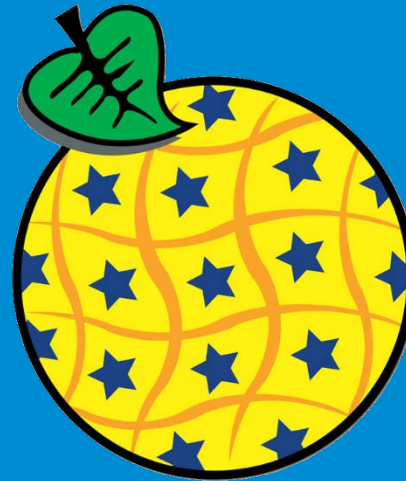


Erfahrungen mit der Datentransformation in INSPIRE Schemata

Florian Hoedt

Thünen-Institut, Zentrum für Informationsmanagement



ETL in a nutshell : Von meinen Daten zu europäischen Daten

- **Ausgang**

- objektrelationale DBMS Schemata, flat-files, proprietäre Formate



- **Ziel**

- Stark hierarchisches XML


Wie sehen ‚meine‘ Daten überhaupt aus

- Hydrographische Daten der Nordsee
- Bodenzustandserhebung Wald
- Biodiversitätsaufnahmen für See / Wald und Land
- ...

Hydrographische Daten der Nordsee

Ausgangsdaten im Detail:

- Punktgeometrien mit Tiefe über water_pressure Feld
- Hierbei ist **je Tiefe** an identischen x/y Koordinaten ein Datensatz (Zeile) vorhanden
- Schiff / Aufnahmefahrt wird über vessel / survey angegeben
- Mehrere chemisch / physikalische Parameter werden je Aufnahmepunkt (x/y + Tiefe) aufgenommen

Id	Name	Alias	Typ
1.2 5	longitude		double
1.2 6	latitude		double
1.2 7	bottom_depth		double
1.2 8	water_pressure		double
1.2 9	water_temperature		double
1.2 10	salinity		double
1.2 11	oxygen_content		double
1.2 12	oxygen_saturation		double
1.2 13	turbidity		double
1.2 14	fluorescence		double
1.2 15	transmission		double
123 3	station		int
 4	datetime		QDateTime
abc 0	survey		QString
abc 1	vessel		QString
abc 2	cruise		QString

Hydrographische Daten der Nordsee



ANNEX III - Oceanographic Geographical Features (OF)

The INSPIRE **Oceanographic Geographical Features** theme (abbreviated to Ocean Features or —OF) describes the physical and chemical characteristics of the sea i.e. properties such as sea surface 'temperature' or 'salinity'. For reporting purposes this type of information is typically presented as a set of point data, e.g. temperature observations from a fixed monitoring station, or as gridded data e.g. sea wave height observations from a satellite.

Besides point and gridded data, other more complex observations are extremely common in oceanography, such as vertical profiles through the ocean depths or trajectories along the ocean surface. [...]

The **Ocean Features theme employs the ISO 19156 Observations and Measurements standard** for consistent encoding of observation-related metadata. There are **three other INSPIRE themes** that are particularly important in their relationships to Oceanographic Geographic Features and these are:

- **Sea Regions (SR):** [...]
- **Environmental Monitoring Facilities (EF):** The process used to derive Oceanographic Geographical Features will involve one or more Environmental Monitoring Facilities (e.g. a ship).
- **Atmospheric Conditions and Meteorological Geographical Features (AC-MF):** On a data level there is a great deal of similarity between measurements made in the seas and oceans and measurements made in the atmosphere so the OF and AC-MF data models have been harmonised as far as possible. Both are based on the same underlying Observations & Measurements model.

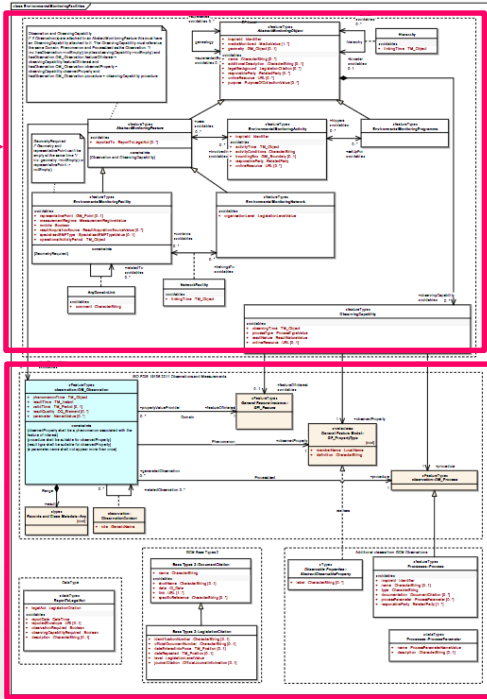


Hydrographische Daten der Nordsee

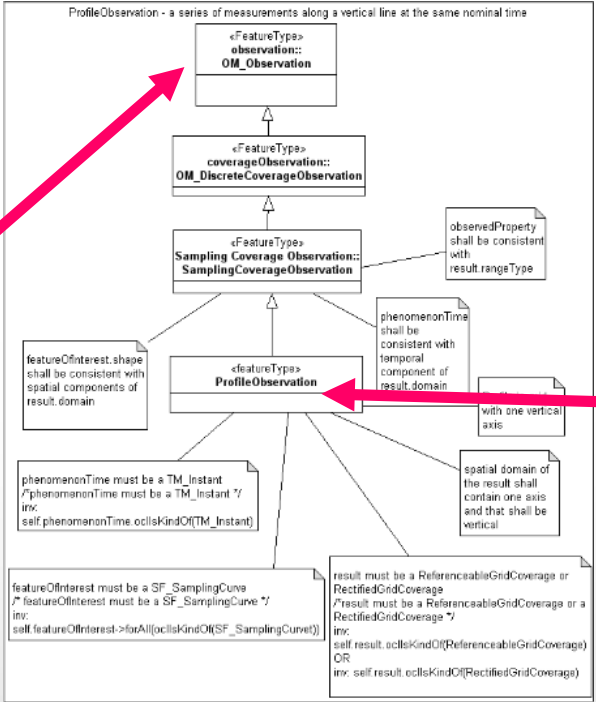
ANNEX III - Oceanographic Geographical Features (OF)



EF →



O&M



OF

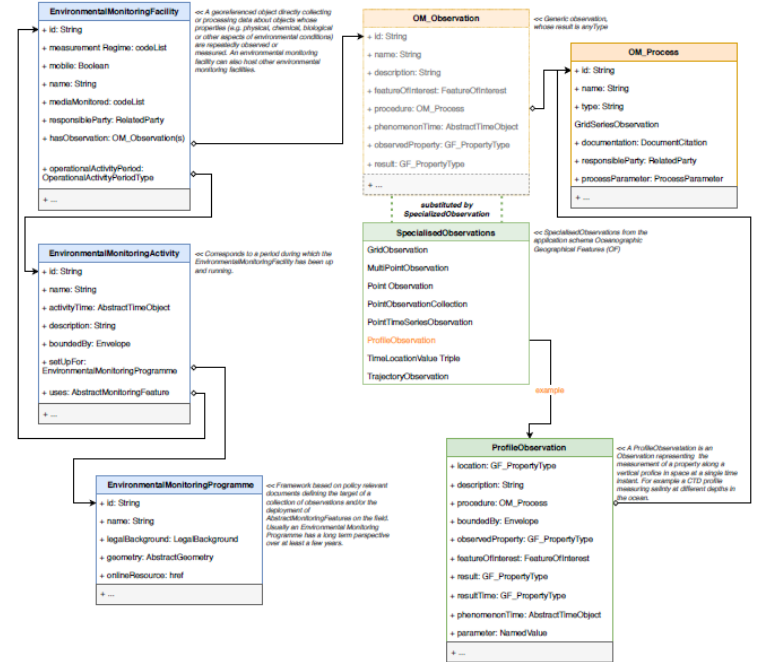
Hydrographische Daten der Nordsee

- Bitte in verständlich
- PDF Diagramm öffnen

Mapping oceanographic survey data to INSPIRE OF schema

The ProfileObservations are SpecializedObservation as defined by OF. These Observations are linked to a Vessel (Ship) which is described by the EF EnvironmentalMonitoringFacility schema via the list hasObservation attribute. Each observation belongs to a specific cruise which will be described through the EnvironmentalMonitoringActivity schema. The legal framework for performing named ProfileObservations could be linked by EnvironmentalMonitoringActivities setByFor Attribute linking to an EnvironmentalMonitoringProgramme

ApplicationSchemas
EF: Environmental Monitoring Facilities
OM: Observations & Measurements
GF: Oceanographic Geographical Features



Hydrographische Daten der Nordsee

Zusammengefasst bedeutet dies

- **Die derzeitigen 6800+ Datensätze müssen zu:**
 - {Anzahl Schiffe} EF.Facilities
 - {Anzahl Fahrten} EF.Activities
 - (n / {Anzahl Tiefen je Profil}) OF.ProfileObservations
 - >> mit jeweils sieben OM.rangeSets (aufgenommene Parameter)

transformiert werden.

Hydrographische Daten der Nordsee

Wir kennen das Ziel!

Transformation mit:



Hydrographische Daten der Nordsee

Transformation mit HALE

- **Naïve Approach**

- OF Schema laden und die jeweiligen Felder zuordnen
 - *Schema lädt EF und O&M nicht mit.*
>> Es kann keine Feldzuordnung stattfinden!

- Mehrere Schemata einladen



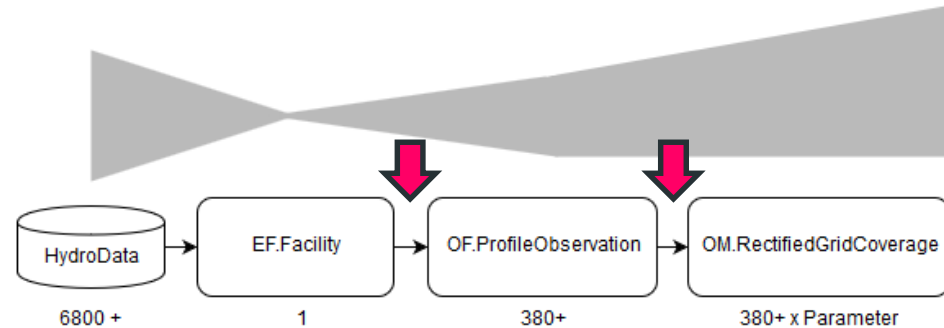
Validation of the provider configuration failed:
Loading multiple XML schemas not supported, please create a combined XML schema instead.

- >> **Von Hand ein kombiniertes Schema erstellen**



Hydrographische Daten der Nordsee

- **1 x EF und 6800+ x ProfileObservation**



- Bei einem ‚normalen‘ Merge werden bestimmte Attribute später nicht mehr aufrufbar im Tool
- Anstelle des ‚normalen‘ Merge muss ein Groovy Script mit Collector eingesetzt werden, um Daten zu sammeln und später auszugeben
 - *Umständlich und nicht gut beschrieben in der Dokumentation*

Hydrographische Daten der Nordsee

- **OF.Result als AnyType im Schema (xsd) definiert**
 - AnyType wird von HALE nicht richtig gelesen; Technical Guidance Document sagt RectifiedGridCoverages sollen verwendet werden (**nicht** in xsd definiert!)
 - >> *Manuelles Anpassen der XML Konfigurationsdateien nötig*



FlorianHoedt 5 months ago

Hello,

I ran into a new issue. As stated in the [Data Specification](#) on page 34 of OF the result of an ProfileObservation should be a RectifiedGridCoverage (or ReferenceableGridCoverage). But the `xsd` only references an AnyType. Therefore HALE is not able to instantiate a RectifiedGridCoverage at `EnvironmentalMonitoringFacility.hasObservation.ProfileObservation.result`

Any Idea how to solve this?



Simon Templar 4 months ago

Hi Florian,

you can configure hale so that it know what kind of element you expect in the anyType field. Sadly not via the UI. An example can be found here: <https://discuss.wetransform.to/#/thread/rjmq/sos-target-schema-without-SWE-schema>

Best,
Simon

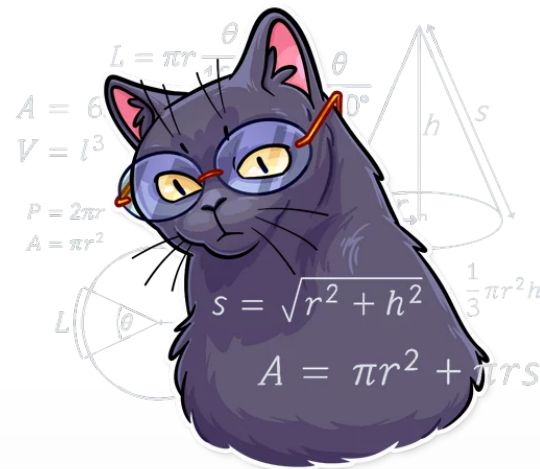
Hydrographische Daten der Nordsee

- **OF.Result als RectifiedGridCoverage**
 - Eulen nach Athen: *INSPIRE Schemata unübersichtlich*

Die Geometrien

Die aufgenommenen Werte

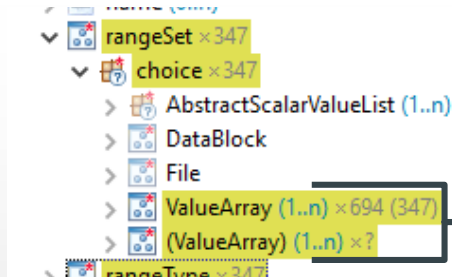
Das „Intervall“ der Geometrien



Hydrographische Daten der Nordsee

- **OF.Result als RectifiedGridCoverage**

- Wieder das Problem HALE verständlich zu machen, dass aus einer Zeile Input, mehrere ValueArray Instanzen gebildet werden müssen:
 - **Instance Context** kann verwendet werden (eigentlich für andere Anwendungsfälle gedacht)



Je Parameter ist jeweils ein ValueArray zu instanzieren



Hydrographische Daten der Nordsee

- Wie sieht das Ergebnis als GML aus?
 - **GML Datei in SublimeText öffnen**

```
1 <!-- .. Schnipp : all der andere OF / EF Kram ... -->
2 <om:result>
3   <gml:cov:RectifiedGridCoverage>
4     <gml:multiPointDomain>
5       <gml:MultiPoint>
6         <gml:pointMember>
7           <gml:Point srsName="urn:ogc:def:crs:EPSG::4326" srsDimension="3">
8             <!-- Die 1. Position (3m Tiefe) | Daten sind *nicht* nach Tiefe geordnet! -->
9             <gml:pos>7.2568 53.8373 3.0</gml:pos>
10          </gml:Point>
11        </gml:pointMember>
12      <!-- .. Schnipp : alle die Anderen 3D Punkte ... -->
13    <gml:rangeSet>
14      <!-- erster Parameter : Temperatur -->
15      <gml:ValueArray>
16        <gml:valueComponent>
17          <!-- gehört zur 1. Position (3m Tiefe) -->
18          <gml:Quantity uom="°">5.277</gml:Quantity>
19        </gml:valueComponent>
20      <!-- .. Schnipp restliche Temperaturen ... -->
21    </gml:ValueArray>
22    <!-- zweiter Parameter : Salinität -->
23    <gml:ValueArray>
24      <gml:valueComponent>
25        <gml:Quantity uom="mg/l">33.536</gml:Quantity>
26      </gml:valueComponent>
27    <!-- .. Schnipp restliche Salzgehalte ... -->
```



Hydrographische Daten der Nordsee

Sie befinden sich Hier?

Und nun?

- **GML veröffentlichen über:**

- Atom Feed
- Deegree FeatureStore als Memory Layer



- **HALE Alignment nutzen für:**

- GeoServer AppSchema Extension
- HALE Connect



Die Welt besteht nicht nur aus HALE



Stefanie Richter,
Dr. Einar Eberhardt



Bundesanstalt für
Geowissenschaften
und Rohstoffe

<https://themes.jrc.ec.europa.eu/groups/profile/1821/soil>

ISSUES:

FME/INSPIRE Solution Pack >> integrate OM_Observation Writers in the INSPIRE Soil Template (>con terra)

Die Welt besteht nicht nur aus HALE



Verwenden der deegree CLI:

```
1 # create ddl for inspire schema
2 # postgres version:
3 $schema="http://inspire.ec.europa.eu/schemas/so/4.0/Soil.xsd"
4 $format='ddl'
5
6 java -jar P:\deegree-cli-utility-1.2-jar-with-dependencies.jar --srid=
   25832 --format=$format --idtype=uuid $schema
7
```

ISSUES:

- Wie bei HALE konnte OF nicht vernünftig bearbeitet werden
>> Fehler des xsd
- Komplettes SQL Schema und daher sehr unübersichtlich
- Keine Validierung des Schemas, viel Eigenarbeit beim Mapping

Andere Tools

Die Welt besteht nicht nur aus HALE

Verwenden von stetl



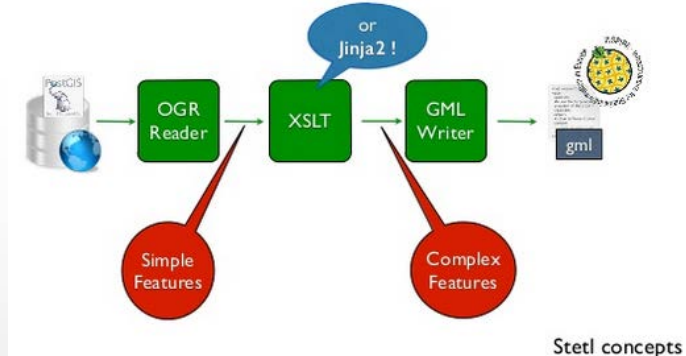
ISSUES:

Ungetestet und ohne Validierung
Transformation muss von Hand per Templating
geschrieben werden (XSLT or Jinja2)

Solution: Add Python to the Equation



Example: Data Model Transform



Transformationen sind komplex

- Hürden in der jeweiligen ETL Software
- Fehler/Unklarheiten in den Zielschemata
- Der Weg von der sauberen ETL Transformation zum sauberen Downloaddienst ist bisher wenig beschrieben

Daher: Lasst uns gemeinsam ‚empor Irren‘

- GESUCHT: Expert*innen für harmonisierte Datenschemata
- Nicht nur Erfolgsgeschichten, auch Fehler publizieren
- Gemeinsame Schulungen zu ETL Software organisieren
- Austausch auf europäischer Ebene:
<https://inspire.ec.europa.eu/forum/discussion/all>