

Progress in SeaDataCloud

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

**ALEXANDRA KOKKINAKI
BRITISH OCEANOGRAPHIC DATA CENTRE**



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Overall aim

To review and specify how the SeaDataNet NetCDF/ODV formats can be used as 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

1. Review feasibility of transforming SeaDataNet formats into INSPIRE O&M data standards (1)

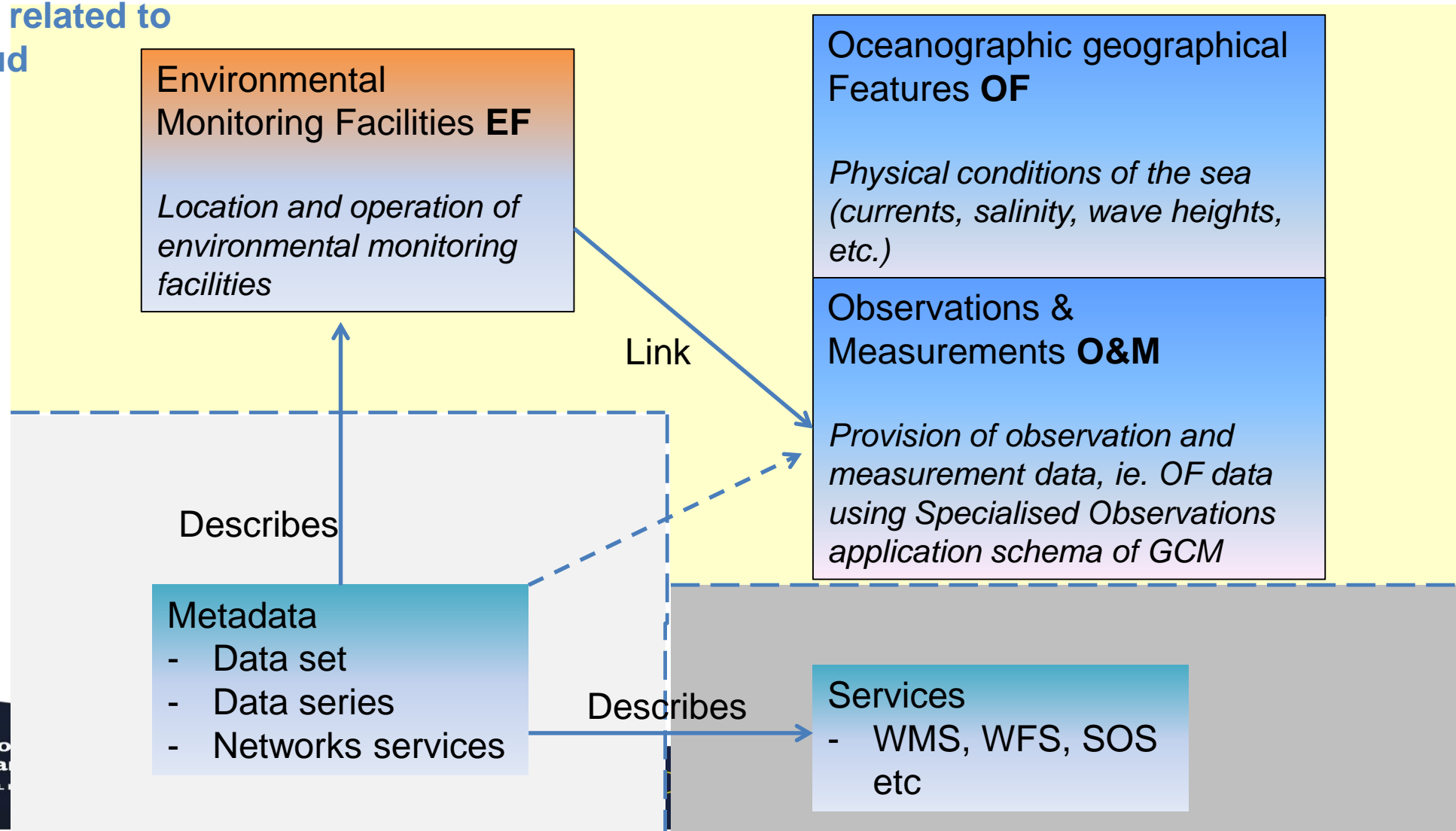
Starting point for the analysis:

- Relevant INSPIRE Technical Guidance documents
 - For example new INSPIRE O&M Guidelines published in 12/2016
- INSPIRE application schemas and examples provided by 52 North
- Previous work
 - INSPIRE Marine Pilot, Geo-Seas project, SeaDataNet II documents
- Example files
 - Finnish Algaline data provided by Seppo Kaitala
 - » raw, processed, SeaDataNet ODV, CDI
 - Examples provided by BODC (Ray Cramer)
 - » profiles, time series, text files (origin NetCDF)



1. Review feasibility of transforming SeaDataNet formats into INSPIRE O&M data standards (2)

INSPIRE themes and other components related to SeaDataCloud

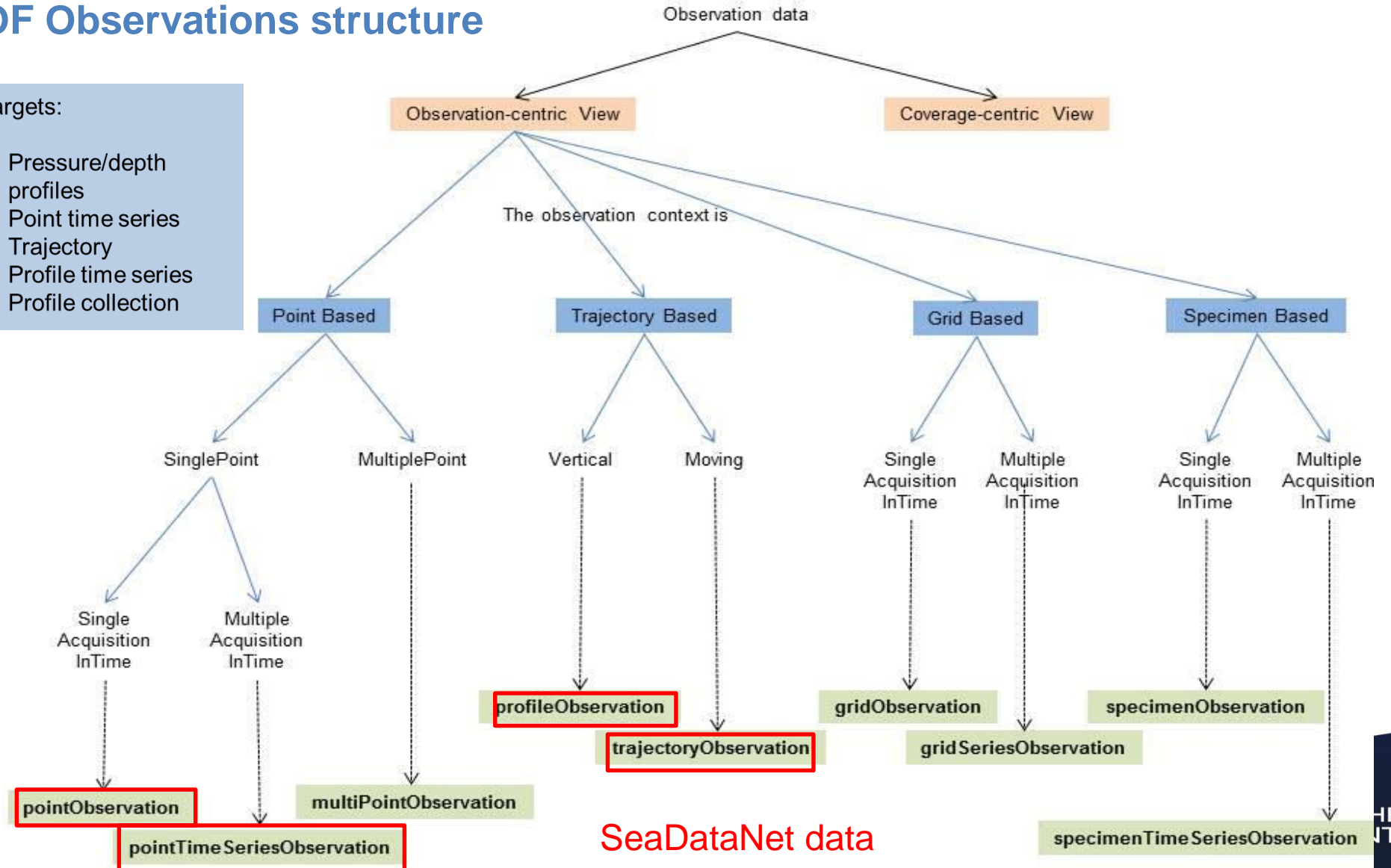


1. Review feasibility of transforming SeaDataNet formats into INSPIRE O&M data standards (3)

OF Observations structure

Targets:

- Pressure/depth profiles
- Point time series
- Trajectory
- Profile time series
- Profile collection



- anyType
- AbstractGML
- AbstractFeature
 - ObservationSet
 - OM_Observation
 - GridObservation
 - GridSeriesObservation
 - MultiPointObservation
 - PointObservation
 - PointTimeSeriesObservation
 - ProfileObservation
 - TrajectoryObservation
 - featureOfInterest
 - metadata (0..1)
 - observedProperty
 - parameter (0..n)
 - phenomenonTime
 - procedure
 - relatedObservation (0..n)
 - result
 - resultQuality (0..n)
 - resultTime
 - type (0..1)
 - validTime (0..1)
 - location (0..1)
 - boundedBy (0..1)
 - description (0..1)
 - descriptionReference (0..1)
 - id
 - identifier (0..1)
 - metaDataProperty (0..n)
 - name (0..n)
- TimeValuePair

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 (1)

- Introduction
 - Open Data
 - Aggregating and sub-setting
- The metadata enrichment options
- Implementing additional metadata
- Implementing 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

Draft text under review

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 (2)

Note: Initial thoughts/challenges gathered from Roy:

- Point NetCDF design allows for packing multiple series into a single file, which makes data to metadata linkages messy.
- NetCDF3 doesn't handle character information (let alone hierarchical information) well.
- Alternatives: URL linkage from NetCDF to the CDI (CSR, EDMED, EDIOS...) XML or O&M file container to enrich the metadata accessible to both data file and CDI.
- **Otherwise – need agreement on what is to be added to the SeaDataNet ODV and NetCDF profiles and how.**

3. Review implications of migrating from NetCDF V3.6 to V4.0 (1)

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."

3. Review implications of migrating from NetCDF V3.6 to V4.0 (2)

Considering 3 perspectives/use cases:

1. SeaDataCloud/SeaDataNet

- Software implications – e.g. NEMO, OCTOPUS, ODV, DIVA,...
- Conversion of existing data file stock
- Product distribution

2. SeaDataNet Users

3. Experience of others, e.g. IMOS, NOAA, ...

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

Benefits of moving to netCDF-4:

- Strongly recommended by the existing Data Transport Formats manual
- 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 formats issues thanks to the feature offered by netCDF-4 of creating 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 for the new format 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 (1)

- Various discussions
- Draft document that specifies a basic CF grid profile with the SeaDataNet extensions added to point data (P01/P06 semantic labelling etc.) incorporated
- Reviewed some examples of existing 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)
- Input from ODIP II requested...

Thank you
QUESTIONS / COMMENTS



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT