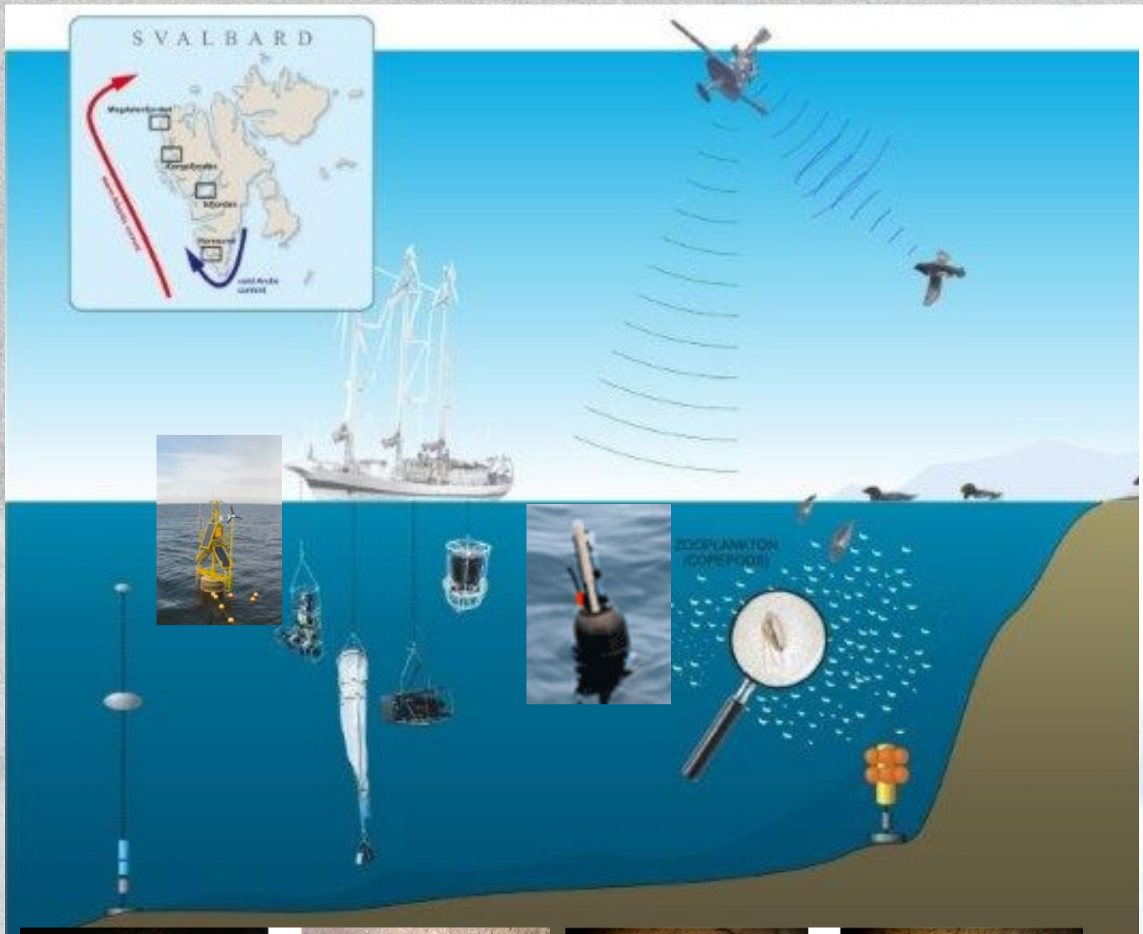
How to accomplish interoperability of heterogeneous concepts? Multifunctional sensors and systems for in-situ monitoring of marine environment

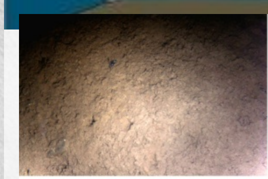
Marcin Wichorowski **Sławomir Sagan**

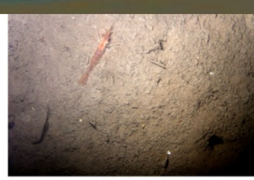
Institute of Oceanology, Polish Academy of Sciences

CODATA 2017 - Global Challenges and Data-Driven Science
St. Petersburg, Russia, October 8–13, 2017

Acquisition of oceanographic data





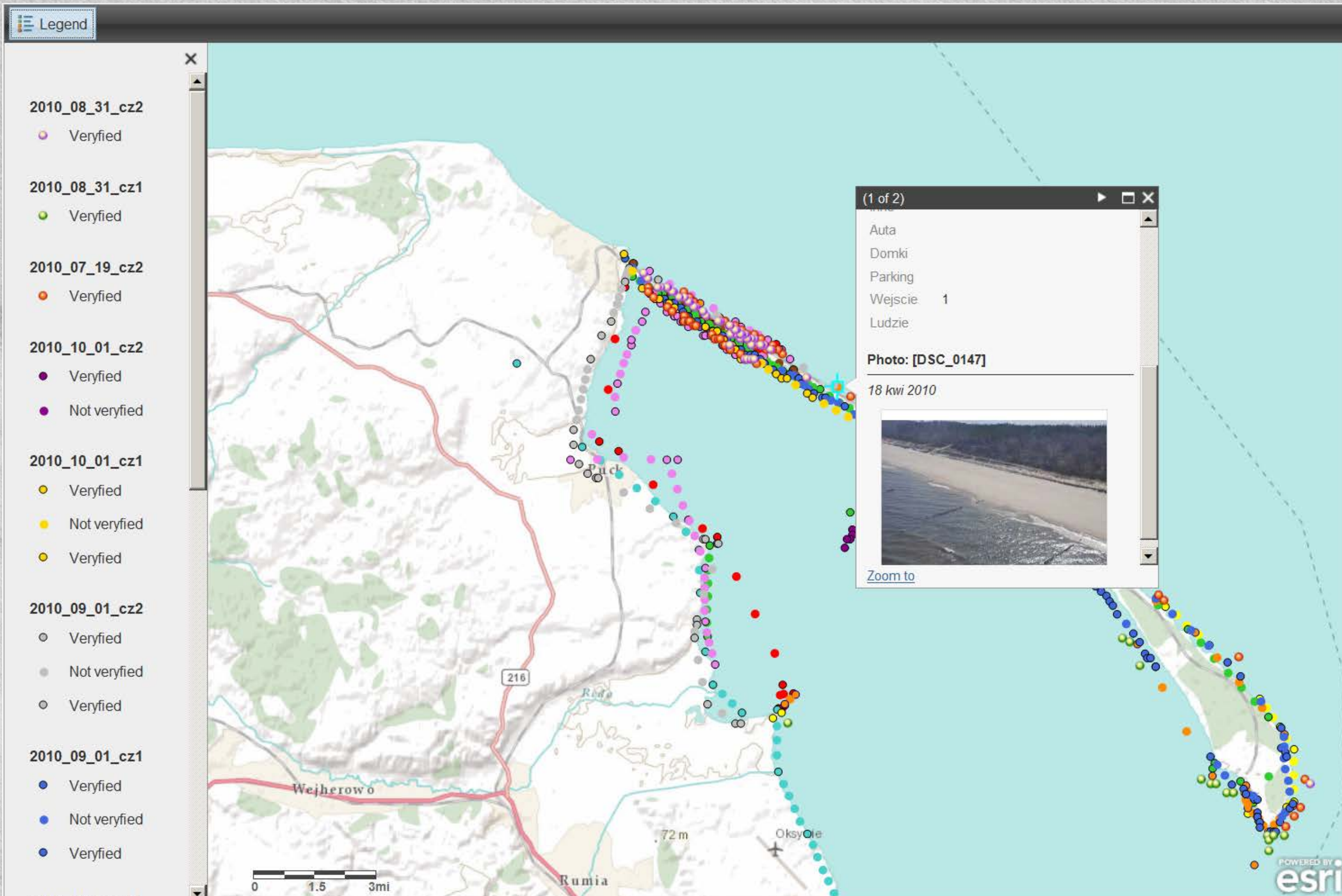


station	year	date	area	depth	longitude	latitude
HG1	2013	30.07.2013	Hornsund	84	16°04,710	77°00,052
HG2	2013	30.07.2013	Hornsund	107	16°01,495	77°00,539
HG3	2013	30.07.2013	Hornsund	105	16°01,108	76°59,599

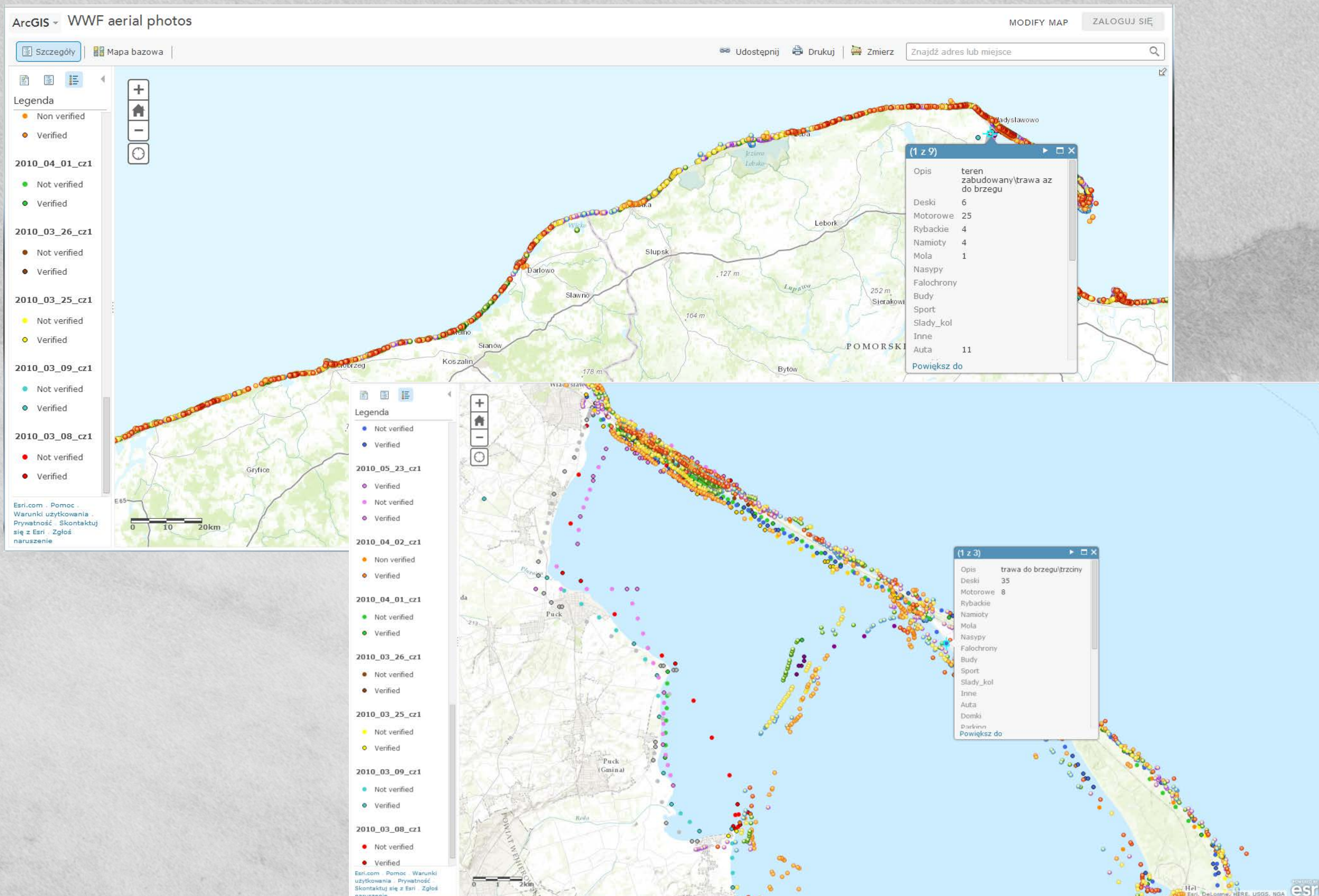
station	year	date	area	depth	longitude	latitude
KG1	2013	07.08.2013	Kongsfjord	90	12°08,714	78°55,881
KG2	2013	08.08.2013	Kongsfjord	115	12°04,252	78°55,236
KG3	2013	08.08.2013	Kongsfjord	98	12°09,641	78°54,904

seabed photos by Kajetan Deja

Acquisition of oceanographic data

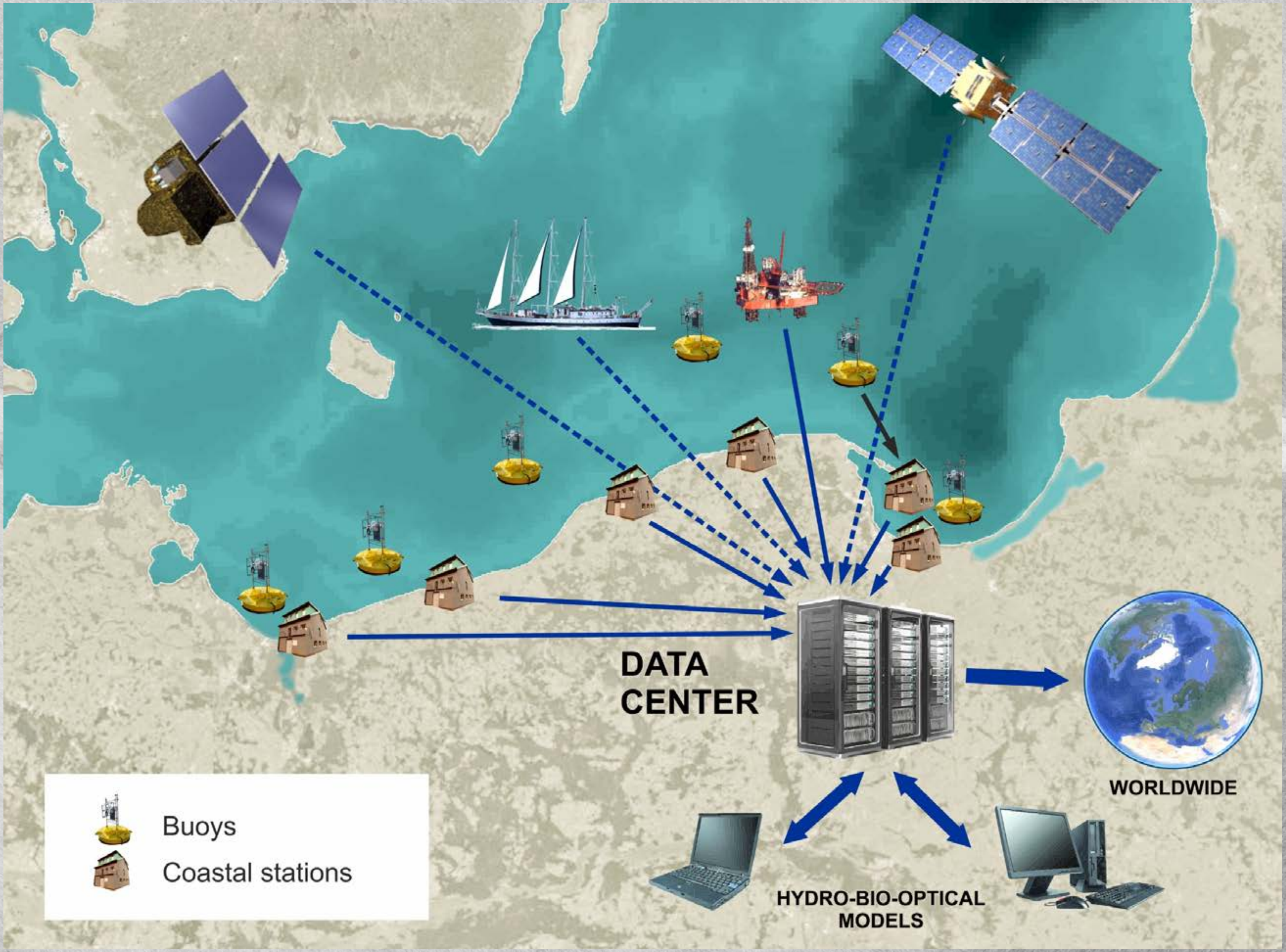


Acquisition of oceanographic data

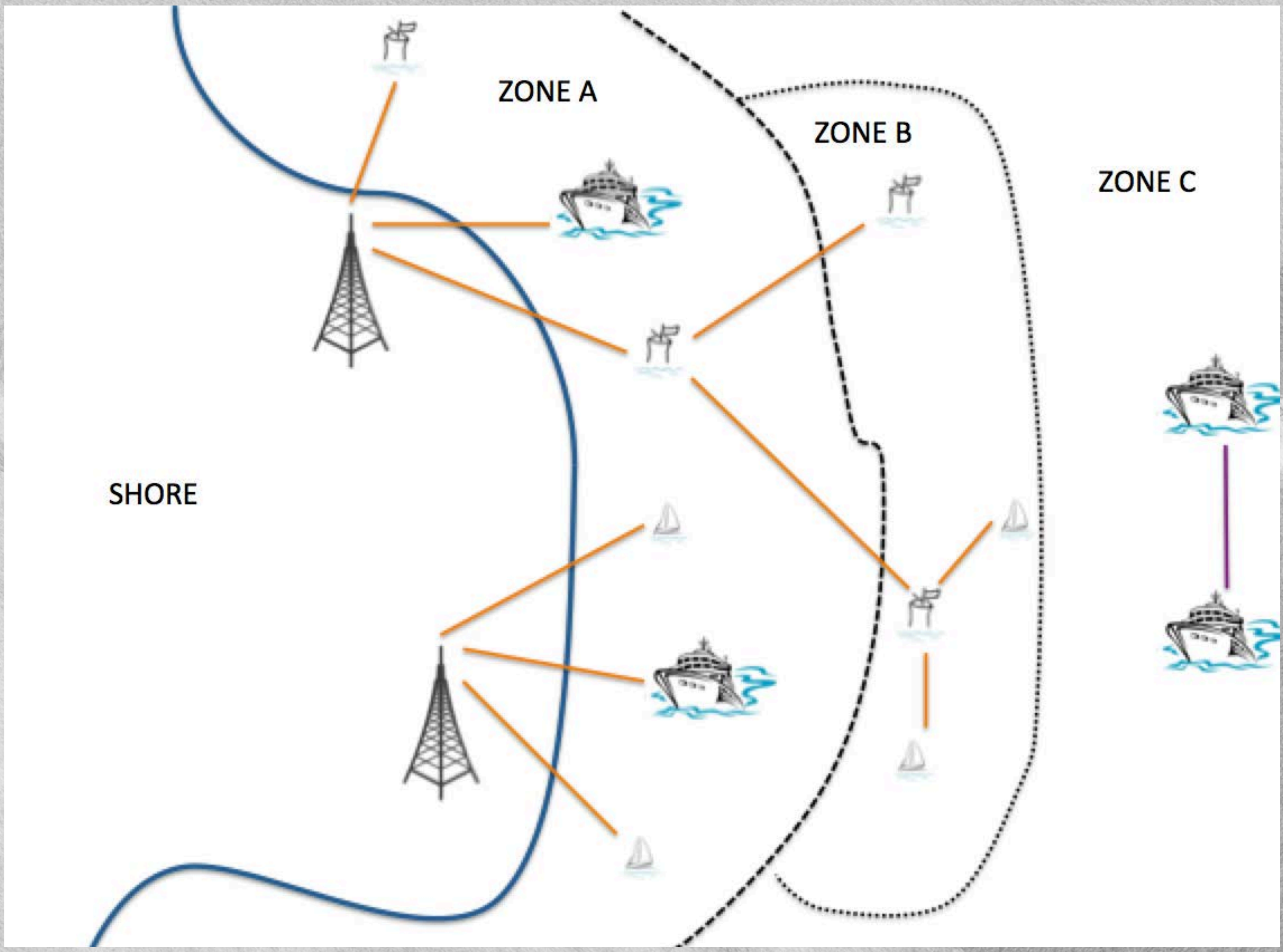


Source: Joanna Pardus, *IO PAN 2012*

Acquisition of oceanographic data



Enabling acquisition of oceanographic data from IoT devices

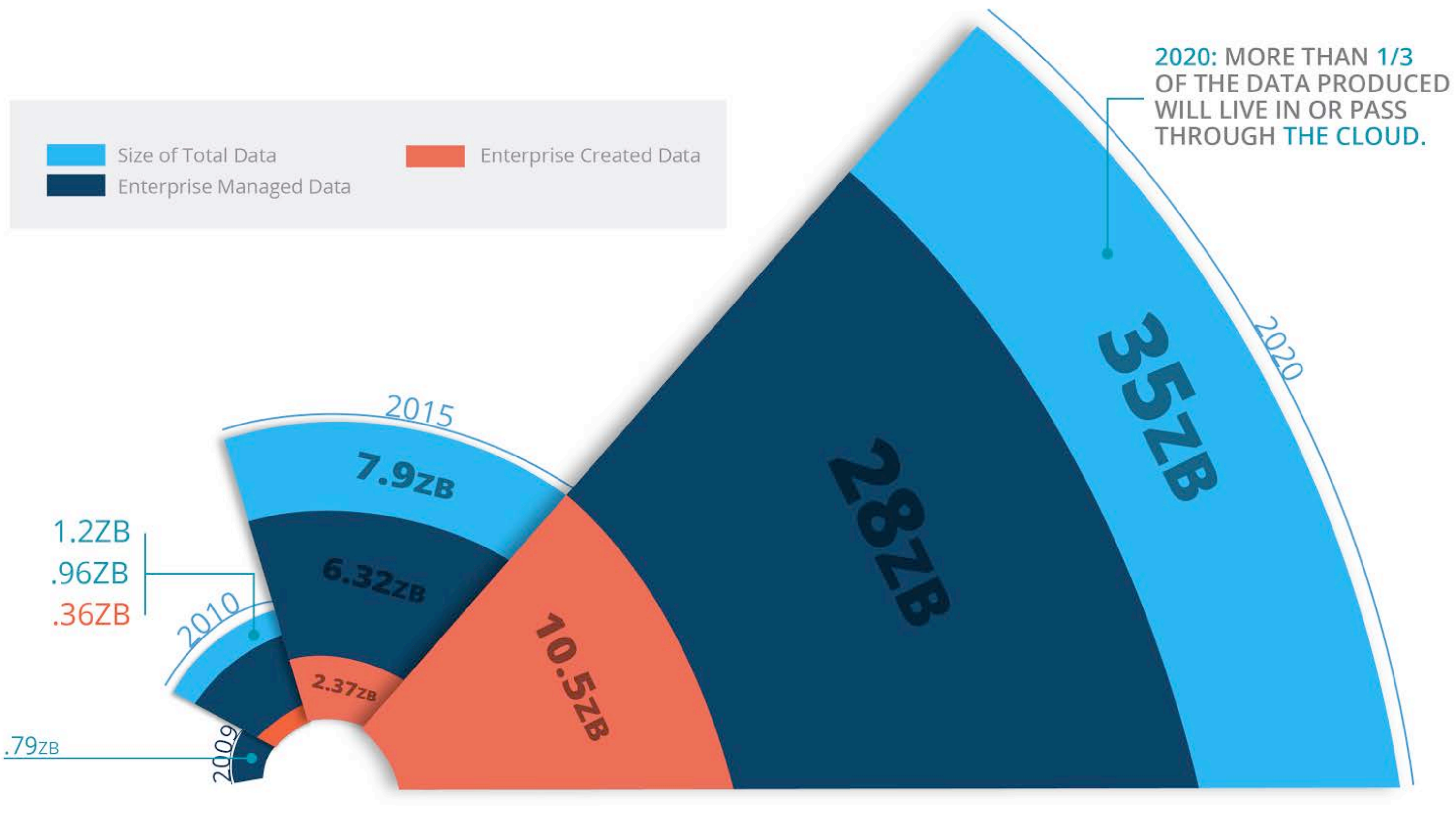


Sentinel-2 Data Volume

- X band downlink with data rate of **490 Mbit/s** (after on-board wavelet compression) → **50x** ENVISAT MERIS/FR
- average **~17 minutes of instrument MSI operation/orbit**
- **800 GB per day** compressed raw data (~170 DVDs), i.e. **400 TB per year** from 1 satellite
- Equivalent **continuous raw data supply rate of ~170 Mbps** (compressed) to be sustained on ground for 2 satellites
- One **100km x 100km portion** of MSI image weights **~ 500 Mbytes** (J2K compressed)

Status of Sentinels -1, -2 and -3 and the Data Policy – CEOS, Frank Martin Seifert, ESA – Earth Observation Programme 4th Space Data Coordination, Caltech, Pasadena, 4 September 2013

Volume of data growth perspective



http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode

4th Paradigm

1. Empirical science
2. Theory and models
3. Numerical modelling
4. Data intensive science

Jim Gray, “The Fourth Paradigm”, Microsoft Research, Redmond

Big Data 7 V's

1. Volume
2. Velocity
3. Variety

Doug Laney's 3 V's

4. Variability
5. Veracity
6. Visualisation
7. Value

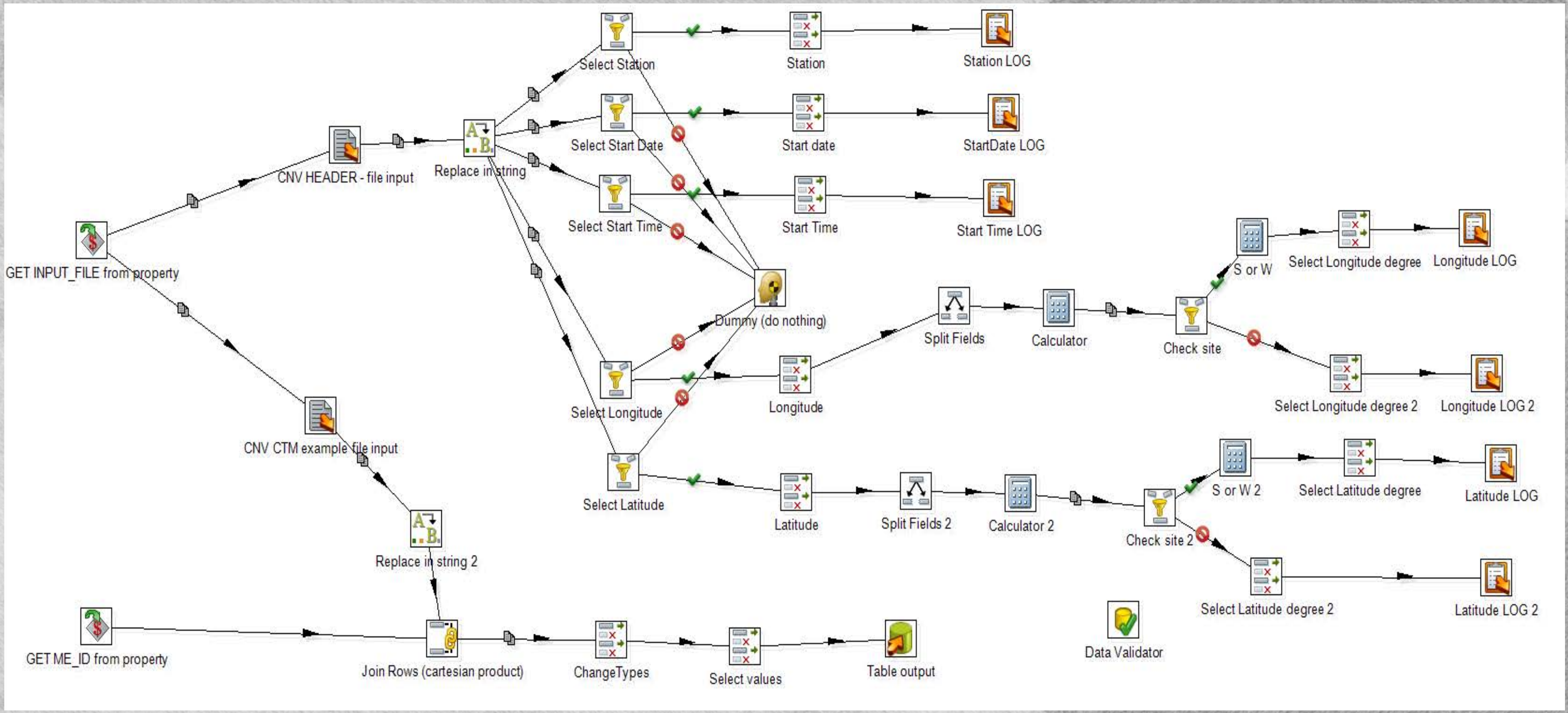
Mark van Rijmenam

Automatization of acquisition processes – ETL tools

```
* Seabird SBE 49 Raw Data File:
* Number of Bytes Per Scan = 6
* System UpLoad Time = Wed Mar 0
* Ship:      Oceania
* Cruise:    Stolpe Channel 2003
* Station:   1rs
* Latitude:  N54 50.707
* Longitude: E19 18.278
* Bottom:    110
# nquan = 7
# nvalues = 210
# units = metric
# name 0 = pr: pressure [deciba
# name 1 = t0: temperature [deg
# name 2 = sal00: salinity, PSS-
# name 3 = sigma-t00: density, s
# name 4 = sva: specific volume
# name 5 = flag:  0.000e+00
# name 6 = nbin: number of scans
# span 0 = 1.000, 105.500
# span 1 = 2.4134, 8.3162
# span 2 = 7.3739, 13.4070
# span 3 = 5.8815, 10.3359
# span 4 = 1709.57, 2148.93
# span 5 = 0.000e+00, 0.000e+00
# span 6 = 13.0000, 201.0000
# interval = decibars: 0.5
# start_time = Wed Mar 07 2007 1
........
# derive_date = Mar 20 2007 09:1
# derive_in = 1RS000.CNV GUILDLI
# strip_date = Mar 20 2007 09:10
# strip_in = 1RS000.CNV
# file_type = ascii
*END*
      1.000     2.5273     7.375
      1.500     2.5464     7.376
      2.000     2.5096     7.375
      2.500     2.4939     7.375
```

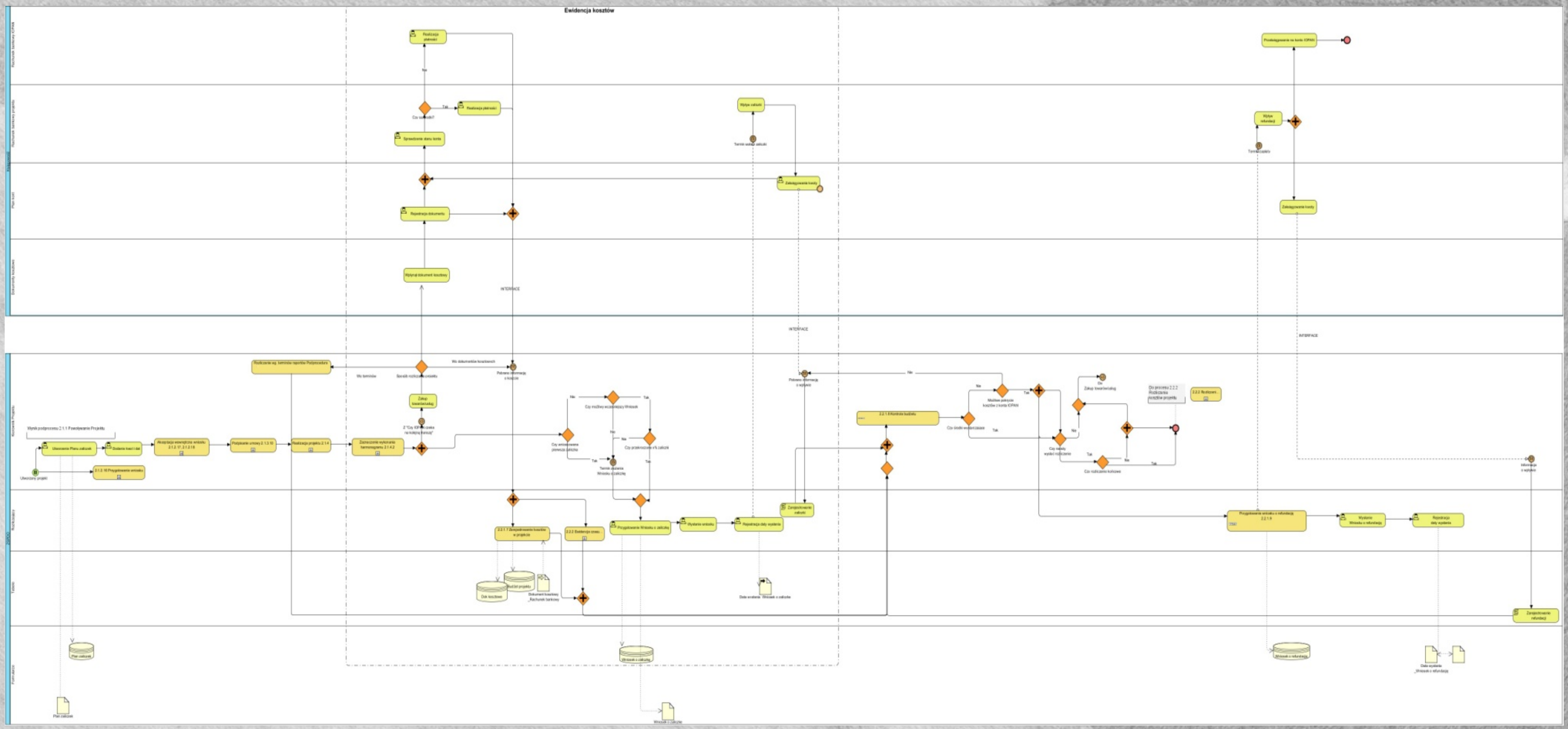

```
<?xml version="1.0"?>
- <transformation>
  - <info>
      <name>Transformation IOPAN test</name>
      <description/>
      <extended_description/>
      <trans_version/>
      <trans_type>Normal</trans_type>
      <trans_status>0</trans_status>
      <directory>/</directory>
    - <parameters>
      - <parameter>
          <name>INPUT_FILE</name>
          <default_value>default</default_value>
          <description>DATA IN</description>
        </parameter>
      - <parameter>
          <name>MEASUREMENTEVENT_ID_PROPERTY</name>
          <default_value>default</default_value>
          <description>Measurement_ID</description>
        </parameter>
      </parameters>
    + <log>
    + <maxdate>
      <size_rowset>10000</size_rowset>
      <sleep_time_empty>50</sleep_time_empty>
      <sleep_time_full>50</sleep_time_full>
      <unique_connections>N</unique_connections>
      <feedback_shown>Y</feedback_shown>
      <feedback_size>50000</feedback_size>
      <using_thread_priorities>Y</using_thread_priorities>
      <shared_objects_file/>
      <capture_step_performance>N</capture_step_performance>
      <step_performance_capturing_delay>1000</step_performance_capturing_delay>
      <step_performance_capturing_size_limit>100</step_performance_capturing_size_limit>
      <dependencies> </dependencies>
      <partitionschemas> </partitionschemas>
      <slaveservers> </slaveservers>
      <clusterschemas> </clusterschemas>
      <created_user>-</created_user>
      <created_date>2013/07/31 09:20:29.637</created_date>
      <modified_user>-</modified_user>
      <modified_date>2013/09/05 10:53:42.204</modified_date>
    </info>
    <notepads> </notepads>
  + <connection>
  + <connection>
  + <connection>
  + <connection>
  - <order>
    - <hop>
        <from>GET INPUT_FILE from property</from>
        <to>CNV BODY file input</to>
        <enabled>Y</enabled>
      </hop>
    - <hop>
        <from>GET INPUT_FILE from property</from>
        <to>CNV HEADER - file input</to>
        <enabled>Y</enabled>
      </hop>
    - <hop>
        <from>CNV HEADER - file input</from>
        <to>Replace in string</to>
```


Building ETL processes with Pentaho Business Analytics



Process management with BPMN v.2

Visual Paradigm / Dysant Framework



Data Integration and Interoperability

What?

Ability of two or more systems to communicate and interact or be used together despite their differences

Why?

- Facilitates exchange and sharing of information
- Increases the availability, access, integration of data
- Facilitates the understanding and usage of data
- Solves heterogeneity (differences)

How?

- **Standards** enable interoperability: standards for data, metadata, services
- Semantic interoperability: **ontology**, **controlled vocabulary**

Heterogeneity at different levels

- **System** (i.e. *interaction between computers* of different OS and databases of different DBMS)
- **Syntactic** (i.e. *differences between formats* such as a GML document and a Shapefile)
- **Schematic** (i.e. *differences in conceptual schemas* such as *sampling* may be defined as a class or as a value of an attribute of a *measurement* class)
- **Semantic** (i.e. *difference of meaning*, e.g. *temperature, is it sea temperature or air temperature; "coastline" vs. "shoreline"*)

Data Integration and Interoperability

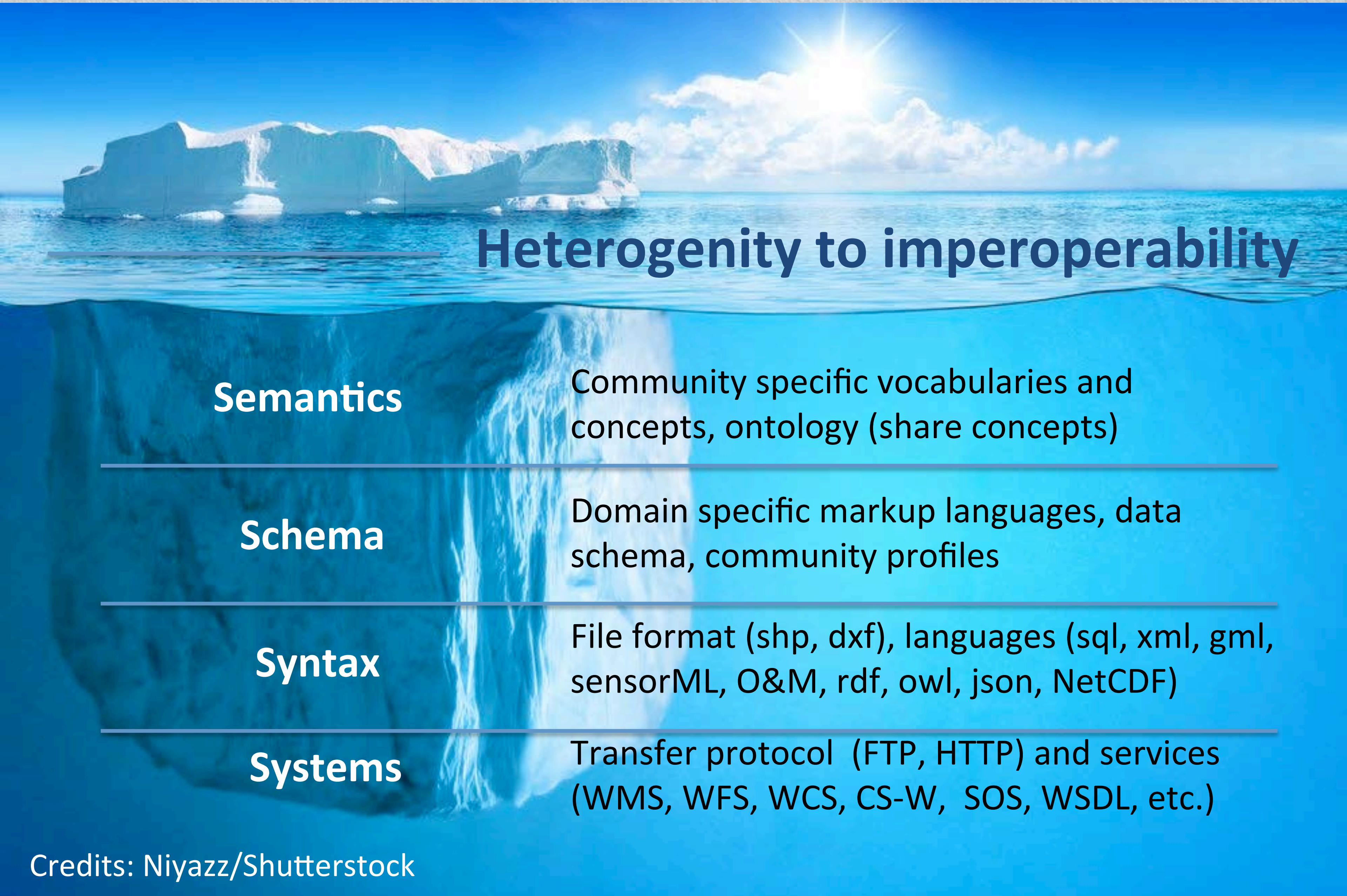
Integration of datasets from heterogeneous data sources (files, databases, web services, etc.), having wide spectra of nature (structured, non-structured data), spatial and temporal resolution, provided via different protocols, unified to enable different systems and platforms to process data and retrieve information

Implementation of monitoring parameters at national levels

Country	Programme	T,P, pH/ pCO2	Heavy Metals	N	µPlastic	uNoise
Denmark	Nation-wide aquatic monitoring programme - monitoring of coastal and open marine waters					
Finland	Monitoring programmes					
France	National sea water quality monitoring network - RNO					
France	French seashore phytoplankton monitoring - REPHY					
Germany	Bund/Länder Messprogramm für die Nordsee					
Greece	MED POL in the Aegean and Ionian Sea and the Saronic Gulf					
Ireland	General Quality of Estuarine and Coastal Receiving Waters					
Ireland	Bathing waters					
The Netherlands	National surface water monitoring programme Monitoring of marine waters					
Norway	Trend monitoring of the Norwegian coastal areas					
Norway	Arctic Monitoring and Assessment (AMAP)the Barents Sea & northern fjords					
Sweden	Nation-wide pelagic frequent monitoring					
UK	UK National Marine Monitoring Plan					

Interoperability on different abstraction layers

- cooperation of data centres,
- data policy enforcement,
- deployment of standards,
- development of controlled vocabularies,
- development of transmission protocols,
- unification of data models
- common sense of datalife cycle plan



Heterogenity to imperoperability

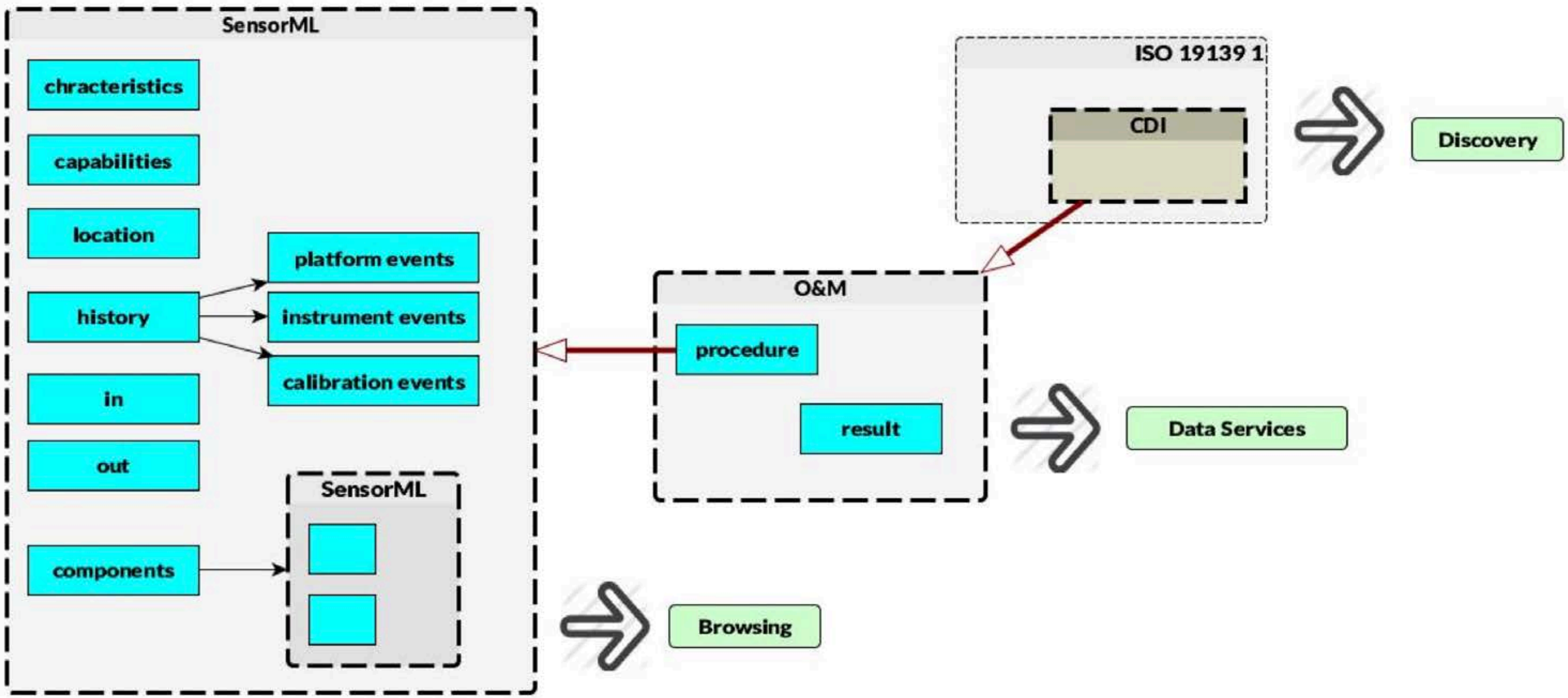
Semantics	Community specific vocabularies and concepts, ontology (share concepts)
Schema	Domain specific markup languages, data schema, community profiles
Syntax	File format (shp, dxf), languages (sql, xml, gml, sensorML, O&M, rdf, owl, json, NetCDF)
Systems	Transfer protocol (FTP, HTTP) and services (WMS, WFS, WCS, CS-W, SOS, WSDL, etc.)

Credits: Niyazz/Shutterstock

OGC's Sensor Web Enablement

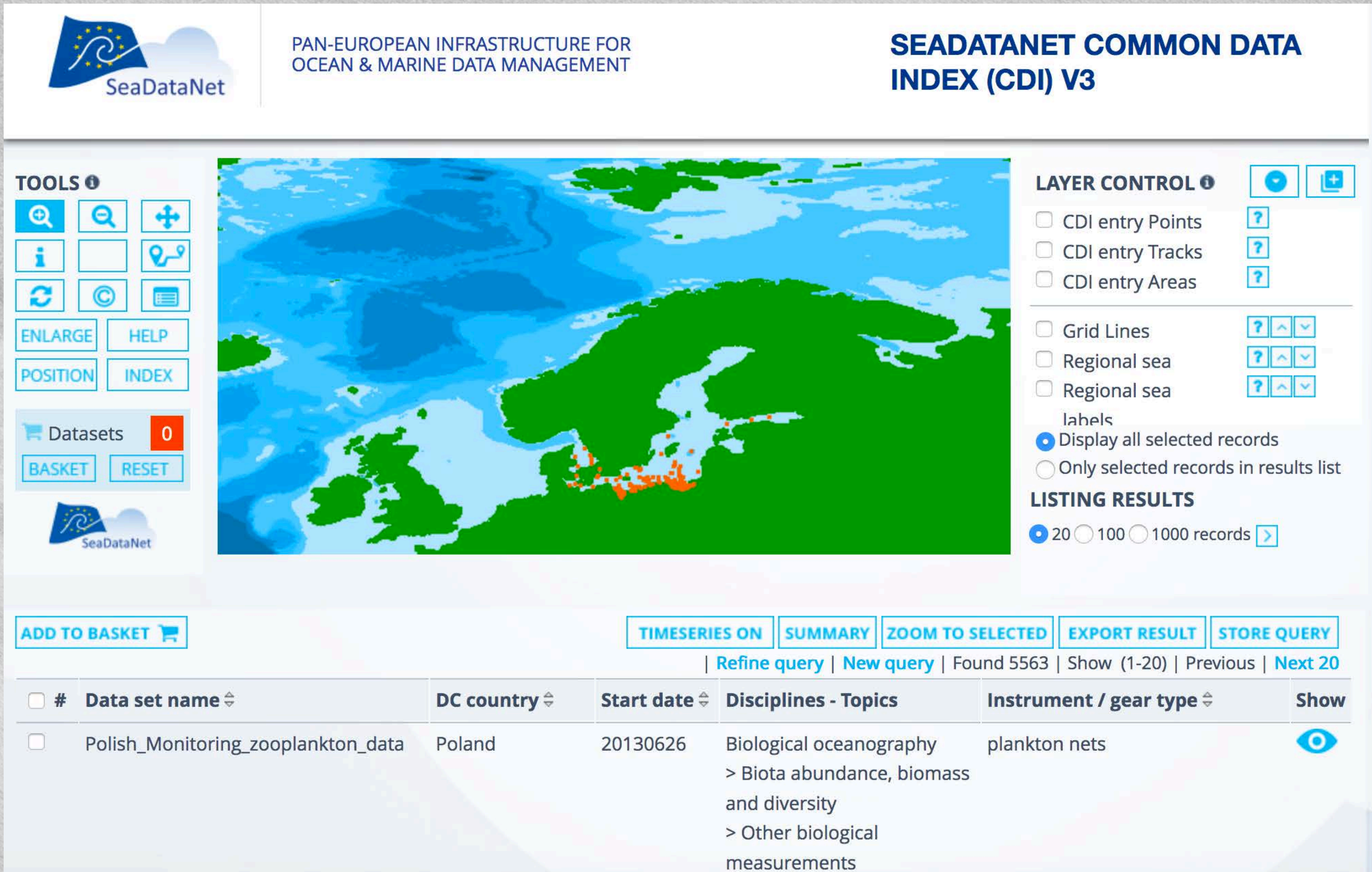
- **SWE Common Data Model** – Defines low-level data models for exchanging sensor related data between nodes of the OGC® Sensor Web Enablement (SWE) framework.
- **Observations & Measurements (O&M)** –The general models and XML encodings for observations and measurements.
- **Sensor Model Language (SensorML)** – Standard models and XML Schema for describing the processes within sensor and observation processing systems.
- **Sensor Observation Service (SOS)** – Open interface for a web service to obtain observations and sensor and platform descriptions from one or more sensors.
- **PUCK Protocol Standard** – Defines a protocol to retrieve a SensorML description, sensor "driver" code, and other information from the device itself, thus enabling automatic sensor installation, configuration and operation
- **Sensor Planning Service (SPS)** – An open interface for a web service by which client can determine the feasibility of collecting data from one or more sensors or models, and submit collection requests.
- **SWE Service Model** -– Defines data types for common use across OGC Sensor Web Enablement (SWE) services. Five of these packages define operation request and response types.

Sensor Observation Service - SeaDataNet CDI service



source: R. Casas, SeaDataNet II – Final Plenary Meeting 16-17 September 2015, Brest

Sea Data Cloud – platform for interoperability assurance



Pan-European infrastructure for ocean & marine data management

https://www.seadatanet.org/


THANK YOU !