SeaDataCloud

Review of data formats, also considering INSPIRE data models (O&M)

Ray Cramer [1], Lena Hallin-Pihlatie [2], Kathi Schleidt [3], Riitta Teiniranta [2], Riikka Repo [2], Lorenzo Cognati [4]

[1] British Oceanographic Data Centre, [2] Finnish Environment Institute,

[3] DataCave, [4] Institute of Marine Sciences of the National Research Council of Italy (CNR-ISMAR)

sdn-userdesk@seadatanet.org – www.seadatanet.org

Overall aim

To review and specify how the SeaDataNet NetCDF/ODV formats can be used as a basis for an INSPIRE compliant data format, following O&M

Main tasks

1. Review feasibility of transforming SeaDataNet formats into INSPIRE O&M data standards (following analysis of INSPIRE data implementation rules)
2. Review feasibility of merging CDI metadata into SeaDataNet ODV and NetCDF files to enable delivery of metadata-enriched data sets as part of the CDI service
3. Review implications of migrating from NetCDF V3.6 to V4.0 (time-permitting)
4. Formulate a SeaDataNet NetCDF (CF) format for gridded data, including CDI metadata

Task

- Introduce SDN partners to INSPIRE
  - "INSPIRE themes of relevance in SeaDataCloud will be identified and the data models will be reviewed and introduced to project partners"
  - Result: this ppt-presentation and the report (D8.6)
- More in depth analysis and Proof of Concept mapping
  - Mapping/matching SeaDataNet data and data formats with O&M and other relevant INSPIRE data models
    - What are the contents of the source data?
      - » Finnish Algaline and BODC data
    - Which are the relevant INSPIRE data models and what are their contents?
    - Is something missing? Can the missing pieces be retrieved from the CDI metadata or from somewhere else?
  - Result: Matching tables and Example GMLs (report D8.6 annexes)

2. Review feasibility of merging CDI metadata into SeaDataNet ODV and NetCDF files to enable delivery of metadata-enriched data sets as part of the CDI service

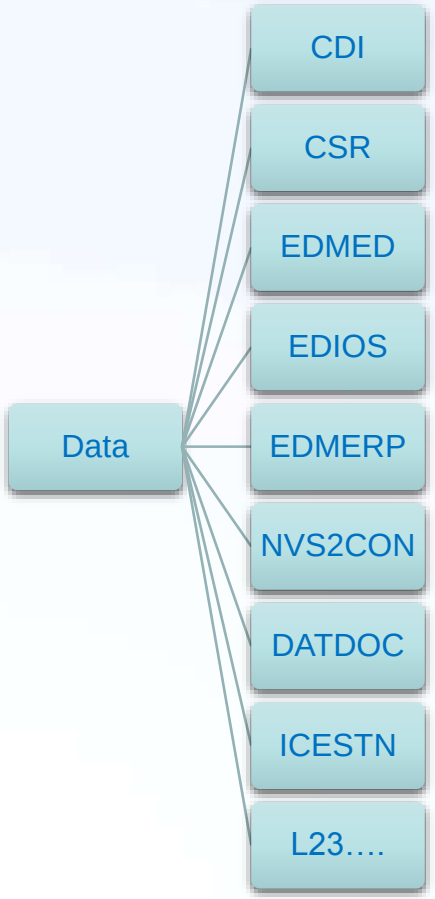
- The metadata enrichment options
  - Open Data
  - Additional metadata, Metadata links
- Current SeaDataNet specification on metadata linkage
- Enhancing the XLINKS system in SeaDataNet
  - Xlink:type
  - Xlink:role
- SeaDataNet software support
- Appendix 1 –The pros and cons of metadata enrichment / linked data

Linking Metadata - need to know three things:

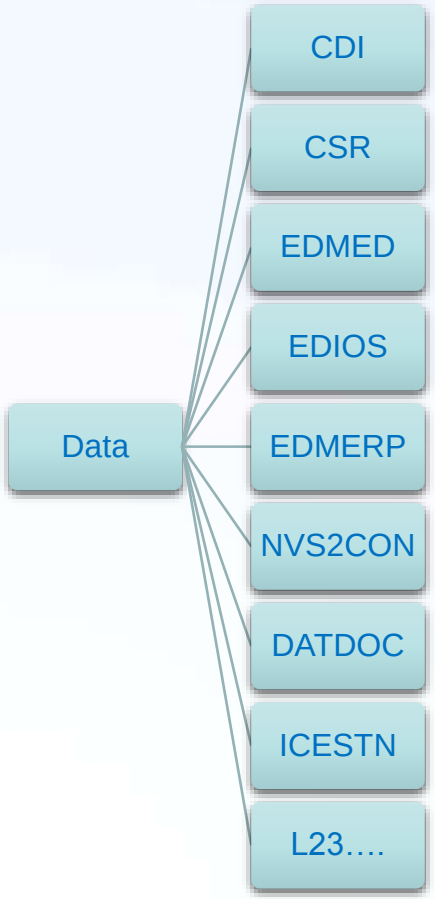
- Where is the metadata resource (URN/URL)
- What is the metadata resource (L23)
- What can we do with the metadata resource - subset of *Internet Assigned Numbers Authority* (IANA) Media Types (aka MIME type)

At least three locations for linked metadata:

- SDN repositories for metadata resources (CDI, CSR, EDMED, EDMERP, NVS2CON)
- SDN local data centre metadata resources (DATDOC)
- Non-SDN – other metadata resources (ICESSTN)



MARIS	**http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?**	**XML**
BSH	**http://seadata.bsh.de/cgi-csr/XML/xmlDownload_V2.pl?**	**XML**
BODC	**https://www.bodc.ac.uk/resources/inventories/edmed/report/**	**XHTML**
MARIS	***http://seadatanet.maris2.nl/v_edios_v2/***	**XHTML**
MARIS	***http://seadatanet.maris2.nl/v_edmerp/***	**XHTML**
BODC	**http://vocab.nerc.ac.uk/collection/C17/current/**	**XML**
LOCAL	**https://www.bodc.ac.uk/data/documents/series/**	**XHTML**



MARIS	**http://cdi.seadatanet.org/[n_code]**
BSH	**http://csr.seadata.org/[n_code]**
BODC	**http://edmed.seadatanet/[n_code]**
MARIS	**http://edios.seadatanet.org/[n_code]**
MARIS	**http://edmerp.seadatanet.org/[n_code]**
BODC	**http://vocab.nerc.ac.uk/collection/C17/current/**
LOCAL	**https://www.bodc.ac.uk/data/documents/series/**

3. Review implications of migrating from NetCDF V3.6 to V4.0

- **Current situation: SeaDataNet NetCDF profile based on CF 1.6 defined for:**
- **Profile** (x, y, t fixed; z variable, e.g. single CTD, but easily modified to allow multiple profiles)
- **TimeSeries** (x, y, z fixed; t variable, e.g. single current meter record, but easily modified to allow multiple time series)
- **Trajectory** (x, y, z, t all variable, specified for a single trajectory, but easily modified to allow multiple trajectories)

- **Noting:** Data Transport Formats manual says "*Significant list discussion focussed on the version of NetCDF that should be used for SeaDataNet. The conclusion was that NetCDF 4 should be used wherever possible, but that NetCDF 3, although strongly discouraged, should not be totally forbidden*."

Benefits of moving to netCDF-4:

- Strongly recommended by the existing SDN Data Transport Formats manual for profile data
- Data files may have to be reprocessed anyway to add attributes for INSPIRE compliance and metadata enrichment
- Allows for data compression
- Required for grids; advantageous for grey areas like VM ADCPs
- Could solve some current format issues thanks to the netCDF-4 feature of user-defined groups and variables

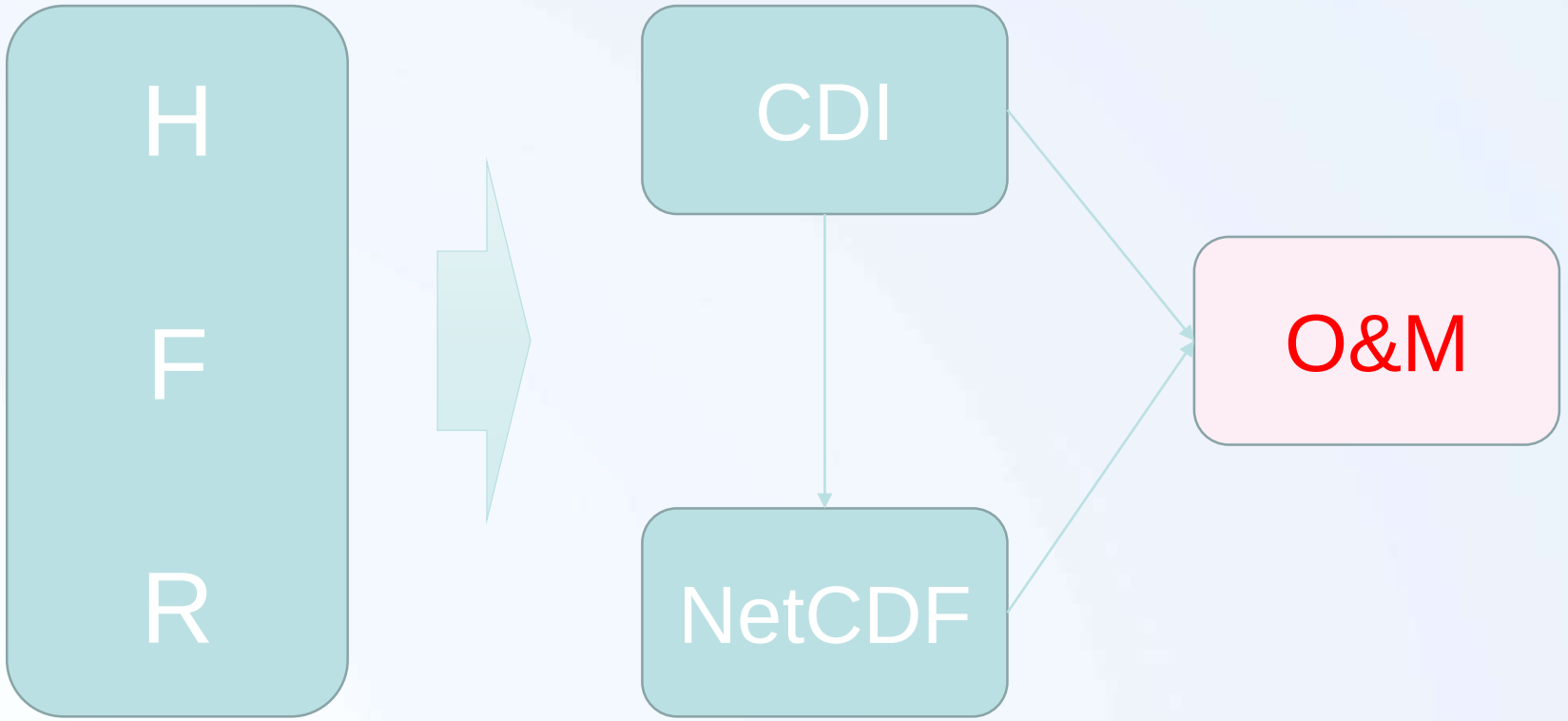
Disadvantages:

- Reprocessing of data file stock
- Upgrading software
- No advantage for some data types (e.g. profile - CTD)

- Note: if the maximum backward compatibility with netCDF-3 datasets and software is required, the best choice would be the netCDF-4 classic model. This solution will not support multiple unlimited dimensions, user-defined types, groups, etc., but acts just like a classic netCDF file.

4. Formulate a SeaDataNet NetCDF (CF) format for gridded data, including CDI metadata

- Draft document that specifies a basic CF grid profile with the SeaDataNet extensions added (P01/P06 semantic labelling etc.)
- Reviewed simple NetCDF gridded data (e.g. GEBCO, numerical model output)
- Scheme defined as European standard model for HF Radar data - CF-1.6, OceanSITES and INSPIRE compliant. (See: Jerico-Next D5.13, http://www.jerico-ri.eu/download/jerico-next-deliverables/JERICO-NEXT-Deliverable-5.13_V1.pdf)


H
F
R
CDI
NetCDF
O&M

SDN extensions to CF (profiles spec)

LATITUDE:sdn_parameter_urn = "SDN:P01::ALATZZ01" ;
LATITUDE:sdn_parameter_name = "Latitude north" ;
LATITUDE:sdn_uom_urn = "SDN:P06::DEGN" ;
LATITUDE:sdn_uom_name = "Degrees north" ;

LATITUDE:grid_mapping = "crs";

-- CRS Linkage - mandatory for SeaDataNet

```
int crs ;
    crs:grid_mapping_name = "latitude_longitude";
    crs:epsg_code = "EPSG:4326“ ;
    crs:semi_major_axis = 6378137.0 ;
    crs:inverse_flattening = 298.257223563 ;
```

Profile data file transport format [two specification options] –

The first is to have a fully descriptive format that includes explicit statements of time zone and CRS. The second is to specify the time zone and CRS that must be used.

For pragmatic reasons SeaDataNet has adopted the latter approach.

TIME - Whilst the full ISO8601 syntax addresses time zone, we are also allowing CJD data format which does not.

POSITION - In the case of position, whilst it is possible to build CRS conversions into data tooling, this is a major task considered beyond SeaDataNet's current resources.

Worthy of note?

INSPIRE CRS - ETRS89-LAEA

```
int crs ;
        crs:grid_mapping_name = " lambert_azimuthal_equal_area";
        crs:longitude_of_projection_origin = 10;
        crs:latitude_of_projection_origin = 52;
        crs:false_easting = 4321000;
        crs:false_northing = 3210000;
```

Thank you

Questions / Comments

sdn-userdesk@seadatanet.org – www.seadatanet.org