



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

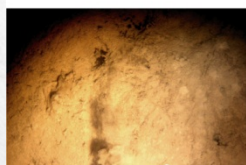
CODATA 2017 - Global Challenges and Data-Driven Science
St. Petersburg, Russia, October 8–13, 2017



Acquisition of oceanographic data



station	year	date	area	depth	longitude	latitude
HG1	2013	30.07.2013	Homsund	84	16°04.710	77°00.052
HG2	2013	30.07.2013	Homsund	107	16°01.495	77°00.539
HG3	2013	30.07.2013	Homsund	105	16°01.108	76°59.599



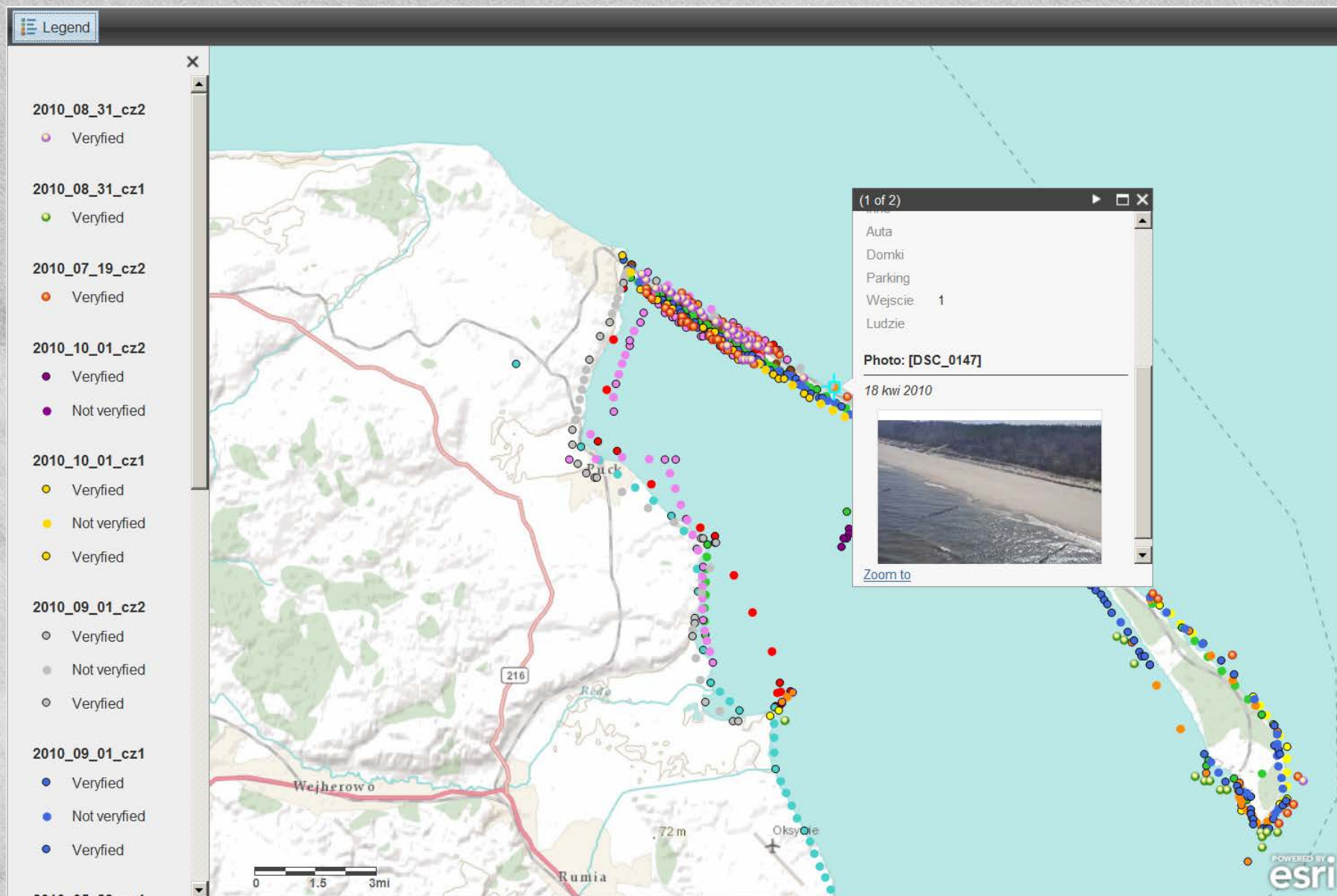
station	year	date	area	depth	longitude	latitude
KG1	2013	07.08.2013	Kongsford	90	12°08.714	78°55.881
KG2	2013	08.08.2013	Kongsford	115	12°04.252	78°55.236
KG3	2013	08.08.2013	Kongsford	98	12°09.641	78°54.904

seabed photos by Kajetan Deja

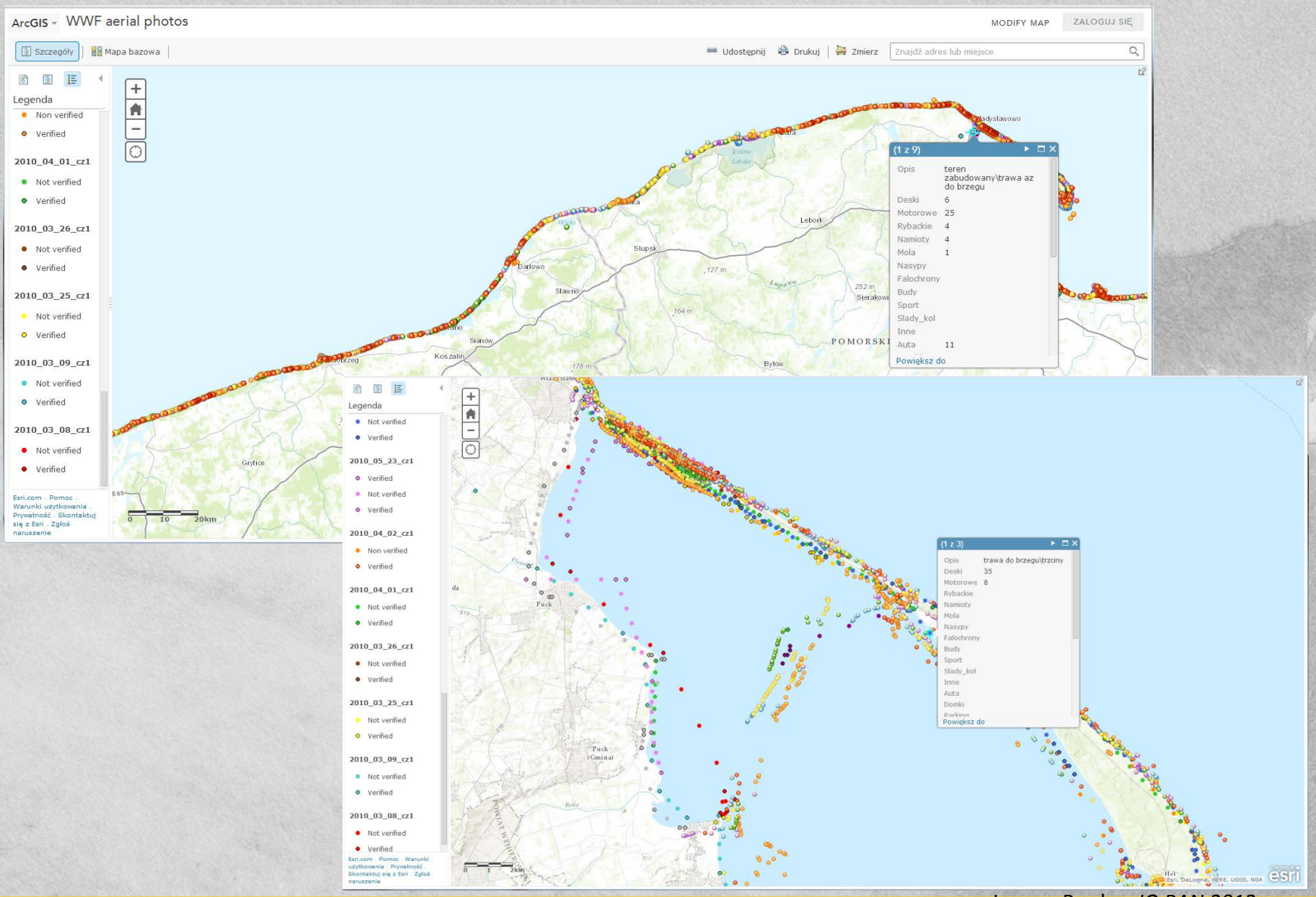


Foto: Jerzy Dabrowski

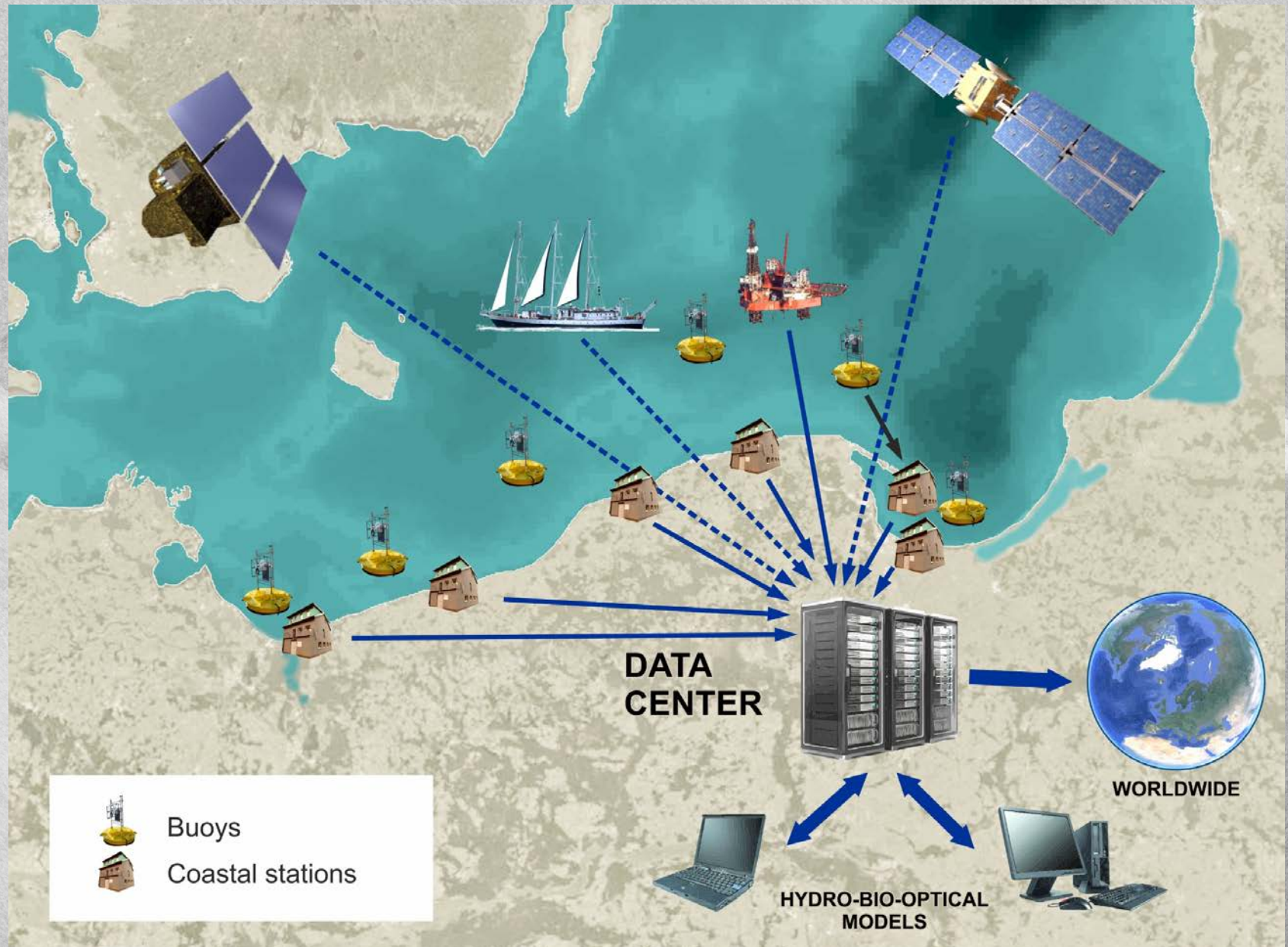
Acquisition of oceanographic data

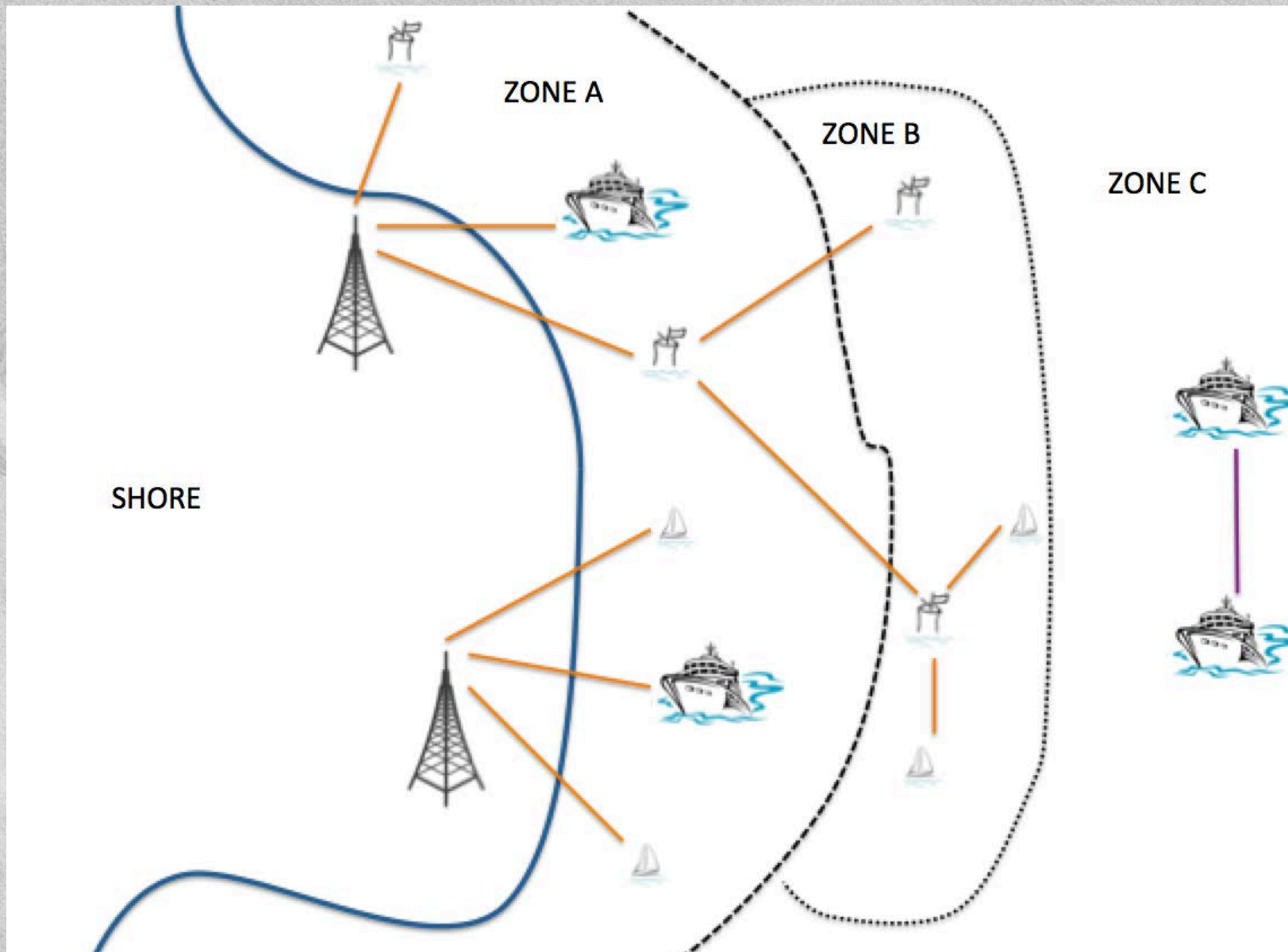


Acquisition of oceanographic data



Acquisition of oceanographic data



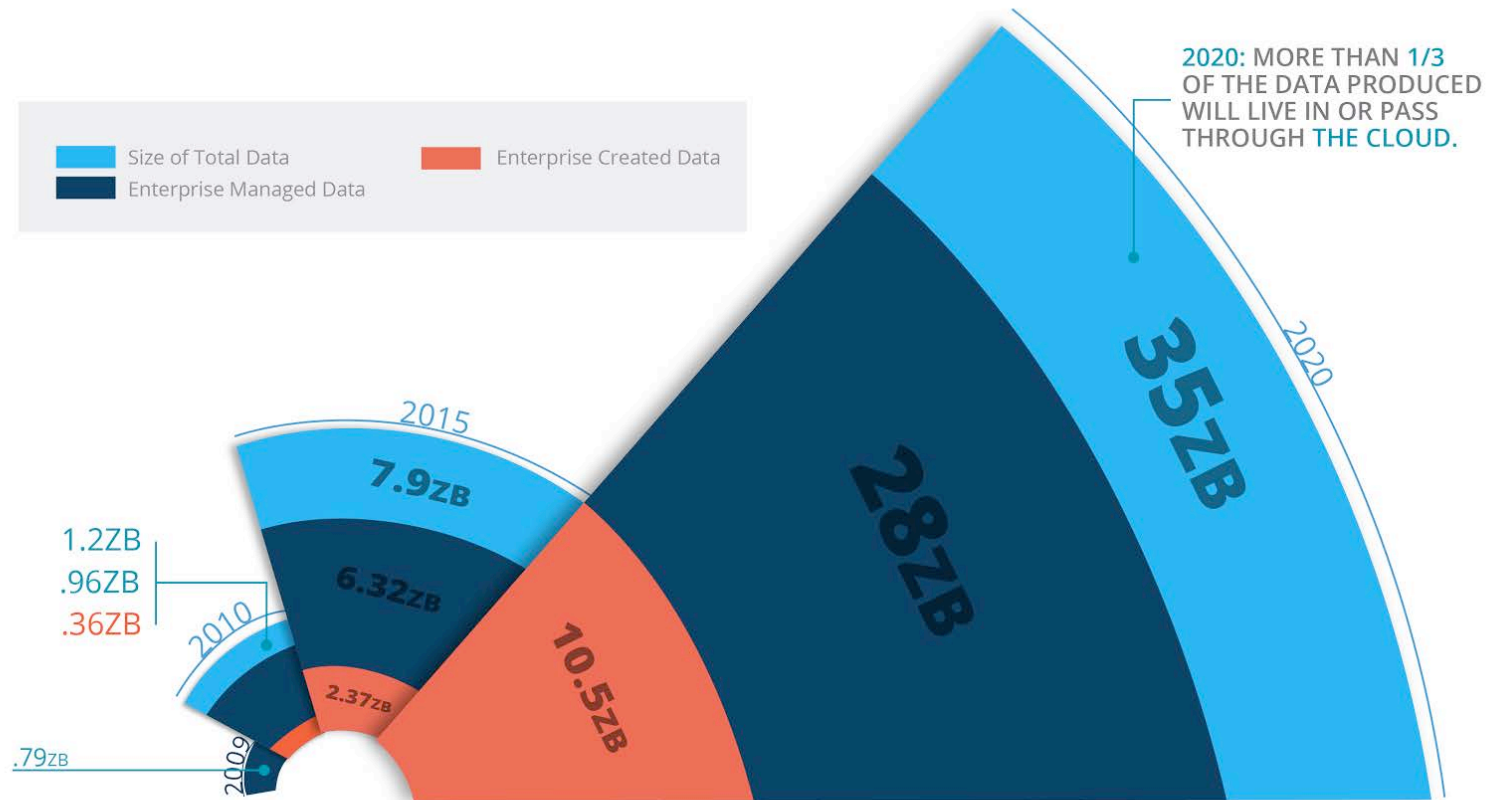


Sentinel-2 Data Volume



- 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)

Volume of data growth perspective



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

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

Jim Gray, "The Fourth Paradigm", Microsoft Research, Redmond

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 07 2007 10:10:10
* 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 [decibars]
# name 1 = t0: temperature [deg C]
# name 2 = sal00: salinity, PSS-78
# name 3 = sigma-t00: density, sigma-t
# 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 10:10:10
# derive_date = Mar 20 2007 09:10:10
# derive_in = 1RS000.CNV GUILDLIN
# strip_date = Mar 20 2007 09:10:10
# strip_in = 1RS000.CNV
# file_type = ascii
*END*

```

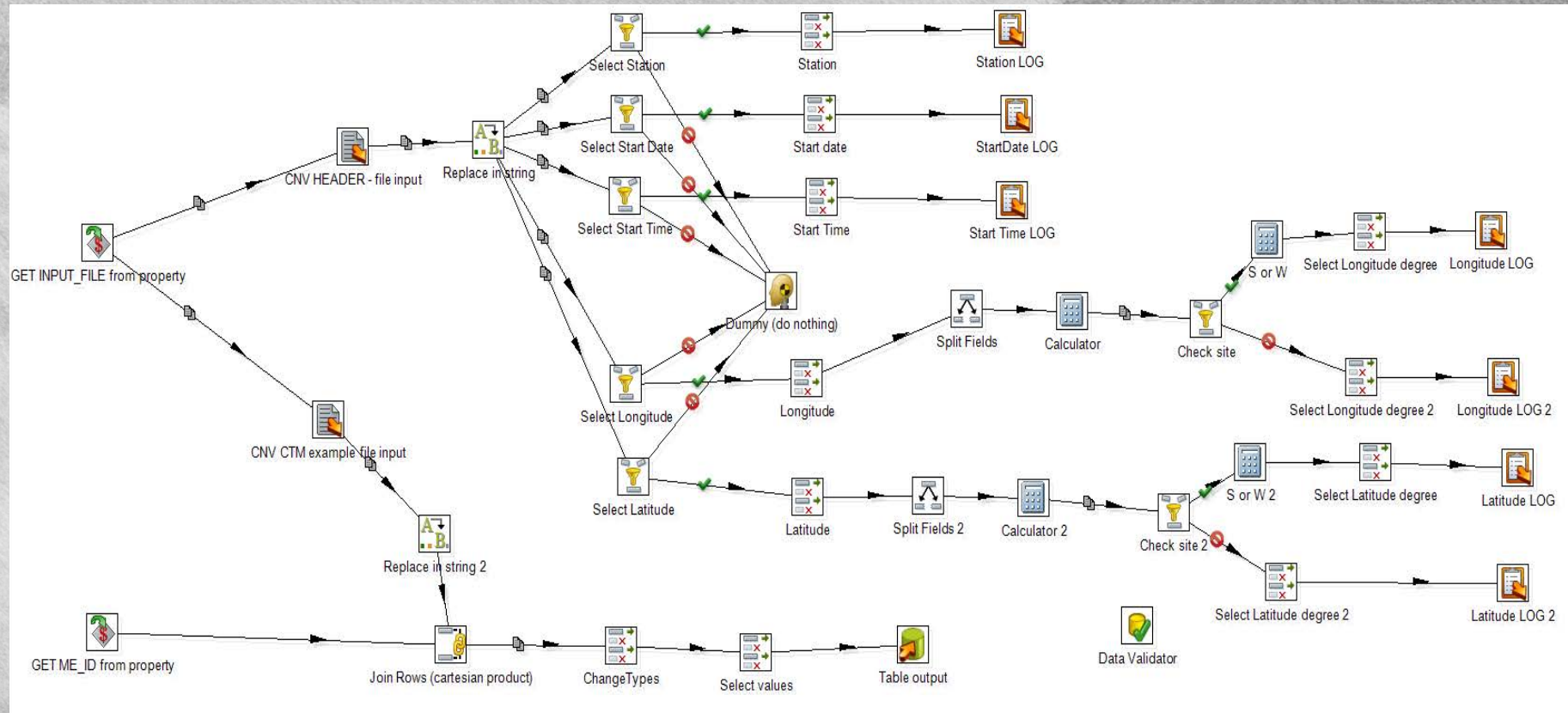
```

<?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/>
    </info>
    <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/>
        <partitionschemas/>
        <slaveservers/>
        <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>
      </log>
    <notepads>
      <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>
      </hop>
    </order>
  </transformation>

```

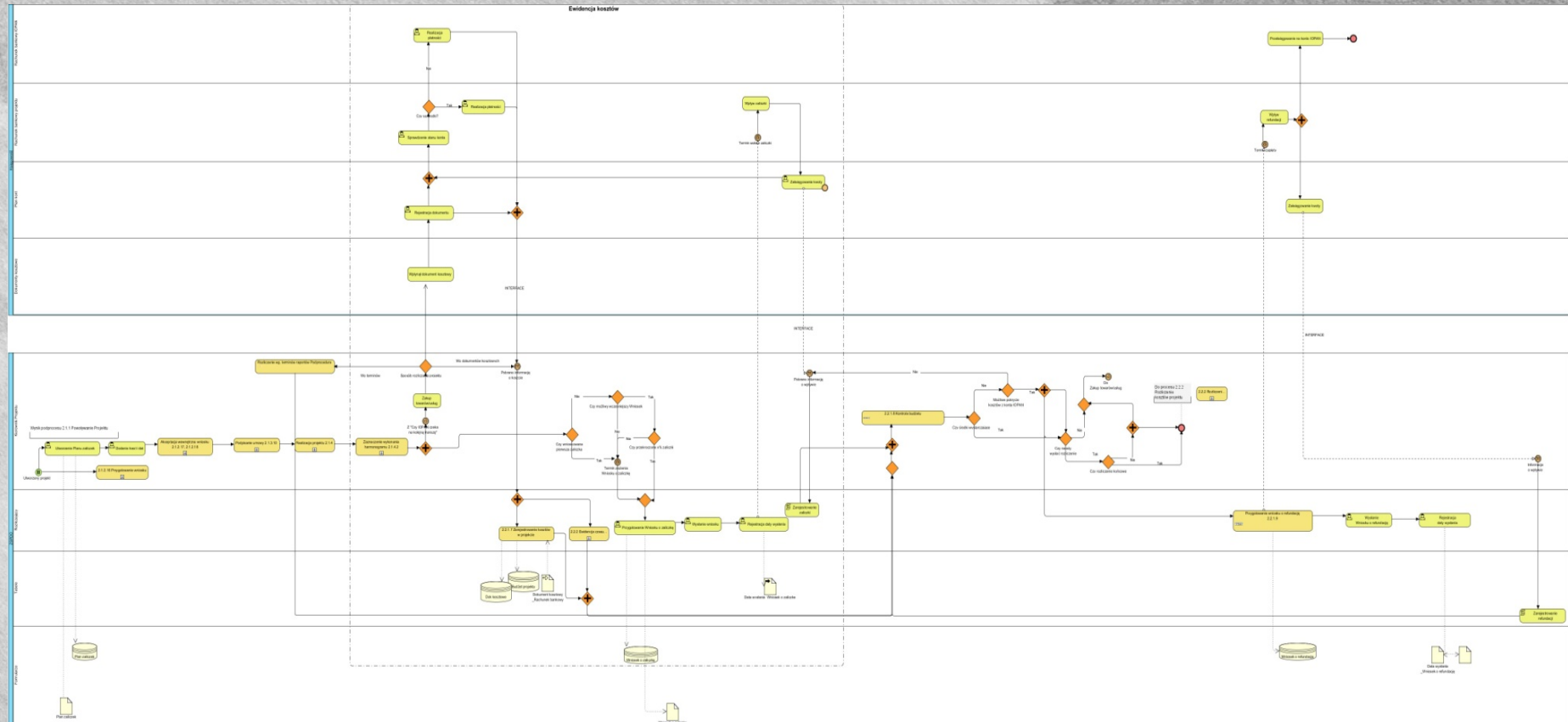


Analytics



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/ pCO ₂	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 data life 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)

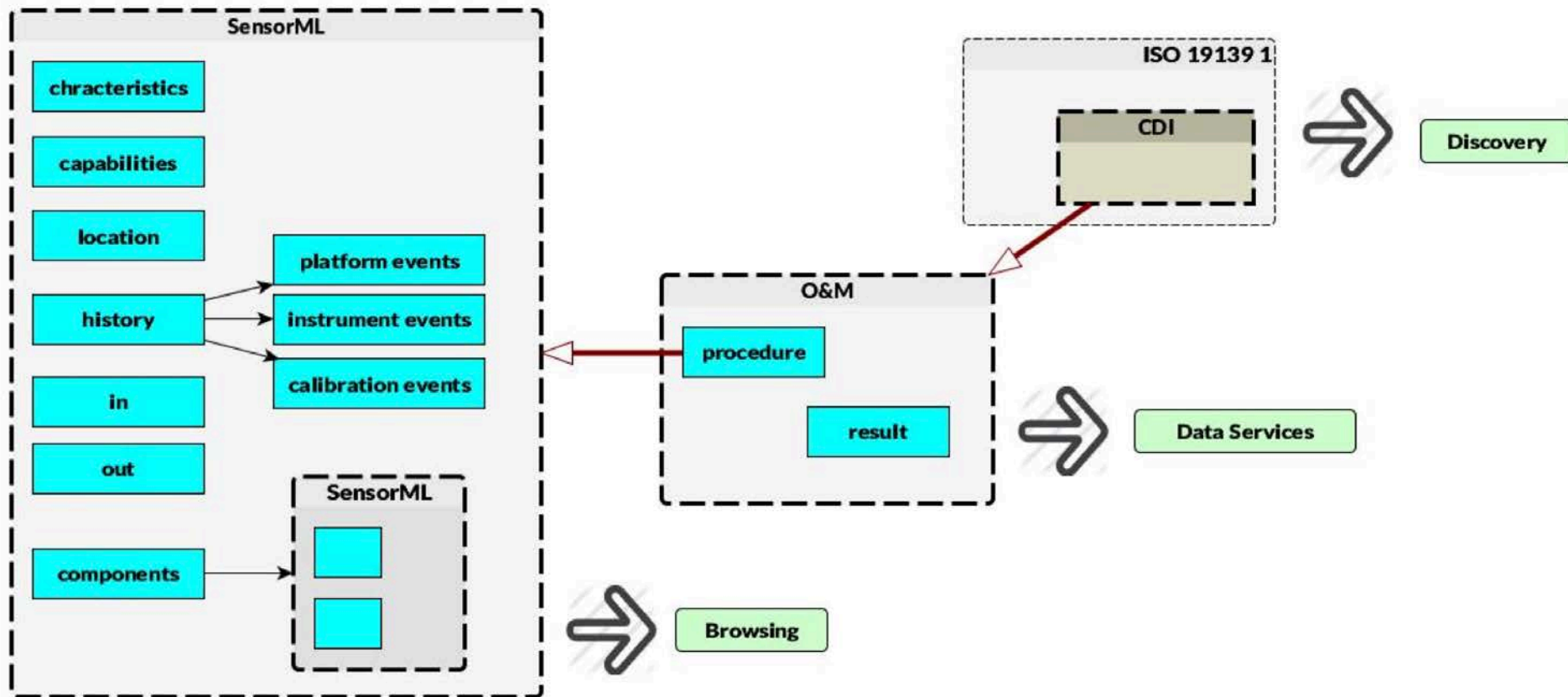
Systems

Transfer protocol (FTP, HTTP) and services (WMS, WFS, WCS, CS-W, SOS, WSDL, etc.)

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



PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT

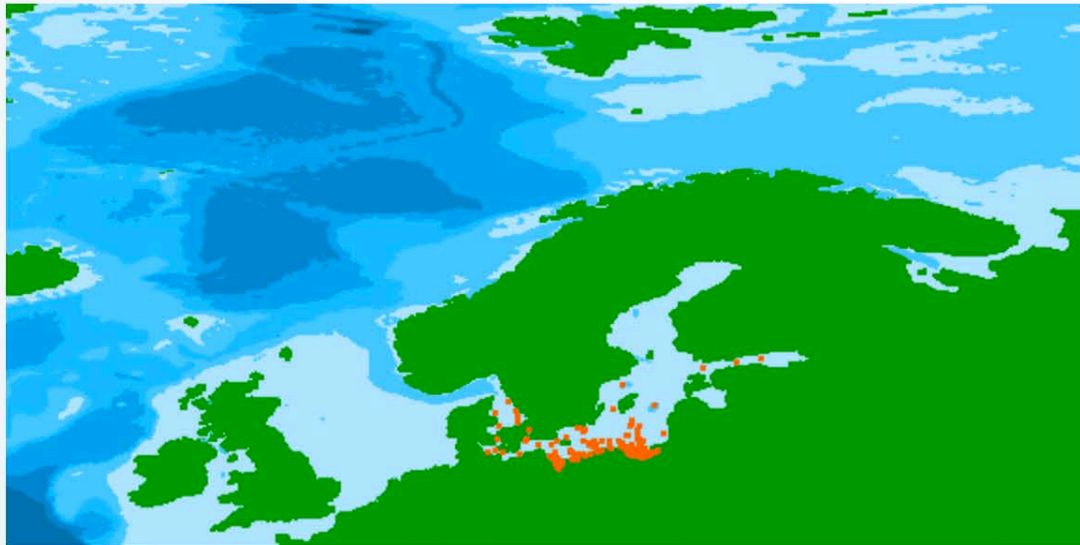
SEADATANET COMMON DATA INDEX (CDI) V3

TOOLS



ENLARGE HELP
POSITION INDEX

Datasets 0
BASKET RESET



LAYER CONTROL

- ☐ CDI entry Points
- ☐ CDI entry Tracks
- ☐ CDI entry Areas

- ☐ Grid Lines
- ☐ Regional sea
- ☐ Regional sea

Labels

- ☒ Display all selected records
- ☐ Only selected records in results list

LISTING RESULTS

☒ 20 ☐ 100 ☐ 1000 records

ADD TO BASKET

[TIMESERIES ON](#)


[SUMMARY](#)

[ZOOM TO SELECTED](#)

[EXPORT RESULT](#)

[STORE QUERY](#)

| [Refine query](#) | [New query](#) | Found 5563 | Show (1-20) | Previous | [Next 20](#)

<input type="checkbox"/> #	Data set name	DC country	Start date	Disciplines - Topics	Instrument / gear type	Show
<input type="checkbox"/>	Polish_Monitoring_zooplankton_data	Poland	20130626	Biological oceanography > Biota abundance, biomass and diversity > Other biological measurements	plankton nets	

Pan-European infrastructure for ocean & marine data management



The screenshot shows the SeaDataNet website interface. At the top, there is a header with the SeaDataNet logo on the left, the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT" in the center, and navigation links for "PARTNERS" and "USERS" on the right. Below the header is a main navigation bar with links: "ABOUT US", "METADATA", "DATA ACCESS", "STANDARDS", "SOFTWARE", "PRODUCTS", "EVENTS", and "PUBLICATIONS". The main content area features a large blue banner with a background image of a ship's deck and a rocky coastline. On the left side of the banner, the word "SERVICES" is displayed with a vertical list of three blue arrows. The central part of the banner contains the heading "ACCESS METADATA CATALOGUES" in large white letters, followed by a paragraph: "Access or Contribute to the SeaDataNet metadata catalogues of Marine organisations (EDMO), datasets (EDMED), projects (EDMERP), observing systems (EDIOS), Research cruises (CSR) and data description (CDI)". To the right of this text is a grid of seven blue buttons with white icons and text: "SEARCH DATA" (magnifying glass icon), "BROWSE DATA" (list icon), "DOWNLOAD SOFTWARE" (download icon), "LOOK-UP VOCABULARIES" (book icon), "ACCESS PRODUCTS" (cloud icon), "ACCESS METADATA CATALOGUES" (document icon), and "HOW TO CONTRIBUTE?" (upload icon). The grid is flanked by large blue arrows pointing left and right.

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT

PARTNERS   

USERS    

ABOUT US **METADATA** **DATA ACCESS** **STANDARDS** **SOFTWARE** **PRODUCTS** **EVENTS** **PUBLICATIONS**

SERVICES

ACCESS METADATA CATALOGUES

Access or Contribute to the SeaDataNet metadata catalogues of Marine organisations (EDMO), datasets (EDMED), projects (EDMERP), observing systems (EDIOS), Research cruises (CSR) and data description (CDI)

SEARCH DATA **BROWSE DATA** **DOWNLOAD SOFTWARE**

LOOK-UP VOCABULARIES **ACCESS PRODUCTS** **ACCESS METADATA CATALOGUES**

HOW TO CONTRIBUTE?

<https://www.seadatanet.org/>

A wide-angle photograph of a massive glacier wall, likely the Perito Moreno Glacier, meeting a body of water. The glacier is a deep blue-white color, showing vertical crevasses and a jagged edge. In the foreground, there are large, rounded icebergs. The water is a dark, calm blue. In the background, dark, snow-dusted mountains rise under a heavy, grey sky with some blue patches. The text "THANK YOU !" is overlaid in white, bold, sans-serif font on the right side of the image.

THANK YOU !