensor, which is a measure of the ground area viewed by a single detector element in a given instance in time	M#	II
	6-06	Temporal sampling interval	Time period between the beginning of consecutive sampling periods	M#	III
	6-07	Diurnal base time	Time to which diurnal statistics are referenced	C#	I
	6-08	Schedule of observation	Schedule of observation	M#	I

WIGOS Controlled Vocabularies

Code table 1-02 - measurement units
Code table 1-01 - Observed variable - measurand
Code table 2-01 – application areas
Code table 2-02 – Programme/Network affiliation
Code table 3-04 - Station/platform (Observing facility) type
Code table 3-08 – Data communication method
Code table 3-09 - Station operating status
Code table 4-02 - Surface cover classification scheme
Code table 5-01 – Source of observation
Code table 5-02 - Measurement/observing method
Code table 5-04 – Instrument Operating Status
Code table 5-08-01 - control standard type
Code table 5-08-02 - control location
Code table 5-08-03 - Instrument control result
Code table 5-14 - Status of observation
Code table 5-15 - Exposure of instrument
Code table 6-03 - Sampling strategy
Code table 7-07 – Data format
Code table 8-04 – Quality Flag System



Code table: 6-03
Code table title: Sampling strategy

#	Name	Definition
6-03-1	Continuous	Sampling is done continuously, but not necessarily at regular time intervals. Sampling is integrating, i.e., none of the medium escapes observations..
6-03-2	Discrete	Sampling is done at regular time intervals for certain sampling periods that are smaller than the time interval. Sampling is not integrating, i.e., parts of the medium escape observation.
6-03-3	Event	Sampling is done at irregular time intervals.

NorDataNet
Norwegian Scientific Data Network

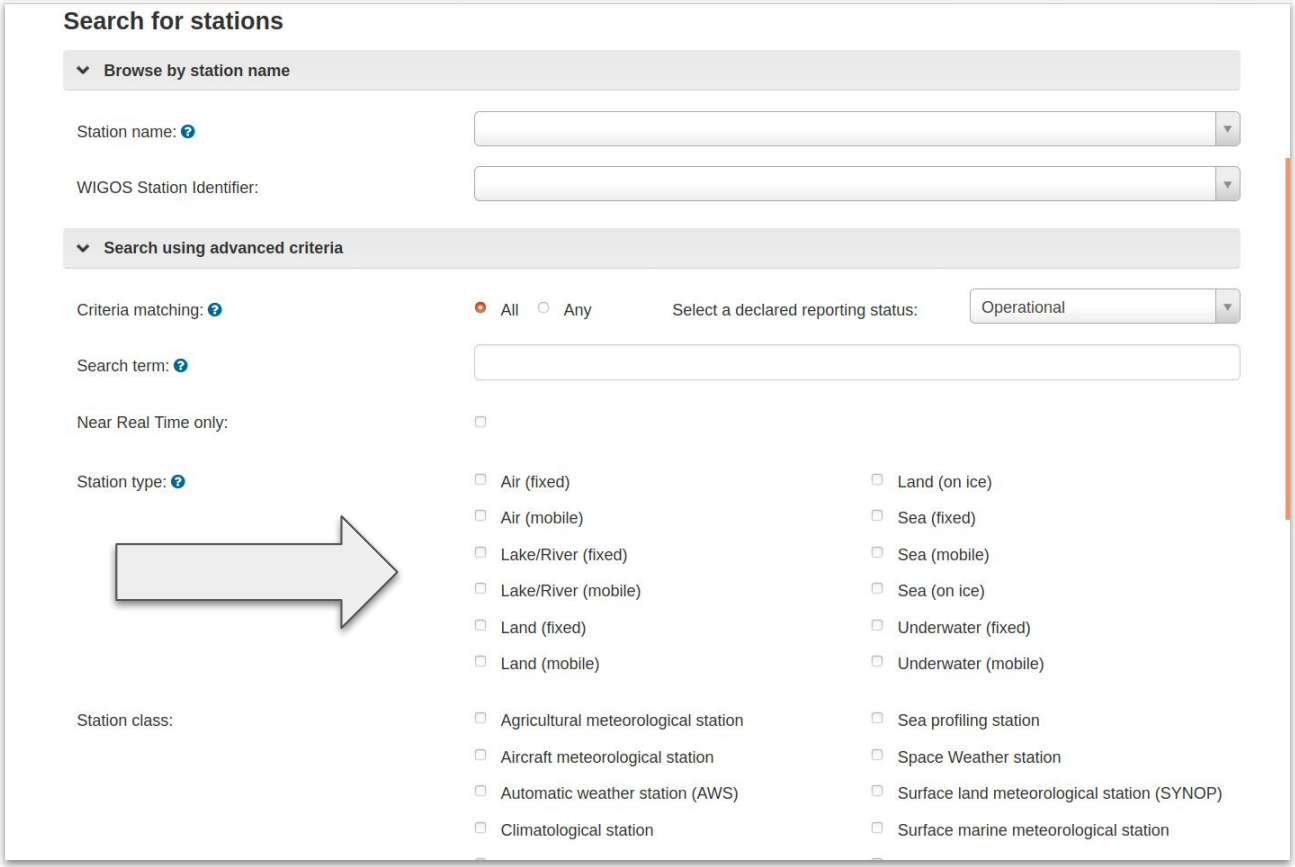
WIGOS Controlled Vocabularies

Code table 1-02 - measurement units
Code table 1-01 - Observed variable - measurand
Code table 2-01 – application areas
Code table 2-02 – Programme/Network affiliation
Code table 3-04 - Station/platform (Observing facility) type
Code table 3-08 – Data communication method
Code table 3-09 - Station operating status
Code table 4-02 - Surface cover classification scheme
Code table 5-01 – Source of observation
Code table 5-02 - Measurement/observing method
Code table 5-04 – Instrument Operating Status
Code table 5-08-01 - control standard type
Code table 5-08-02 - control location
Code table 5-08-03 - Instrument control result
Code table 5-14 - Status of observation
Code table 5-15 - Exposure of instrument
Code table 6-03 - Sampling strategy
Code table 7-07 – Data format
Code table 8-04 – Quality Flag System

Name	Notation	Description	Types	Status
Air (fixed)	airFixed	Airborne station/platform, at fixed position	Concept	stable
Air (mobile)	airMobile	Airborne station/platform, moving around	Concept	stable
inapplicable	inapplicable	None of the codes in the table are applicable in the context ...	Concept	stable
Lake/River (fixed)	lakeRiverFixed	Station/platform at lake/river surface, at fixed position	Concept	stable
Lake/River (mobile)	lakeRiverMobile	station/platform at lake/river surface, moving around	Concept	stable
Land (fixed)	landFixed	Station/platform on solid terrain, at fixed position	Concept	stable
Land (mobile)	landMobile	Station/platform on solid terrain, moving around	Concept	stable
Land (on ice)	landOnIce	Station/platform on ice-covered ground, moving with the ice	Concept	stable
Sea (fixed)	seaFixed	Station/platform at sea surface, at fixed position	Concept	stable
Sea (mobile)	seaMobile	Station/platform at sea surface, moving around	Concept	stable
Sea (on ice)	seaOnIce	Station/platform on floating ice, moving with the ice	Concept	stable
Space-based	spaceBased	Satellite platform in orbit	Concept	stable
Underwater (fixed)	underwaterFixed	Station/platform under water, at fixed horizontal position	Concept	stable
Underwater (mobile)	underwaterMobile	Station/platform under water, moving around also horizontally	Concept	stable
unknown	unknown	The station/platform type is unknown.	Concept	stable

NorDataNet
Norwegian Scientific Data Network

OSCAR



https://oscar.wmo.int/surface/#/search/station

NorDataNet
Norwegian Scientific Data Network

Definition of data provenance

Data provenance is information about entities, activities, and people involved in producing (influencing or delivering) a piece of data.

This information is used for assessments about the data regarding:

- its quality
- its reliability
- Its trustworthiness

Barbara Magana: Summer school 2019

NorDataNet
Norwegian Scientific Data Network

Provenance questions

Provenance is the entire information that ran through the whole history of data process, including all data sources and all processes generating these data.

- Who played a role when creating the data?
- Who owned the data?
- Who contributed to the data?
- How data was modified from its first revision?
- How other data affected the current data?
- Which tools were used to generate each version of the data?

Barbara Magana: Summer school 2019

NorDataNet
Norwegian Scientific Data Network

Provenance vs Metadata

Provenance is a kind of metadata, BUT not all metadata is provenance

- The title or format of a book is metadata, but it is not part of its provenance.
- The date of creation, the author, the publisher or the license of a book are part of its provenance.

Provenance and FAIR

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

- R1.2. (Meta)data are associated with detailed provenance

For others to reuse your data, they should know where the data came from, who to cite and/or how you wish to be acknowledged. Include a description of the workflow that led to your data:

Who generated or collected it? How has it been processed? Has it been published before? Does it contain data from someone else that you may have transformed or completed? Ideally, this workflow is described in a machine-readable format.

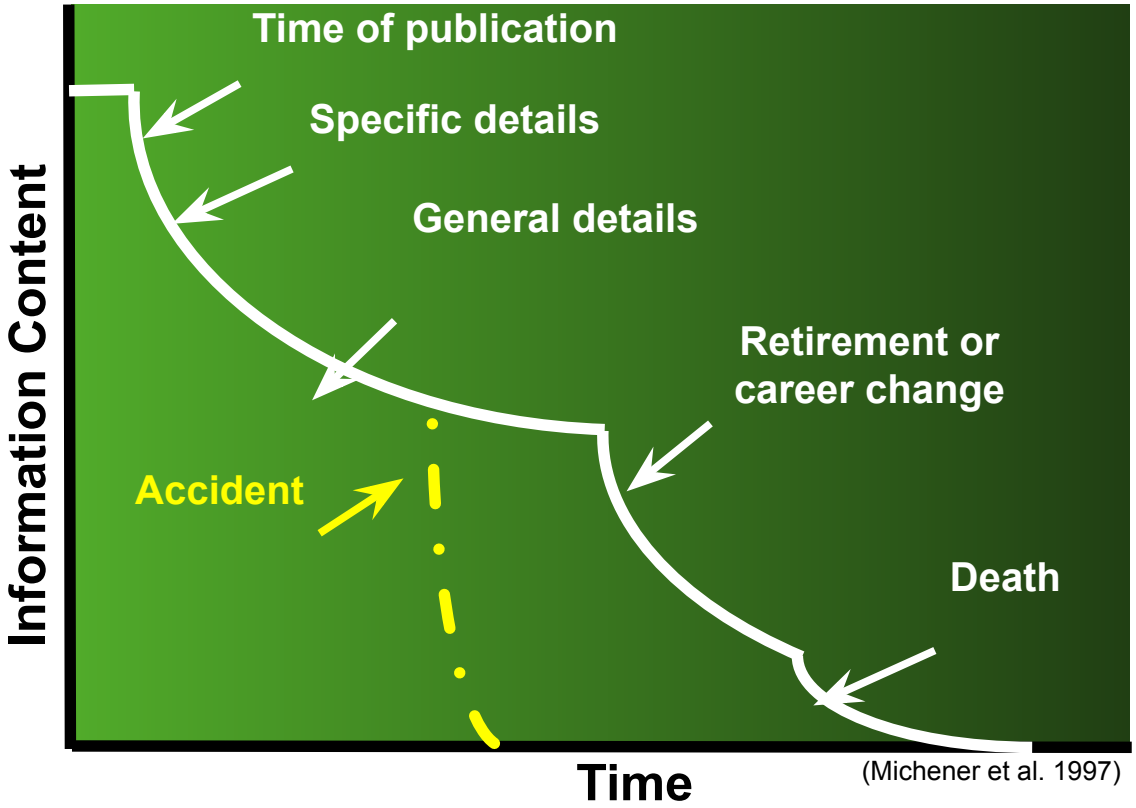
Barbara Magana: Summer school 2019

NorDataNet
Norwegian Scientific Data Network

Uses of Provenance

- Data reuse
- Estimation of data quality
- Attribution
- Data discovery
- Comparison
- Debugging

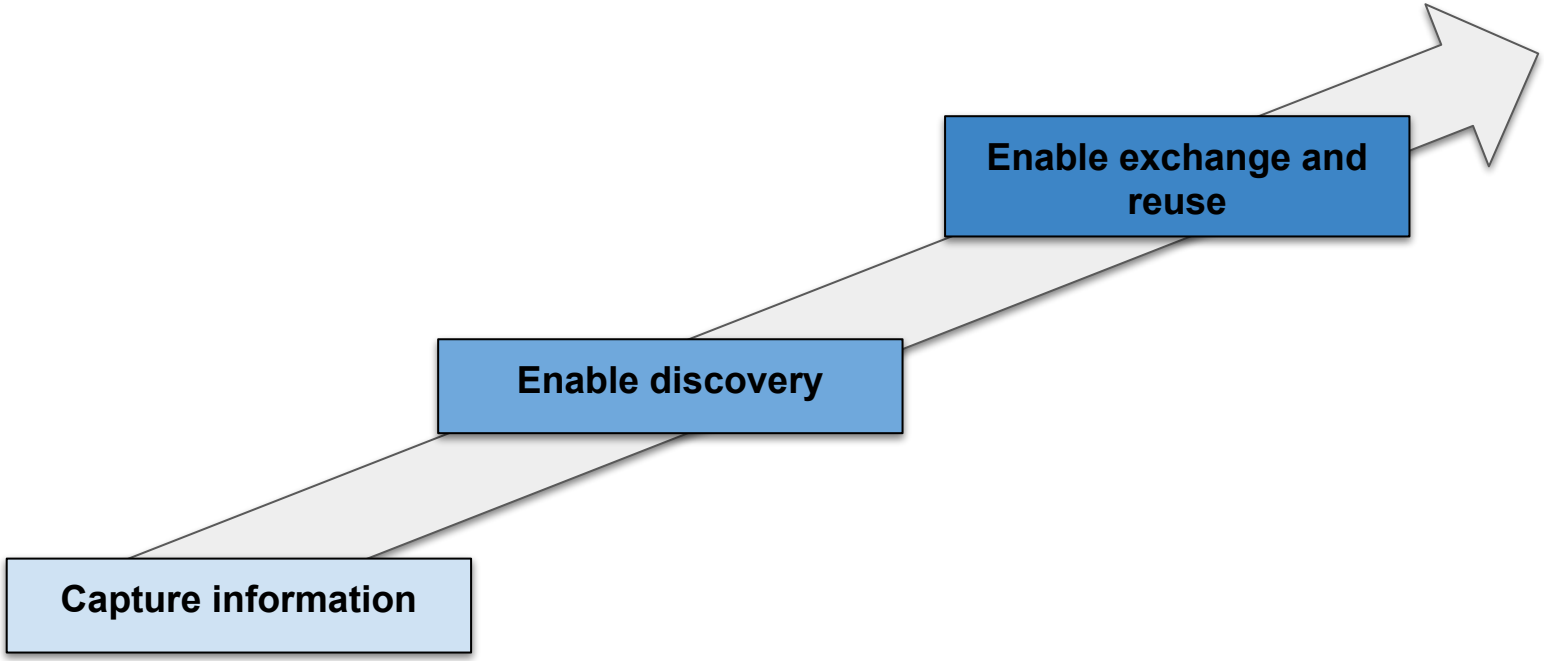
Knowledge about data in time



(Michener et al. 1997)

Metadata: what are they good for?

DataONE
Data Observation Network for Earth



NorDataNet
Norwegian Scientific Data Network

Concerns about writing metadata

DataONE
Data Observation Network for Earth

Concern	Solution
workload required to capture accurate robust metadata	incorporate metadata creation into data development process – distribute the effort
time and resources to create, manage, and maintain metadata	include in grant budget and schedule
readability / usability of metadata	use a standardized metadata format
discipline specific information and ontologies	Use a standard ‘profile’ that supports discipline specific information

NorDataNet
Norwegian Scientific Data Network

Tips for good metadata

- Titles are critical in helping readers find your data
  - While individuals are searching for the most appropriate datasets, they are most likely going to use the title as the first criteria to determine if a dataset meets their needs.
  - Treat the title as the opportunity to sell your dataset.
- A complete title includes: What, Where, When, Who, and Scale
  - An informative title includes: topic, timeliness of the data, specific information about place and geography

A Clear Choice: Which title is better?

• Rivers Maps

OR

• Greater Yellowstone Rivers from 1:126,700 U.S. Forest Service Visitor Maps (1961-1983)

The goal of a metadata record is to give the user enough information to know if they can use the data without contacting the dataset owner.

Tips for good metadata

- Use descriptive and clear writing
- Fully document geographic locations
- Select keywords wisely
- Use thesauri for keywords whenever possible
- Be detailed: there’s no such thing as too much metadata!

ESIP Commons. doi:10.7269/P37D2S2J

Tips for good metadata

Tips for the short-term: Get your own house in order

- Use common date formats, codes, smart file names, etc.
- WRITE EVERYTHING DOWN! (Keep good readme files)
- Put in the time early on to start implementing a standard
- Most have minimum compliance levels with options to get more detailed
- Stay flexible

Tips for the long-term:

- Get help! Librarians, standards groups, data centers, and domain communities have tools and expertise to share
- Watch for Best Practices and standards in your field

NorDataNet
Norwegian Scientific Data Network

Metadata templates

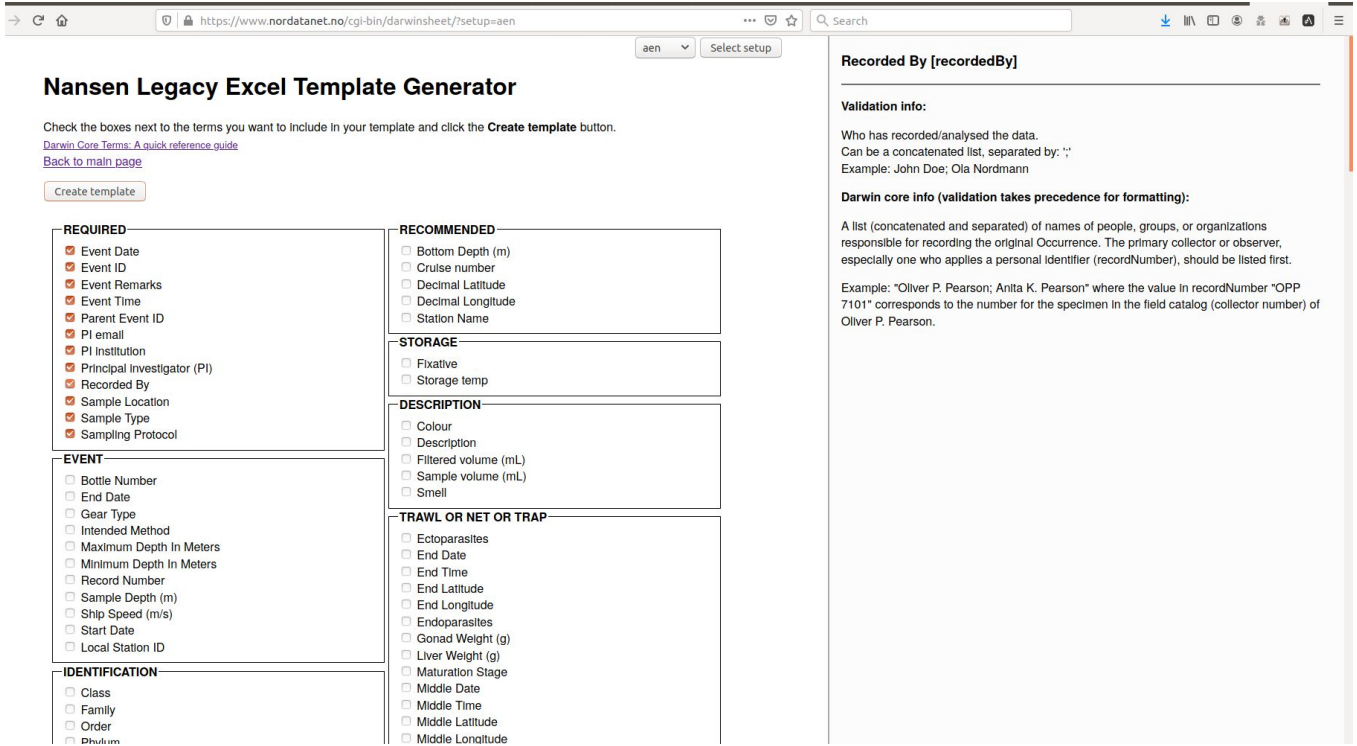
- There are tools available to help recording data in a proper way
  - Darwin Core template

    Darwin Core is a standard maintained by the Darwin Core maintenance group. It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to **facilitate the sharing of information about biological diversity** by providing identifiers, labels, and definitions.

    https://www.nordatanet.no/cgi-bin/darwinsheet/?setup=aen
  - *EBAS*

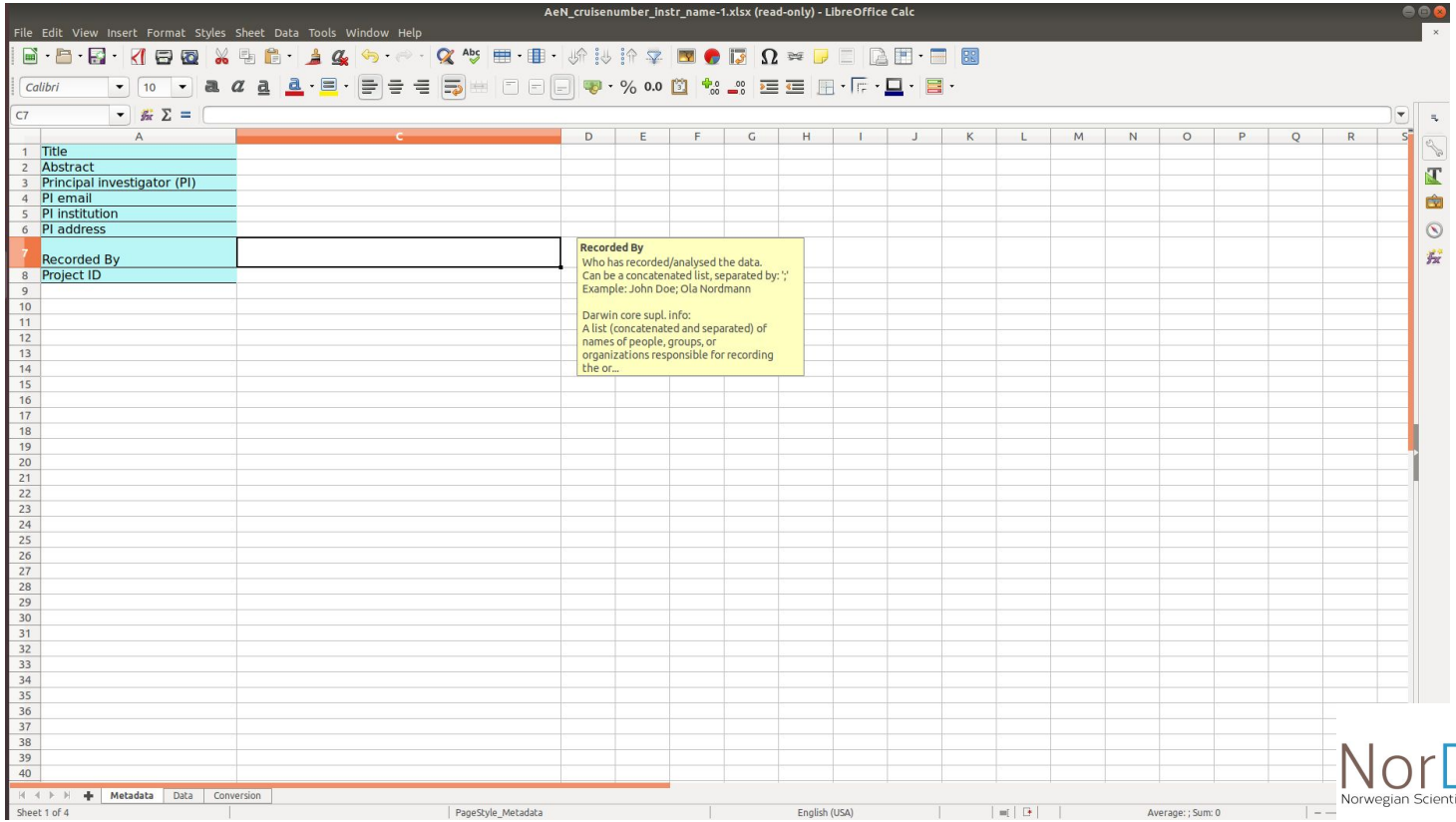
NorDataNet
Norwegian Scientific Data Network

Darwin Core Template



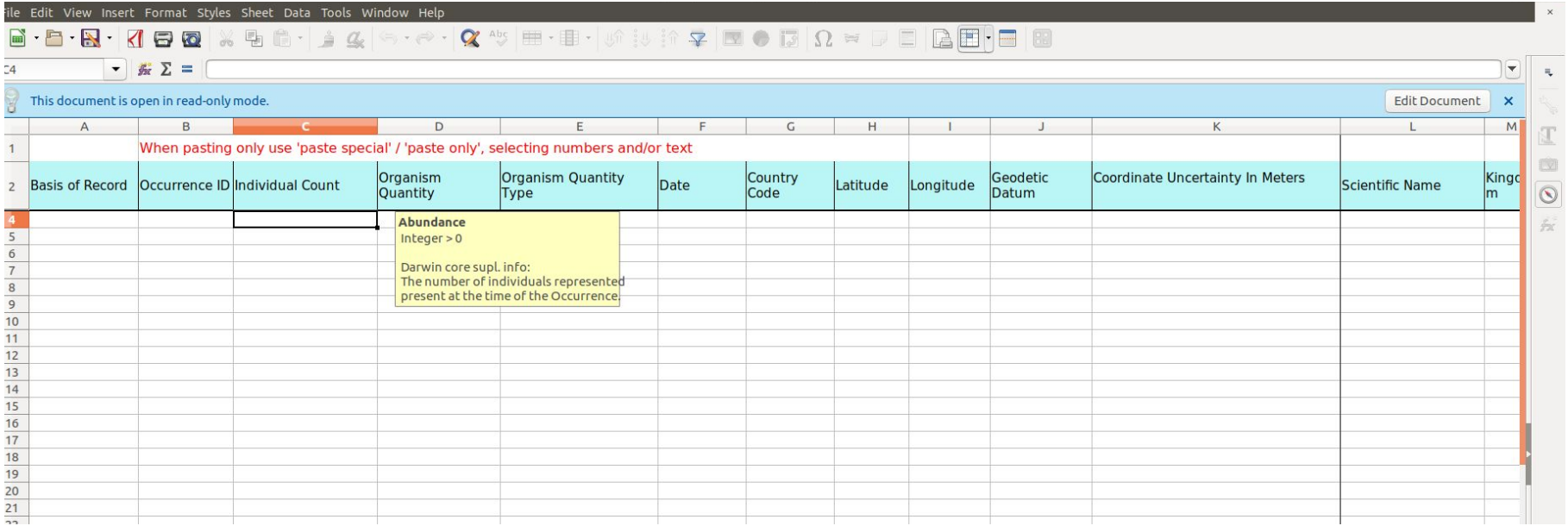
NorDataNet
Norwegian Scientific Data Network

Darwin Core Template



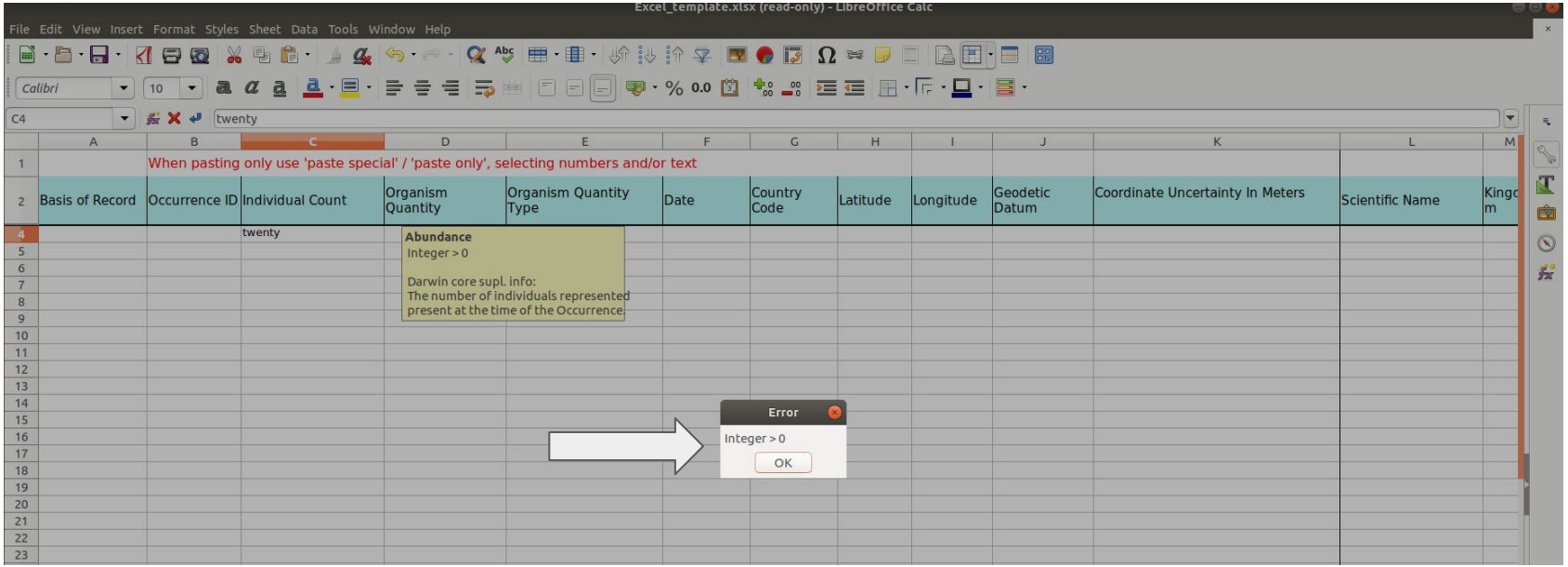
NorDataNet
Norwegian Scientific Data Network

Darwin Core Template



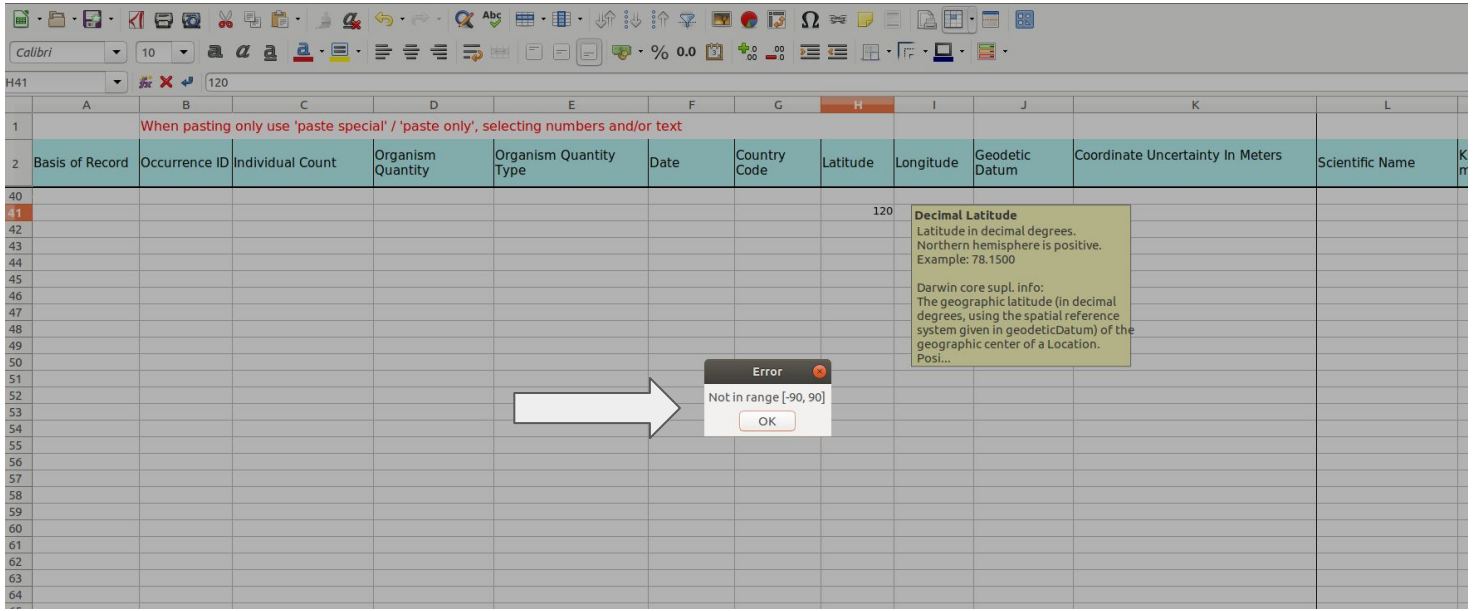
NorDataNet
Norwegian Scientific Data Network

Darwin Core Template



NorDataNet
Norwegian Scientific Data Network

Darwin Core Template



NorDataNet
Norwegian Scientific Data Network

Resources

DataOne: https://old.dataone.org/education-modules

https://old.dataone.org/all-best-practices

ESIP: https://commons.esipfed.org/node/1422

Data management training