Capacity Building Workshop on Hydrological Data Exchange, Standardization, Interoperability in RA VI
Zagreb, Croatia
January 30th 2024

WHOS metadata quality

Juan F. Bianchi, INA, Argentina

What is metadata and why?

Metadata is the data describing the data.

- Observations without metadata are of very limited use
- Necessary to provide users with confidence that the data are appropriate for their application
- Can be data depending on user needs and objectives

Metadata should be documented and treated with the same care as the data

Metadata types

Discovery metadata allow users to discover data and answer the following questions:

What does the data set contain?

Where were the data collected?

Who is the data provider?

When are the observations taken?

Observational metadata enable data to be interpreted in context

Measurement units

Station operating status

Measurement/observing method

Diurnal base time

Use metadata allow end-users to access and use data

Data policies

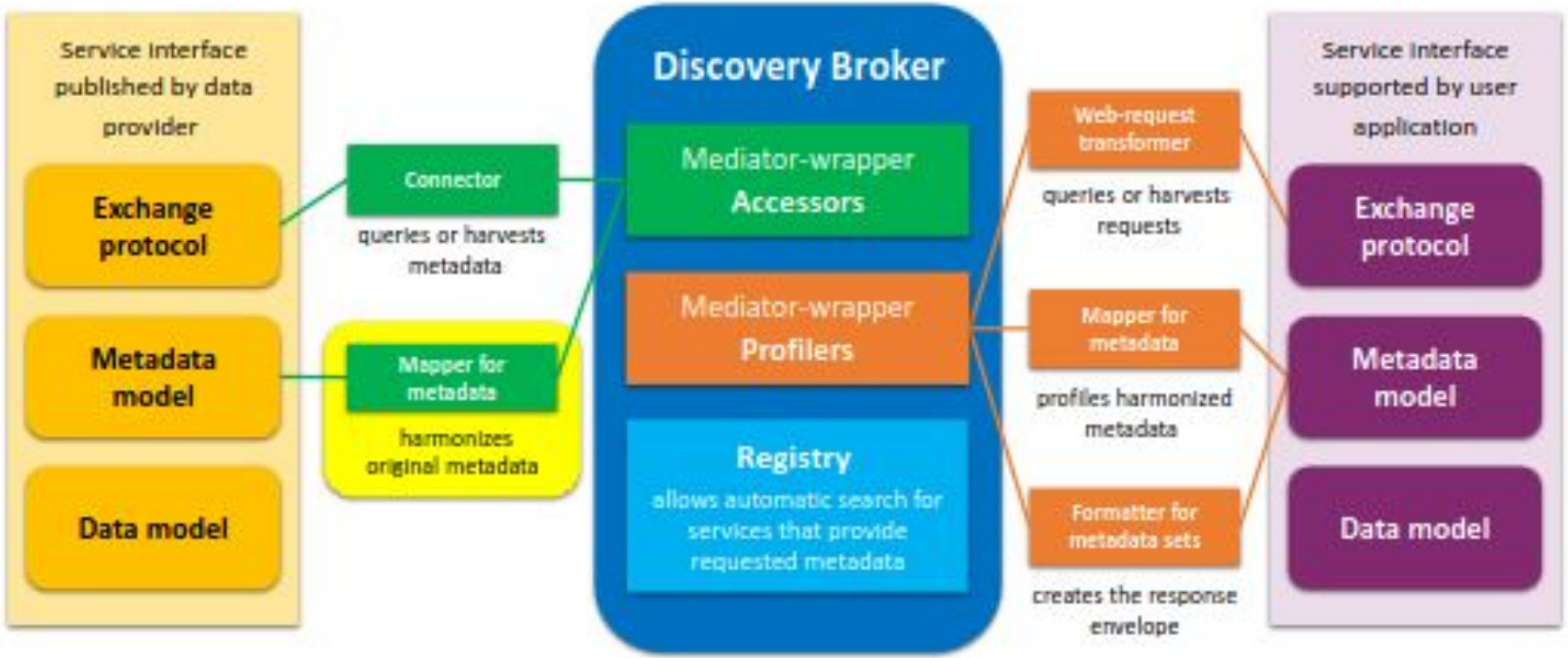
Access service endpoint

Access service protocol

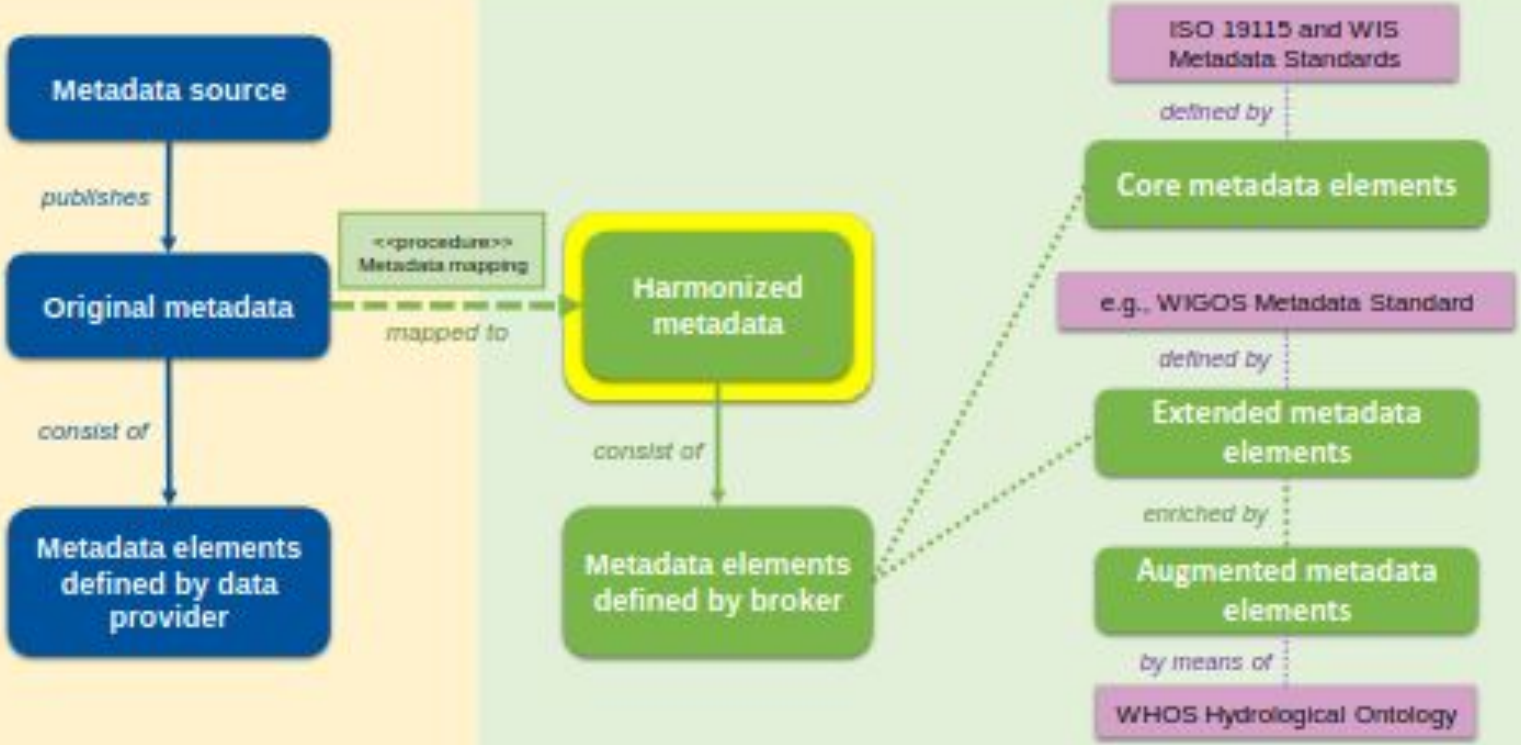
Key concepts

- Metadata may be used to describe data at different *aggregation levels*, such as:
  - Value
  - Time series
  - Dataset
  - Dataset series (collections)
- At a given aggregation level, a set of *metadata elements* (sometimes grouped forming metadata *classes*) may be defined, where each element definition normally includes a name, a domain (of possible values) and an indication of allowed multiplicity
- *Complex elements* can serve as a metadata class, allowing for the creation of *structured metadata*
- A formal definition of different metadata classes and elements constitutes a metadata *schema*
- A schema that is stored and maintained and published in a controlled manner, and includes semantic definitions of the elements and ways of encoding them constitutes a *metadata standard*

Discovery Broker Components



WHOS: ad hoc metadata mapping



Metadata in WHOS - the brokering approach

- As in the case of data, WHOS does not impose (but only recommends) the adoptions of standards for metadata publishing.
- It deals with the heterogeneity among the metadata services through brokering, which means that for each new *metadata service type* a new *metadata accessor* of the brokering application must be developed.
- From documentation provided by the provider, plus feedback between IT experts, the correct *metadata mapping* is set.
- Metadata element names are mapped as well as the values (semantic mapping).
- If an important metadata element is missing from the source, a recommendation is so that it is added. If this is not possible, sometimes it can be added on-the-fly by the broker.
- Every brokered metadata service is periodically harvested, harmonized and saved into the broker's metadata repository, which allows for optimized data discovery (including semantic discovery).
- WHOS internal metadata model uses ISO 19115 + WIS + extended metadata elements, augmented by the use of WHOS Hydrological Ontology

Minimum metadata requirements to enable brokering

- Metadata must be in a machine-readable format
- For every dataset, metadata must contain:
  - Location:
    - Coordinates*
    - Name / identifier
  - Observed property (variable):
    - Name / identifier
    - Units name / identifier*
    - Date range of available observations*
    - Time zone*
    - Observation type (Instantaneous / Interpolated / Aggregated)*
    - Observation spacing or aggregation period*
  - Point of contact:
    - Email / address of responsible party*

*: If unavailable online, it may be provided offline by the data provider

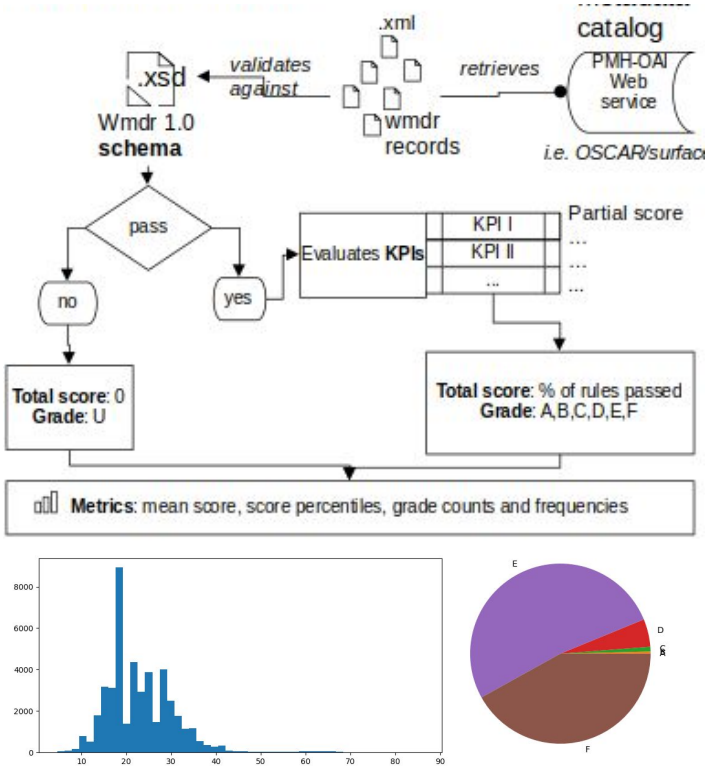
Additional metadata elements (some examples from WMDR and WaterML 2.0)

- In order to enhance the quality of the metadata records, additional elements may be added, such as:
  - Dataset collection:
    - Language
    - Character set (encoding)
    - Metadata standard name + version
    - Date of production (dateStamp)
    - Additional contacts
    - Data policy
  - Dataset:
    - Sampled medium
    - Number of observations
    - Program affiliations
    - Application area(s)
    - Near-real-time availability
  - Location:
    - WSI (WIGOS station identifier)
    - Site description
    - Country
    - WMO Region
    - Facility URL
    - Altitude
  - Observation process:
    - Type
    - Reference (to documentation)
    - Vertical datum
    - Bias
    - Deployments
    - Source (manual/automatic)
    - Instrument details
    - Instrument operating status
    - Uncertainty
    - Equipment specification URL
  - Processing:
    - Data processing method
    - processing/analysis centre
    - Timeliness
    - Numerical resolution
    - Level of data
    - Data format (and version)
    - Traceability to a standard
    - Data quality flagging system
    - Primary observation
  - Value:
    - Quality (data quality code)

Metadata quality assessment

- Schema encoding compliance (within a given standard). Will not pass if the encoding does not correspond to the schema (including the absence of a mandatory element).
- Richness of metadata. Grade according to a scale based on optional elements.
  - Grade higher if a given optional element is present
  - When using controlled vocabularies, grade higher if a given optional element value is present in the CV
  - If multiplicity is allowed, grade higher if more items are present
- WMO's TT-WIGOSMD developed a set of Key Performance Indicators (KPIs) based on these criteria. Additionally an application that computes the KPIs for a given WMDR record has been developed.
- Given that WHOS broker's metadata mapper is able to deliver WMDR records, WHOS metadata quality may be assessed according to the WIGOS standard.
- Additional quality assessment guidelines/tools could be developed according to other standards, such as OGC WaterML 2.0 or ISO 19115

https://github.com/wmo-im/wmdr/tree/issue42/kpi
https://github.com/wmo-im/pywmdr
https://github.com/wmo-im/tt-w4h/tree/main/whos-metadata



Semantic mapping

- When implementing a new metadata service type, each available variable name/identifier should be mapped to the WHOS hydrology ontology
- This may include adding new synonyms and translations of existing concepts, as well as new concepts to the ontology
- The same goes for names/identifiers of measurement units

http://gs-service-production.geodab.eu/gs-service/services/essi/view/whos-plata/semantic

Nivel del Agua	http://hydro.geodab.eu/hydro-ontology/concept/3
Temperatura del Agua	null
Precipitación Acumulada	http://hydro.geodab.eu/hydro-ontology/concept/65
Conductividad	null
Caudal	http://hydro.geodab.eu/hydro-ontology/concept/76
Humedad del Suelo	http://hydro.geodab.eu/hydro-ontology/concept/138
Permitividad del Suelo	null

Hydro-ontology http://hydro.geodab.eu/hydro-ontology/concept/scheme

Concept URI: http://hydro.geodab.eu/hydro-ontology/concept/138

Concept name: Soil Moisture

Concept id: 138

Synonyms

- Moisture content (@en)
- Moisture soil (@en)
- Humedad del suelo (@es)
- Umidità del suolo (@it)
- Humidité du sol (@fr)
- Moisture content (@en)
- Moisture soil (@en)

Broader concepts:

- http://hydro.geodab.eu/hydro-ontology/concept/137 Water content, soil

Hands-on activity (if we have spare time)

- Pick up a machine-readable metadata document that your organisation generates or one from an external source that you frequently use (it may be a data file that contains metadata)
- Identify its format (binary, text (xml, json)…?) and the standard it follows (if any).
- Try to identify the *station* name(s) and identifier(s) and its coordinates.
- Try to identify the *variable* identifier(s) and the source of the vocabulary used (if present). Can we tell which variables are available at each station?
- Try to find the *date range* of the dataset(s). Check the format of the date strings.
- Look for other important metadata elements (country, organisation, time support, contact, units)
- For each element found, try to write down its *path* (its location within the file structure)
- According to the previous slides, how would you qualify this record and how would you improve it? (Container / content)

Questions?

Many Thanks!!

Juan F. Bianchi

Researcher

Instituto Nacional del Agua

Argentina

jbianchi@ina.gob.ar