Management, analysis, and visualisation of information in Scientific Infrastructures Architectural Patterns for web-based information processing

Rainer Haener, Martin Hammitzsch, Joachim Waechter

Management, analysis, and visualisation of information in Scientific Infrastructures

Architectural Patterns
for web-based information processing

Rainer Haener, Martin Hammitzsch, and Joachim Waechter
Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany

GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

Overview and briefly introducing the European Plate Observing System (EPOS)

- Introduction and Challenges
- Establishing Collaboration & Interoperability
- Architectural Patterns & Design Principles
  - Semantic Enablement
  - Encoding
  - Integration & Communication
- Automation of Integration & Infrastructures
- Conclusion

- aims at creating a **pan-European infrastructure** for solid Earth-Science to support a safe and sustainable society.
- enables innovative **multidisciplinary research** for a better understanding of the Earth's physical and chemical processes that control earthquakes, volcanic eruptions, ground instability, tsunami, and all those processes driving tectonics & Earth's surface dynamics.
- allows scientists to make a step change in developing **new** geo-hazards and geo-resources **concepts** and Earth-Science applications to help address key societal challenges (through the **integration of research infrastructures and data**)

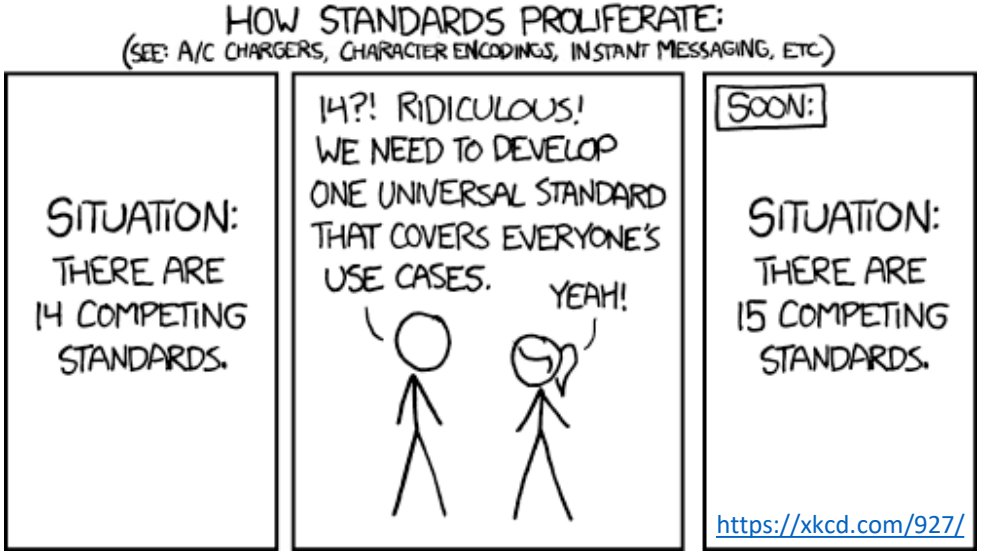
GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

Challenge

Current developments in Scientific Infrastructures reveal …

- a very high variety regarding amount and kind of data as well as
- a very high degree of heterogeneity concerning their interfaces provided by the data suppliers



… and also show a still present lack of standardisation concerning the encoding of data, metadata, and semantics due to a proliferation of domain & even application specific formats:

1. Well established: GML, Observations & Measurements, SensorML
2. Application Profiles: **BoreholeML**, GeoSciML, **GroundwaterML**, WaterML
3. Industrial: WITSML

A Lithology-Log encoded via GroundwaterML

```
<om:result xmlns:ns="http://www.opengis.net/gwml-well/2.2" xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:xlink="http://www.w3.org/1999/xlink"
xsi:type="ns:GW_GeologyLogCoveragePropertyType">
    <ns:GW_GeologyLogCoverage gml:id="gwglc_pv1596478">
        <ns:element>
            <ns:LogValue>
                <ns:fromDepth >…</ns:fromDepth >
                <ns:toDepth>…</ns:toDepth>
                <ns:value>
                    <swe:DataRecord>
                        <swe:field name="earthMaterial">
                            <swe:Text definition="http://www.opengis.net/def/gwml/2.0/observedProperty/earthMaterial">
                                <swe:label>earthMaterial</swe:label>
                                <swe:description>Lithology of the observed level</swe:description>
                                <swe:value>clayey silt</swe:value>
                            </swe:Text>
                        </swe:field>
                    </swe:DataRecord>
```

GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

A Lithology-Log encoded via GroundwaterML cont'd

```
<om:result xmlns:ns="http://www.opengis.net/gwml-well/2.2" xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:xlink="http://www.w3.org/1999/xlink"
xsi:type="ns:GW_GeologyLogCoveragePropertyType">
    <ns:GW_GeologyLogCoverage gml:id="gwglc_pv1596478">
        <ns:element>
            <ns:LogValue>
                <ns:fromDepth>
                    <swe:Quantity definition="fromDepth">
                        <swe:label>fromDepth</swe:label>
                        <swe:uom code="m" xlink:title="metre" xlink:href="#Meter"/>
                        <swe:value>0.6</swe:value>
                    </swe:Quantity>
                </ns:fromDepth>
                <ns:toDepth> ... </ns:toDepth>
                <ns:value> ... </ns:value>
            </ns:LogValue>
        </ns:element>
    </ns:GW_GeologyLogCoverage>
</om:result>
```

GFZ Helmholtz Centre Potsdam

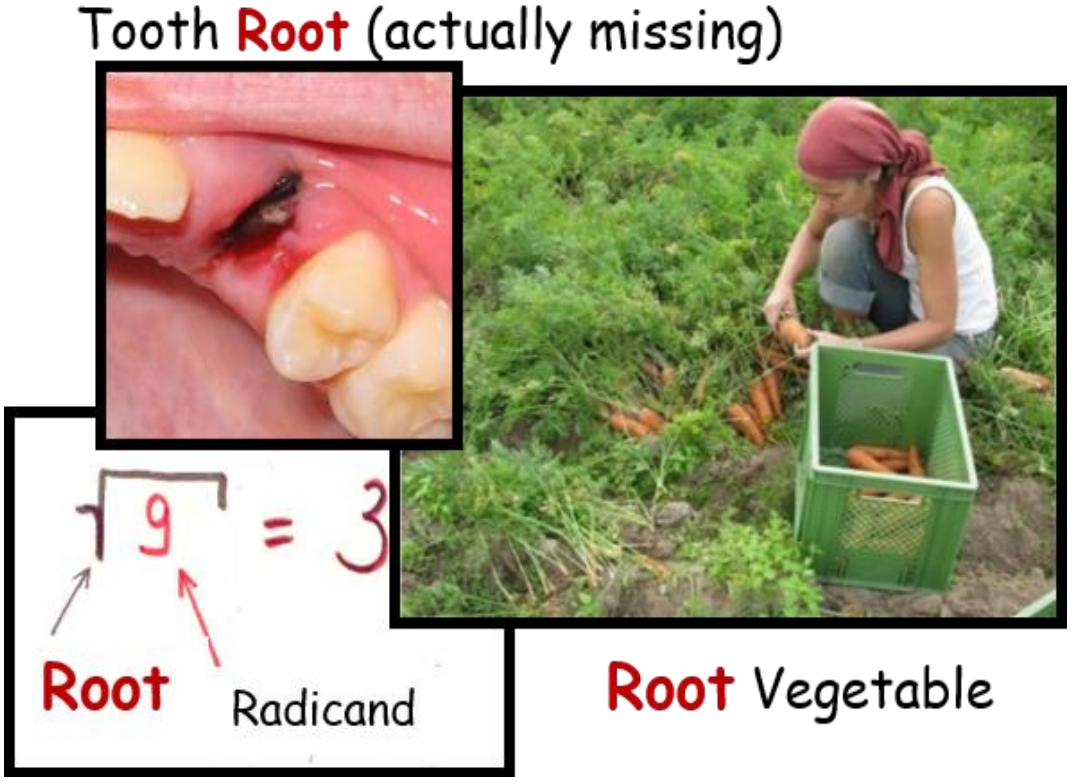
HELMHOLTZ ASSOCIATION

A Lithology-Log encoded via BoreholeML

```
<bml:layer xmlns:bml="http://www.infogeo.de/boreholeml/3.0" xmlns:gmd="http://www.isotc211.org/2005/gmd" >
    <bml:Interval>
        <bml:from uom="m">0.6</bml:from>
        <bml:to uom="m">2.1</bml:to>
        <bml:lithology>
            <bml:Lithology>
                <bml:rockName codeSpace="http://bml/codelists/RockNames">clayey silt</bml:rockName>
            </bml:Lithology>
        </bml:lithology>
        <bml:stratigraphy>
            <bml:Stratigraphy>
                <bml:chronoStratigraphy codeSpace="http://bml/codelists/ChronoStratigraphy">Q</bml:chronoStratigraphy>
                <bml:lithoStratigraphy>
                    <gmd:LocalisedCharacterString>Quarternary</gmd:LocalisedCharacterString>
                </bml:lithoStratigraphy>
            </bml:Stratigraphy>
        </bml:stratigraphy>
    </bml:Interval>
</bml:layer>
```

Semantic Enablement, provided by a Registry establishing …

1. Common understanding of the meaning of information exchanged
2. Discovery of Resources
3. Inferable knowledge (statements) about Resources
4. Domain Controlled Vocabularies (Dictionaries of Terms)
5. Further processing of information
6. Data Mining & Fusion (Correlation)



… across domain specific contexts

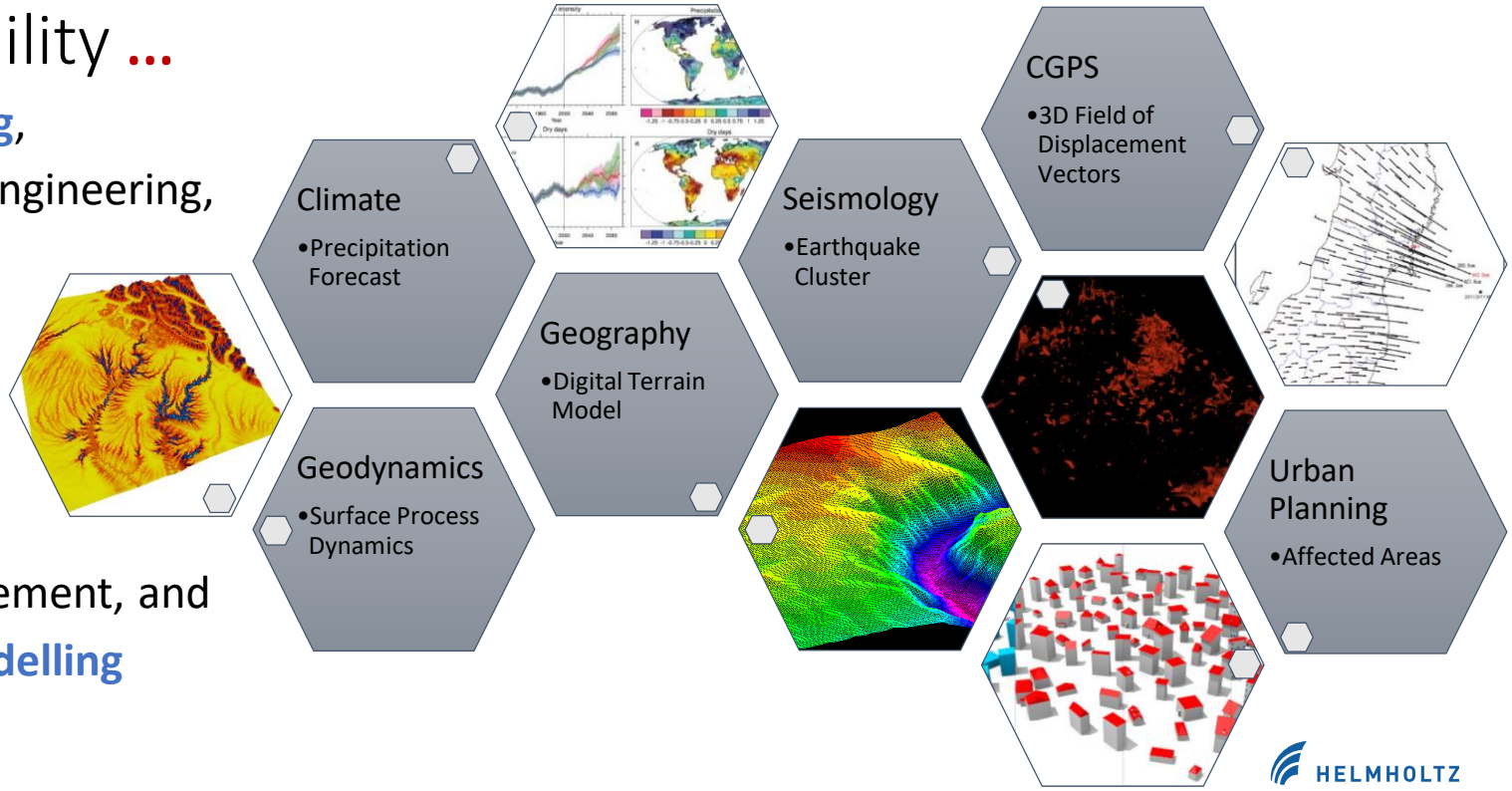
Collaboration & Interoperability

What does that mean?

… on the example of Risk Management

Collaboration & Interoperability …

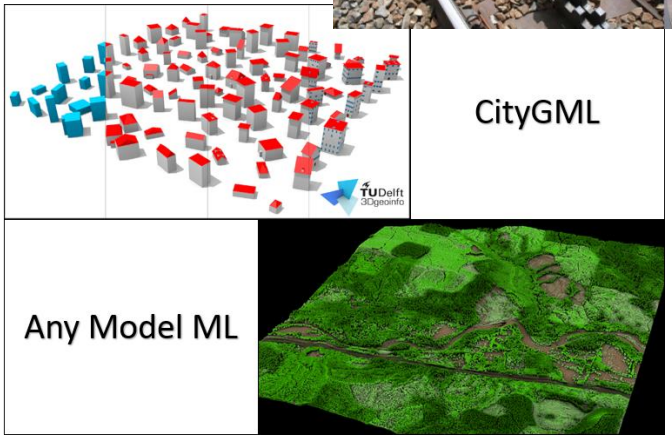
- **Urban Planning**,
- Geotechnical Engineering,
- Climatology,
- Meteorology,
- Geology,
- Geography,
- Geophysics,
- Facility Management, and
- **Numerical Modelling**



Establishing Collaboration & Interoperability

Collaboration in this context requires …

- the seamless & easy exchange of information
- that information & resources can be found (discovered)
- that information can be further processed, and
- easily utilised & integrated into existing workflows

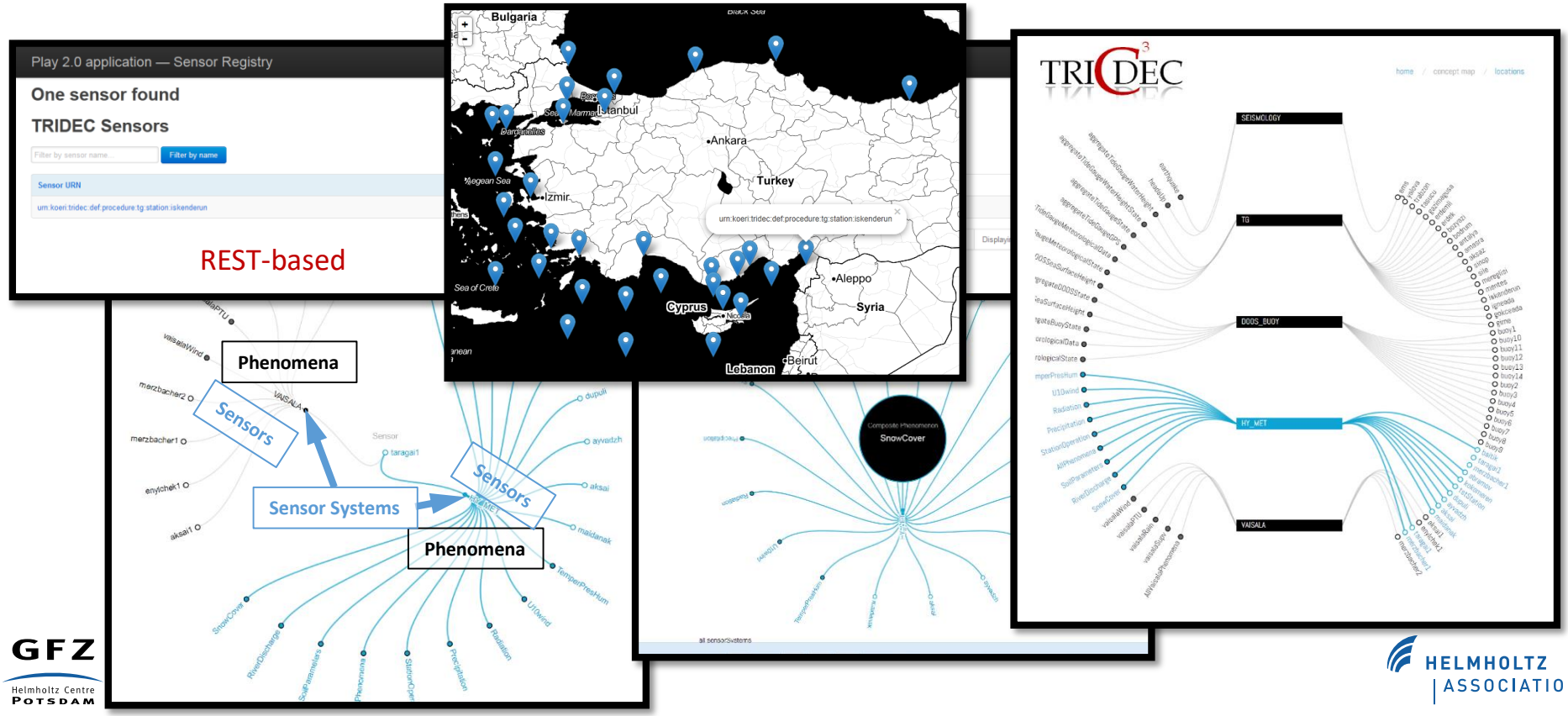


… not only across system,
but also across domain boundaries

Reference Architecture

Semantics

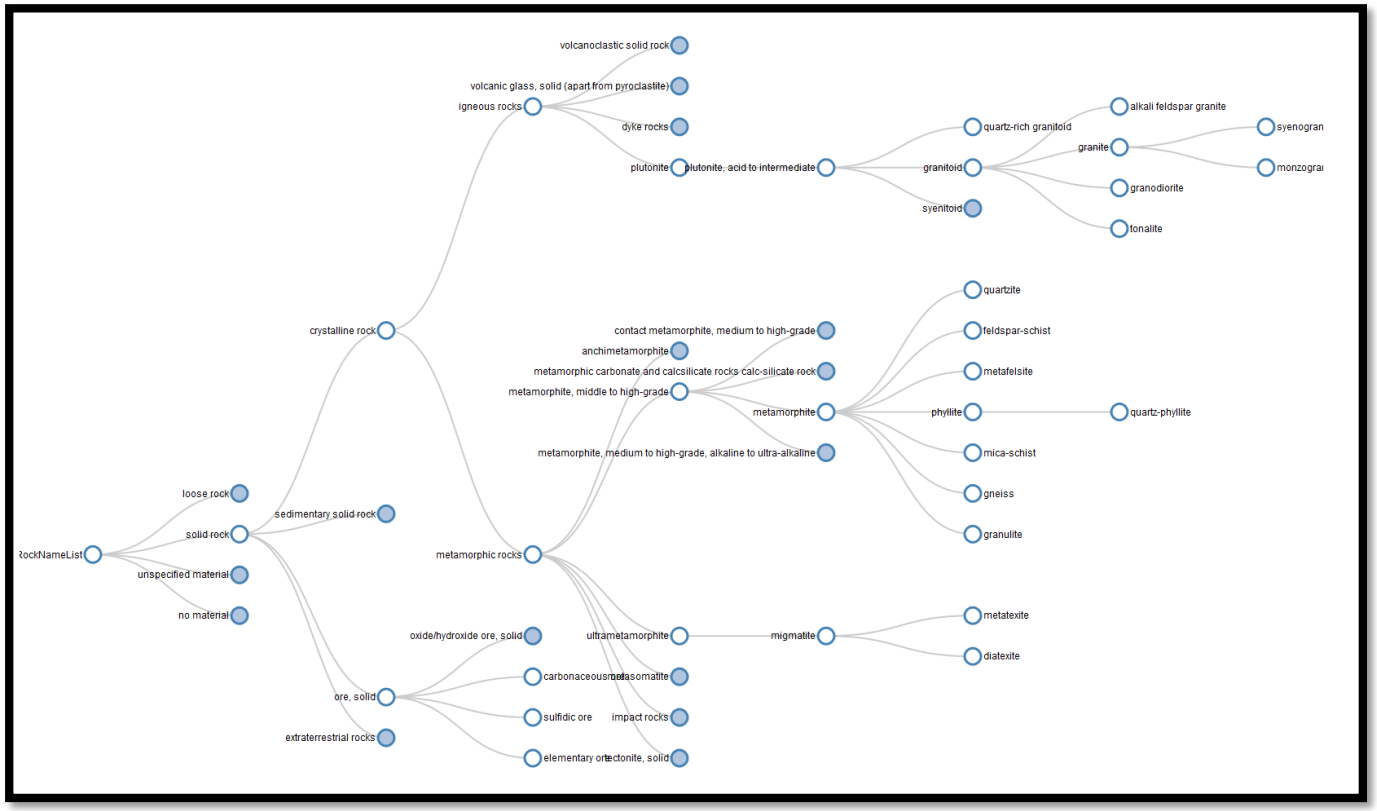
Registry: Discovery of - and Access to Resources



GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

Registry: Domain Controlled Vocabulary

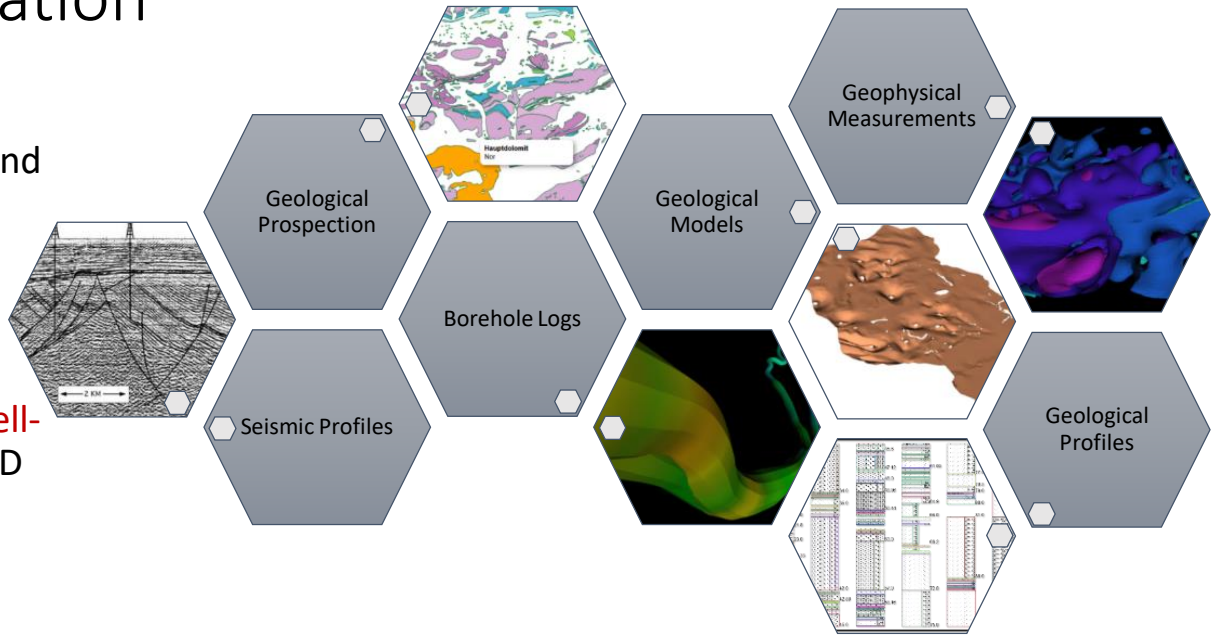


Reference Architecture

Encoding

Design Implications regarding the specification of encodings (DbC) …

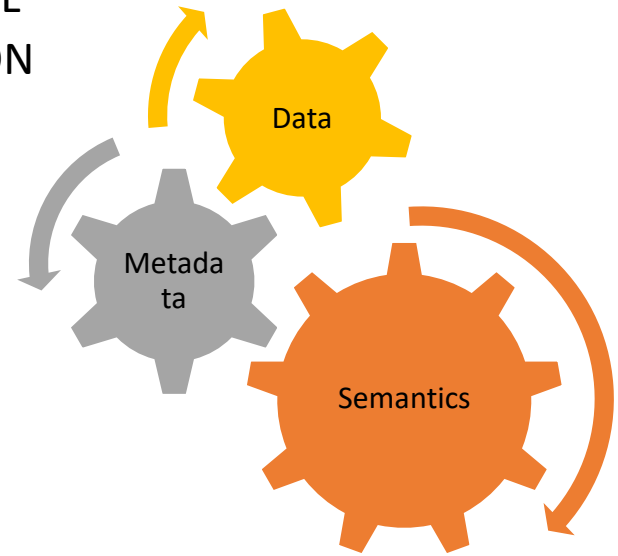
1. Strict separation of Data, Metadata and Semantics
2. Use of a small, reduced set of well-established & commonly accepted encodings
3. Delegate responsibility and re-use well-established standards like GML or X3D for e.g., geometries
4. Favour composition over inheritance
5. Characterise common elements (e.g. measure) via attributes unlike naming the elements themselves



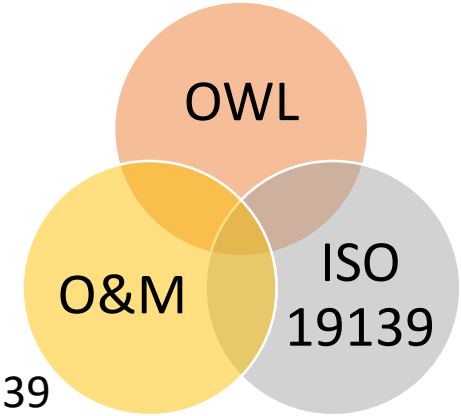
… on the example of Subsurface & Risk Management

Interoperability: Classification of Data & Encoding

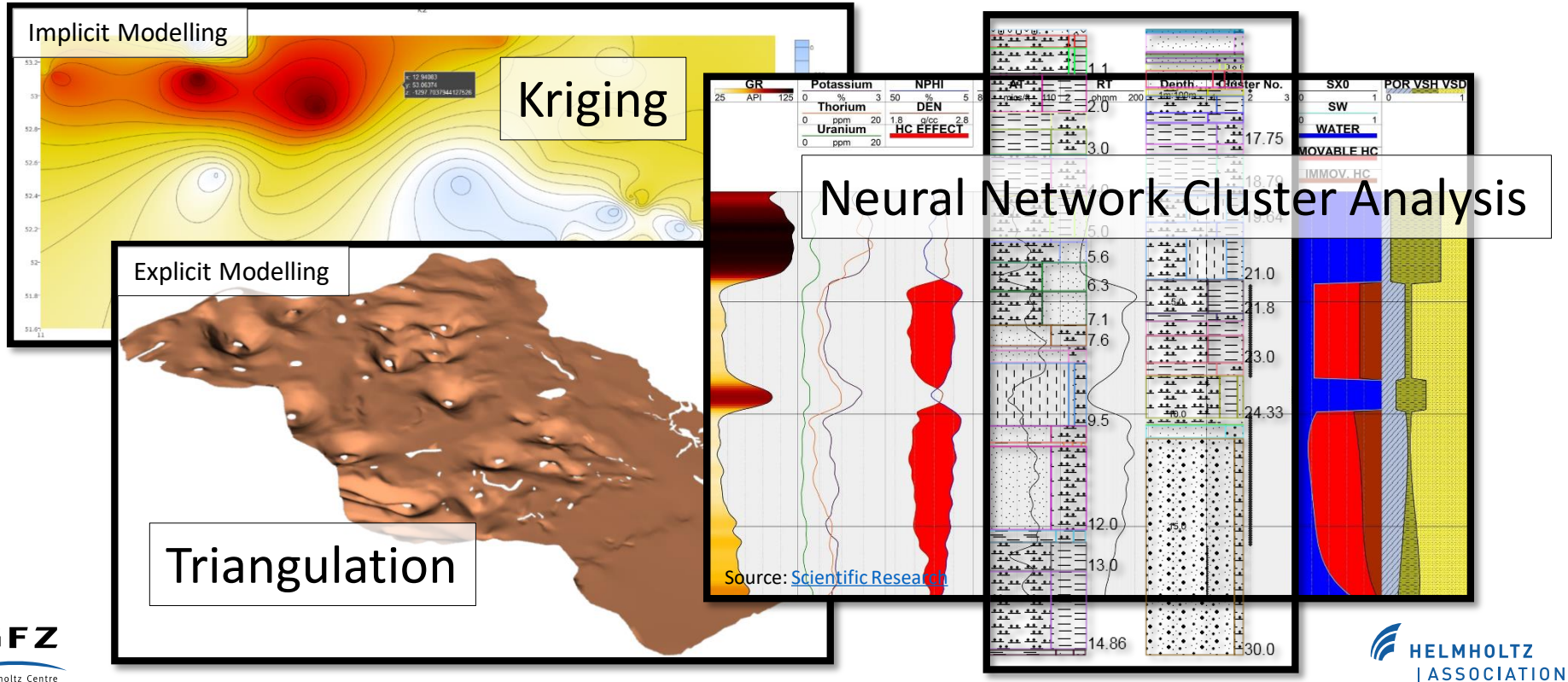
- Structure
  - Charts, Tables
  - XML
  - JSON



- Data
  - CSV
  - O&M, GML, …
  - X3D, SVG, …
- Metadata
  - ISO 19115/19139
  - SensorML
- Semantics
  - RDF/XML, OWL, …
  - ISO 19135



Correlation of Borehole Logs, Geological Models and Machine Learning



GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

A Landslide, observed (numerically modelled)

```
<?xml version="1.0" encoding="UTF-8"?>
<om:OM_Observation gml:id="landslide">
    <om:procedure          xlink:href="Modelling of Ground Instability"/>   <!– Process ("Sensor") Identification -->
    <om:observedProperty   xlink:href="Landslide"/>                        <!– the "Phenomenon" observed -->
    <om:featureOfInterest  xlink:href="Populated Mounteanous Area"/>       <!– Sampling Feature -->

    <om:result xlink:href="Link to Data">
        <gml:TriangulatedSurface gml:id="not of interest here">
            <gml:patches>
                <gml:Triangle>
                    <gml:exterior>
                        <gml:LinearRing>
                            <gml:posList>0 1 1 2 3 5 8 13 21 34 55 89 144</gml:posList>
                        </gml:LinearRing>
                    </gml:exterior>
                </gml:Triangle>
            </gml:patches>
        </gml:TriangulatedSurface>
    </om:result>
</om:OM_Observation>
```

GFZ Helmholtz Centre Potsdam

A Building, observed as well

```
<?xml version="1.0" encoding="UTF-8"?>
<om:OM_Observation gml:id="town">
    <om:procedure           xlink:href="3D Scanner"/>                    <!– Process ("Sensor") Identification -->
    <om:observedProperty    xlink:href="Buildings"/>                     <!– the "Phenomenon" observed -->
    <om:featureOfInterest   xlink:href="Town in area prone to Landslides"/>  <!– Sampling Feature -->

    <om:result xlink:href="Link to Data">
        <gml:Surface gml:id="might not be of interest in our context">
            <gml:patches>
                <gml:PolygonPatch>
                    <gml:exterior>
                        <gml:LinearRing>
                            <gml:posList>4 8 15 16 23 42</gml:posList>
                        </gml:LinearRing>
                    </gml:exterior>
                </gml:PolygonPatch>
            </gml:patches>
        </gml:Surface>
    </om:result>
</om:OM_Observation>
```

GFZ Helmholtz Centre Potsdam

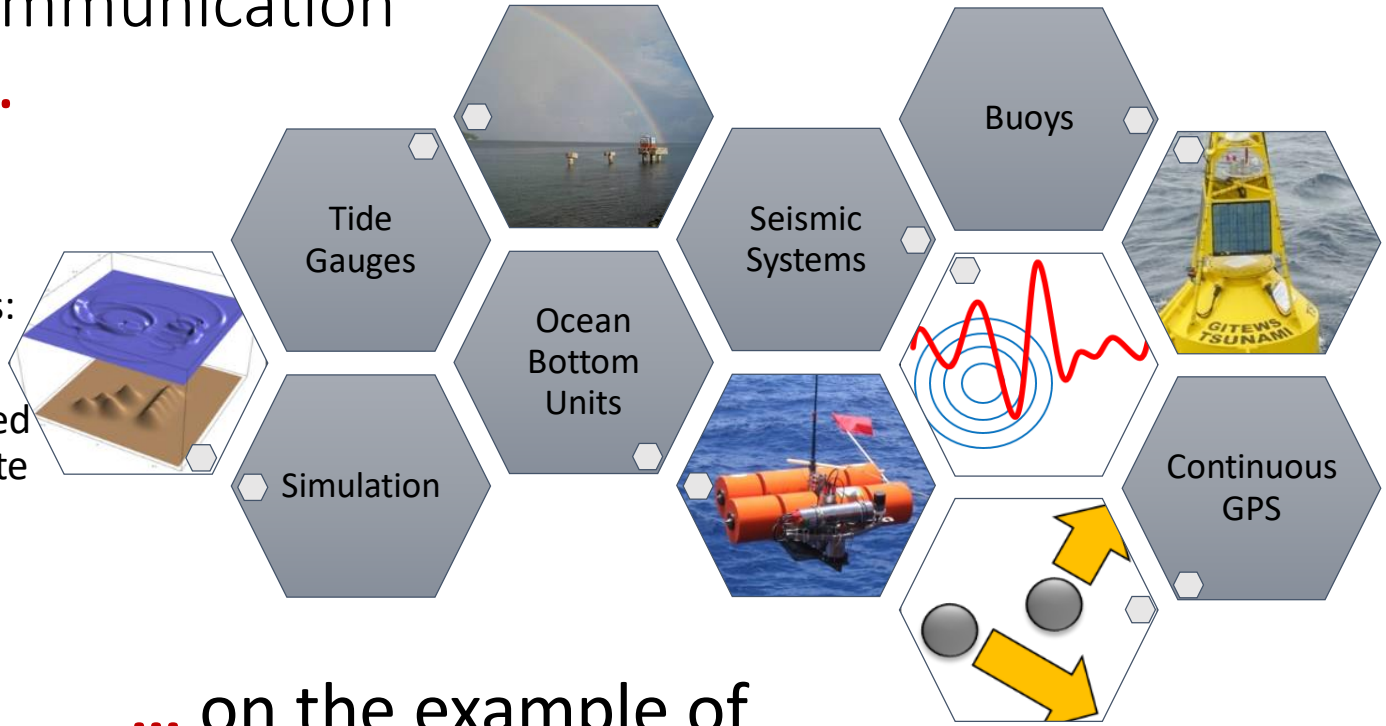
Reference Architecture

Integration & Communication

Design Principles regarding Integration & Communication Infrastructures …

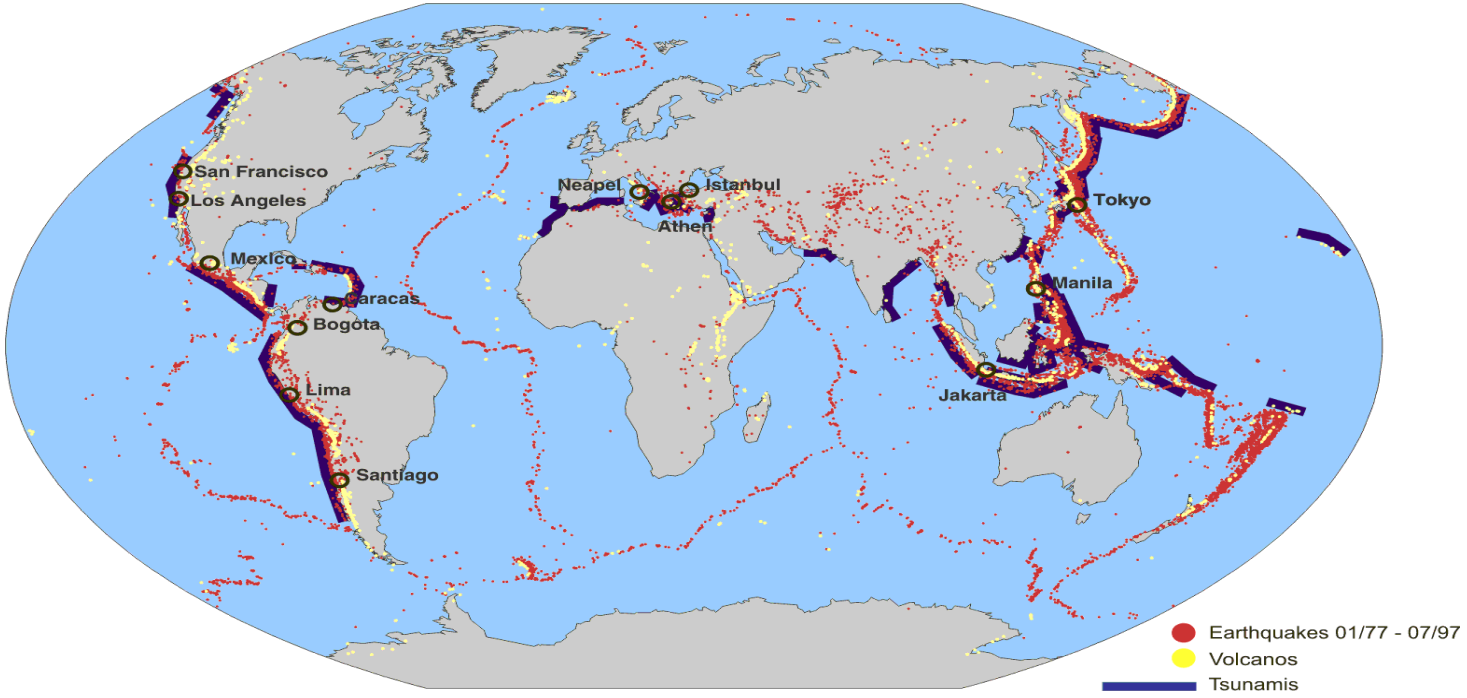
The patterns proposed, are based on the architectural paradigms for a functional integration of web resources:

1. Resource Oriented Architecture (ROA), based on Representational State Transfer (REST), and
2. Event-Driven Service-Oriented Architectures (SOA 2.0)



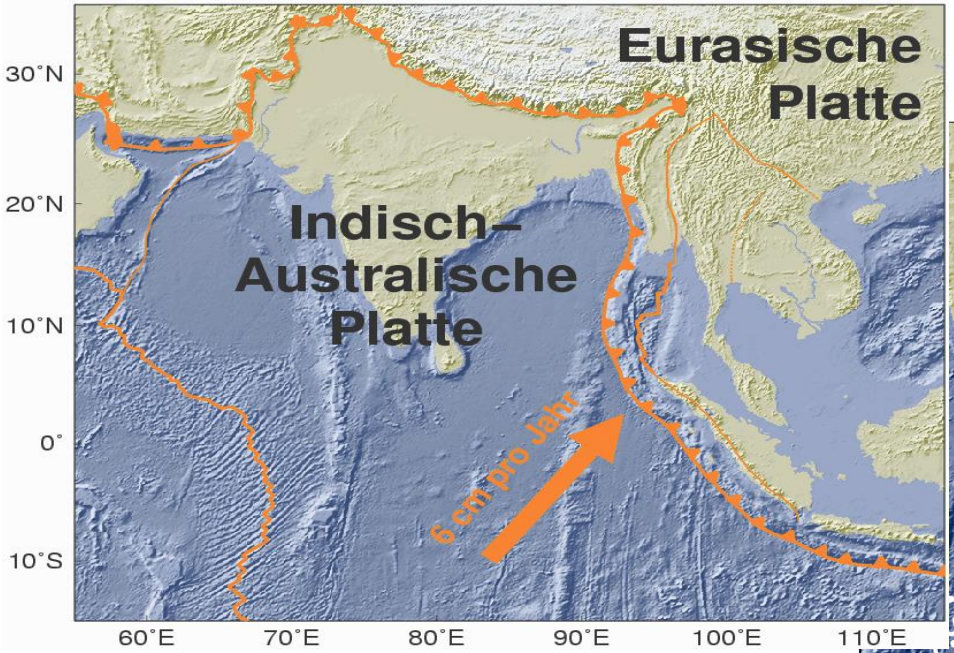
… on the example of Tsunami Early Warning

Tsunami Sources Submarine Earthquakes M > 7.0 (about 90 %)



But: only 20% of strong submarine quakes trigger Tsunamis!

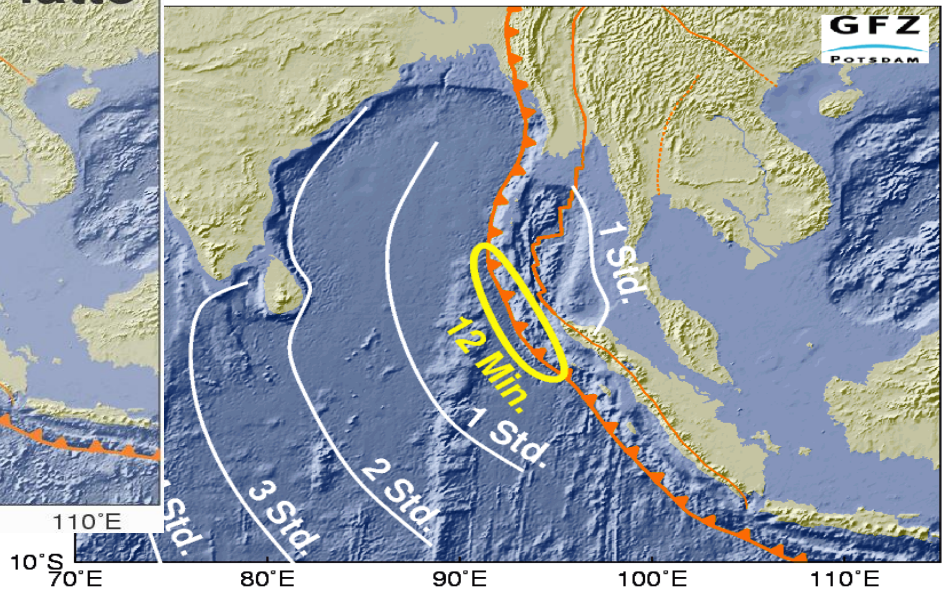
Challenge Indian Ocean



tectonic situation

extremely prone to earthquakes

To make matters worse, extremely exposed

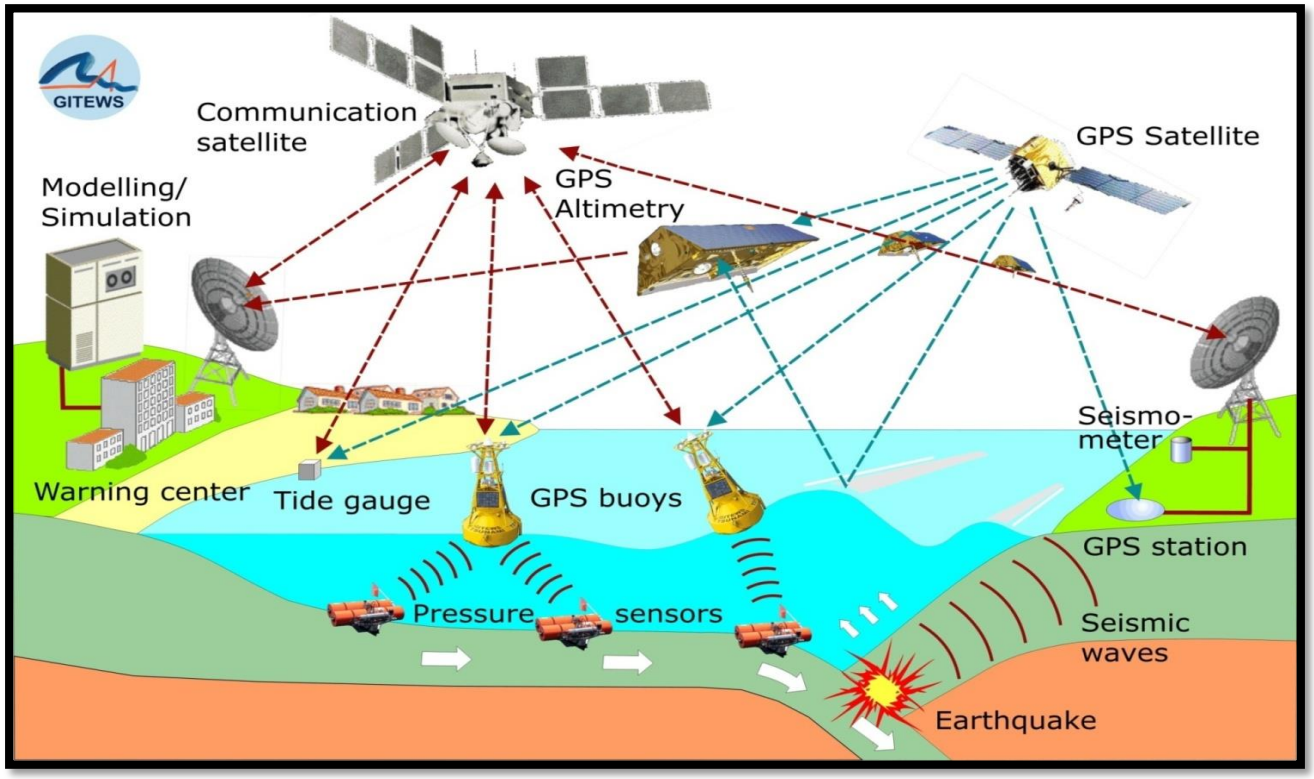


travel times

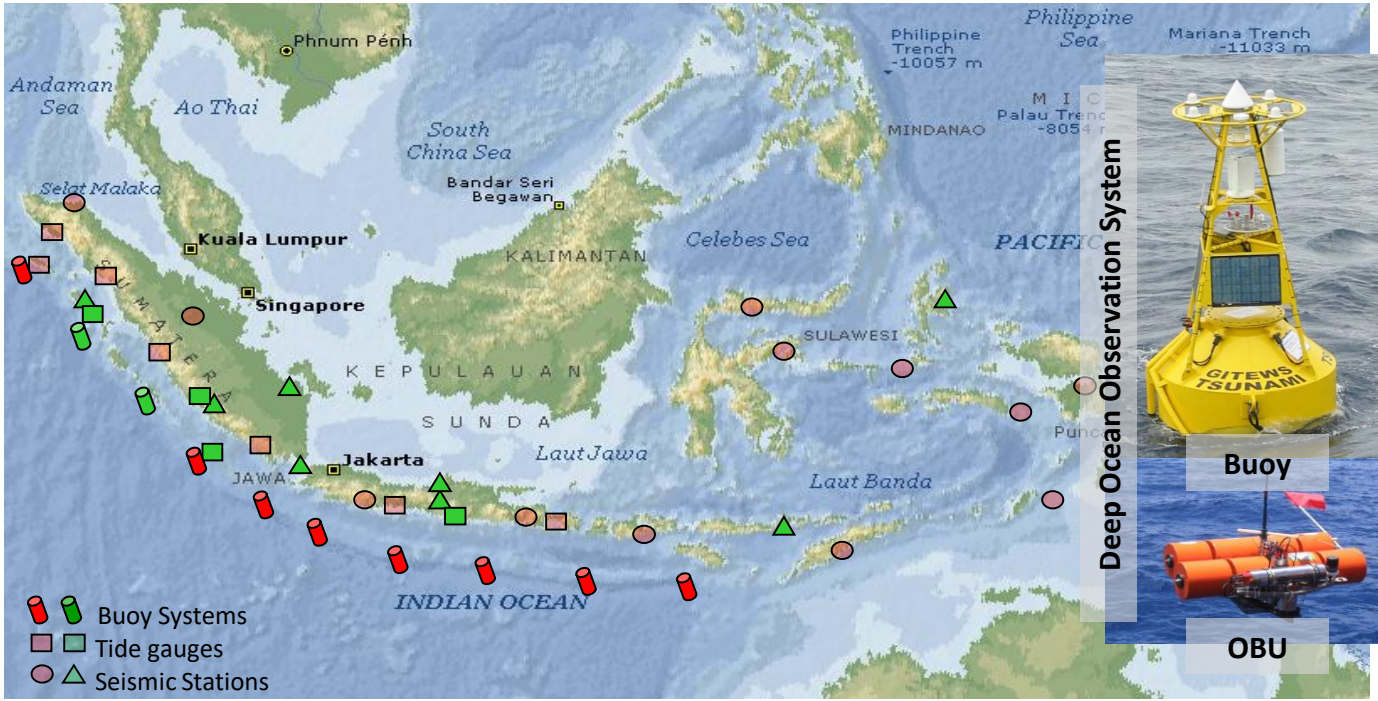
… and the idea is …

a comprehensive view of all phenomena that cause a tsunami and are related to it

Sensor Concept of the German Indonesian Tsunami Early Warning System (GITEWS)

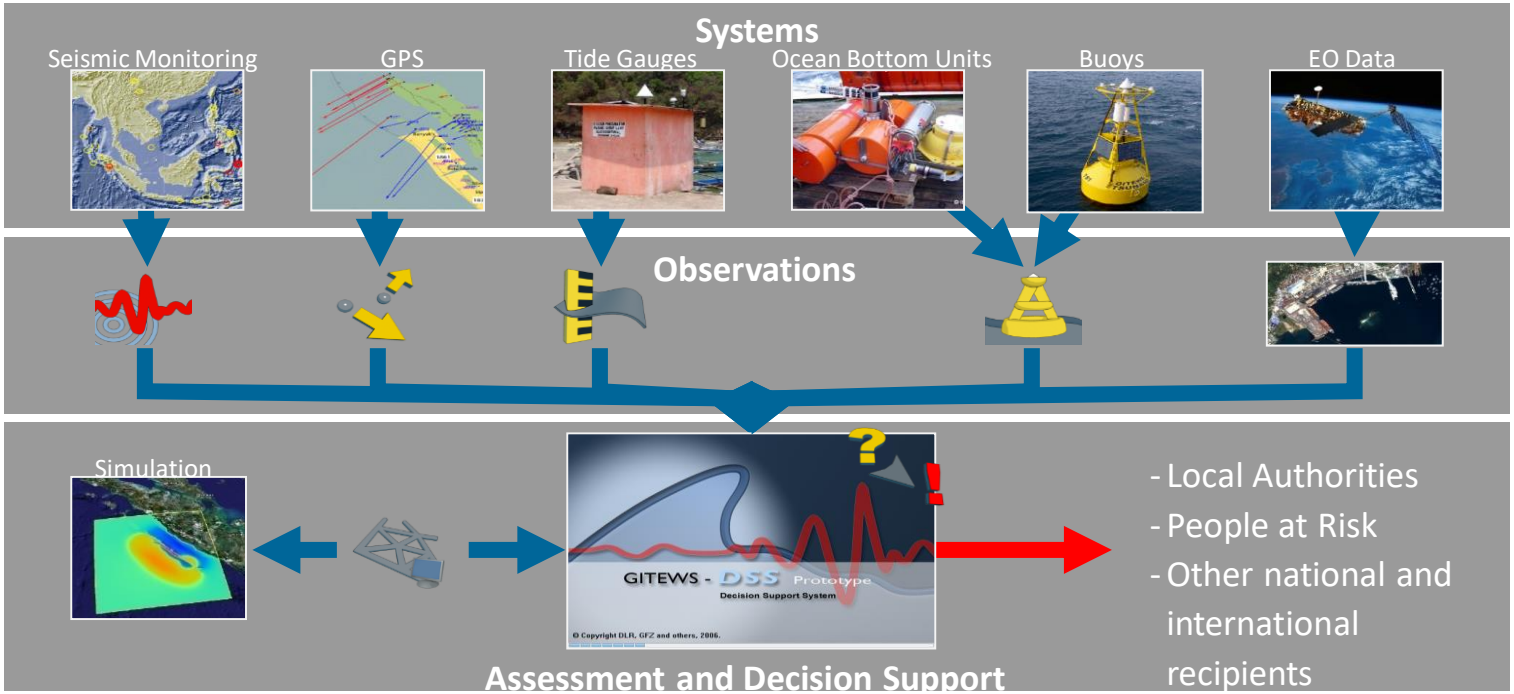


GITEWS Sensors Deep Ocean Observation System (DOOS)



GITEWS Concept of Early Warning

Decision Support System (DSS)



Source: Deutsches Zentrum für Luft und Raumfahrt e.V. (DLR)

Automation of Integration & Infrastructures

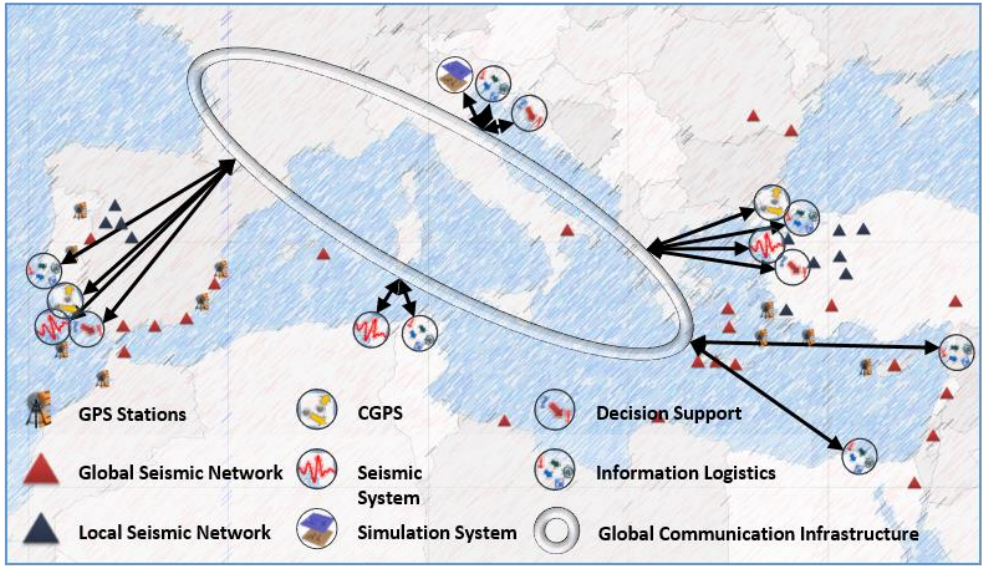
A Blueprint

GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

Automation of Integration & Infrastructures …

- Share pool of physical and/or virtual resources rather than deploying them locally
  - Provide pre-processed results (Resources), e.g., Simulation or Neural Networks
  - Establish Scientific Computing in Research Infrastructures
- Enable collaboration (Workflows) across domain and system boundaries
  - Provide & utilise standardised processing algorithms (Functionality, Services)
  - Establish Virtual Research Environments



Tsunami Early Warning based on SOA 2.0

… on the example of the TRIDEC Cloud

ROA and SOA 2.0 as workflow engine

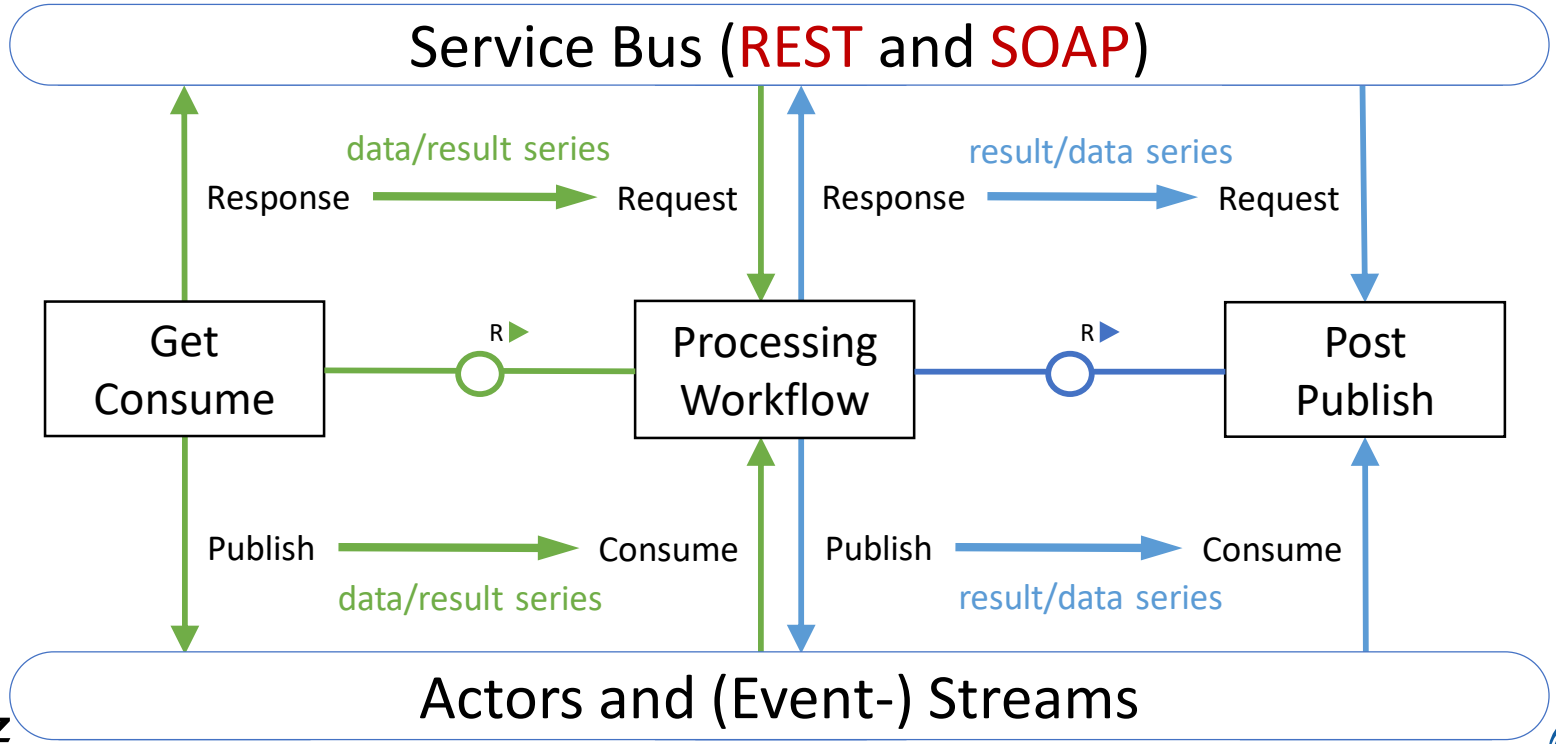
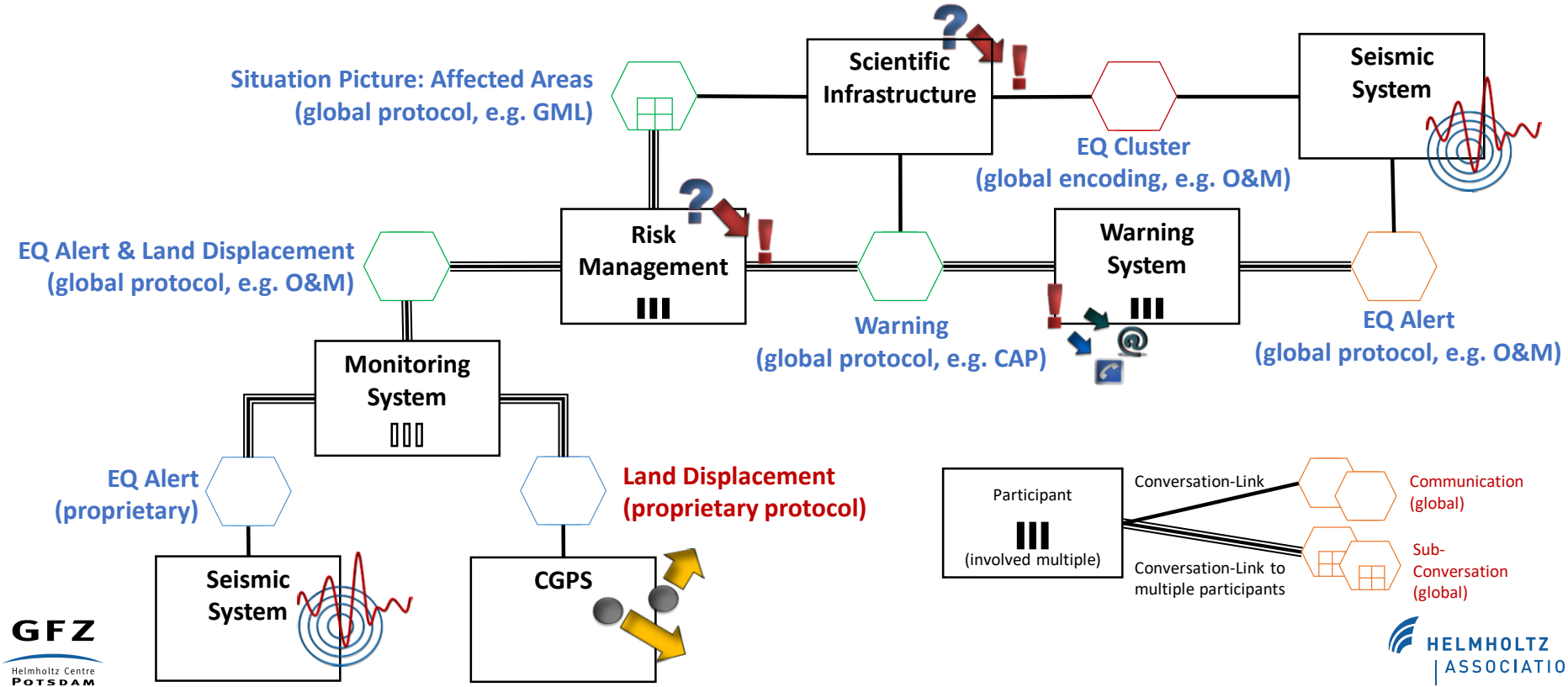


Diagram based on the Fundamental Modelling Concept (FMC) Adapted from Presentation EGU 2018, ESSI2.6

Establishing Interoperability on the basis of conversations (Design by Contract)



GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

Workflow Automation (Choreography), REST & SOA

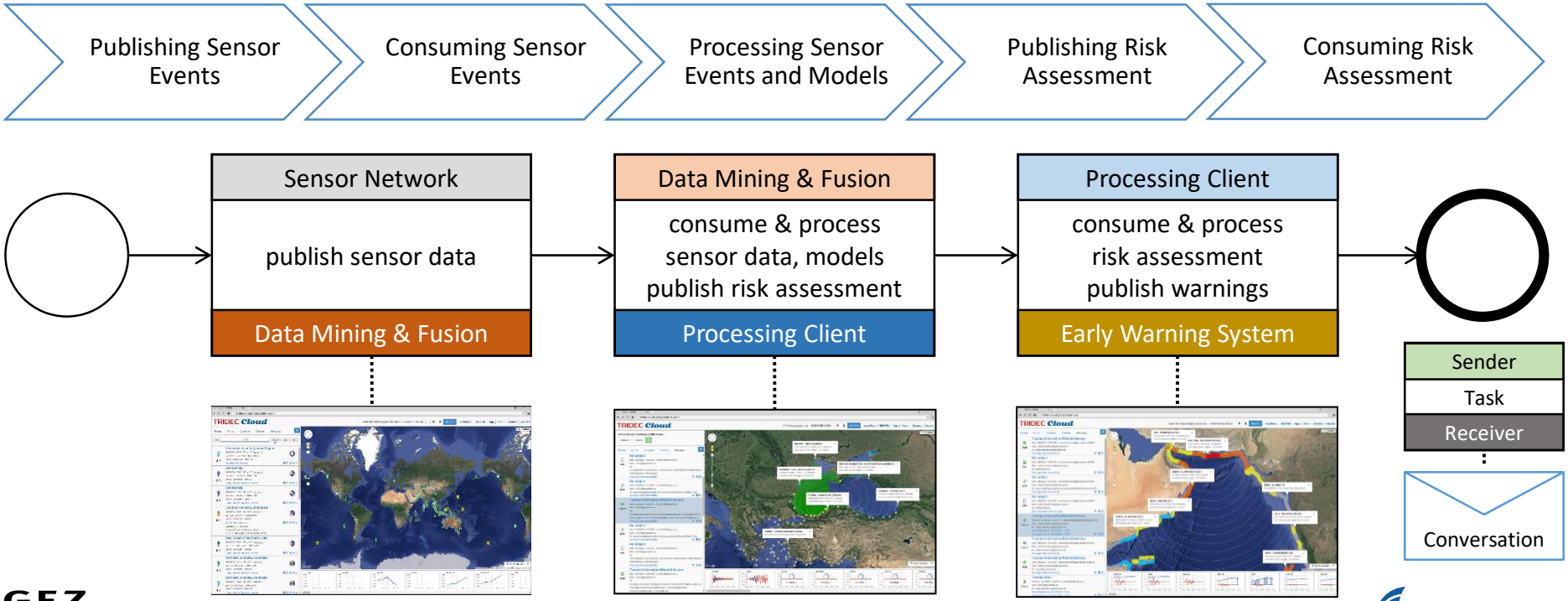
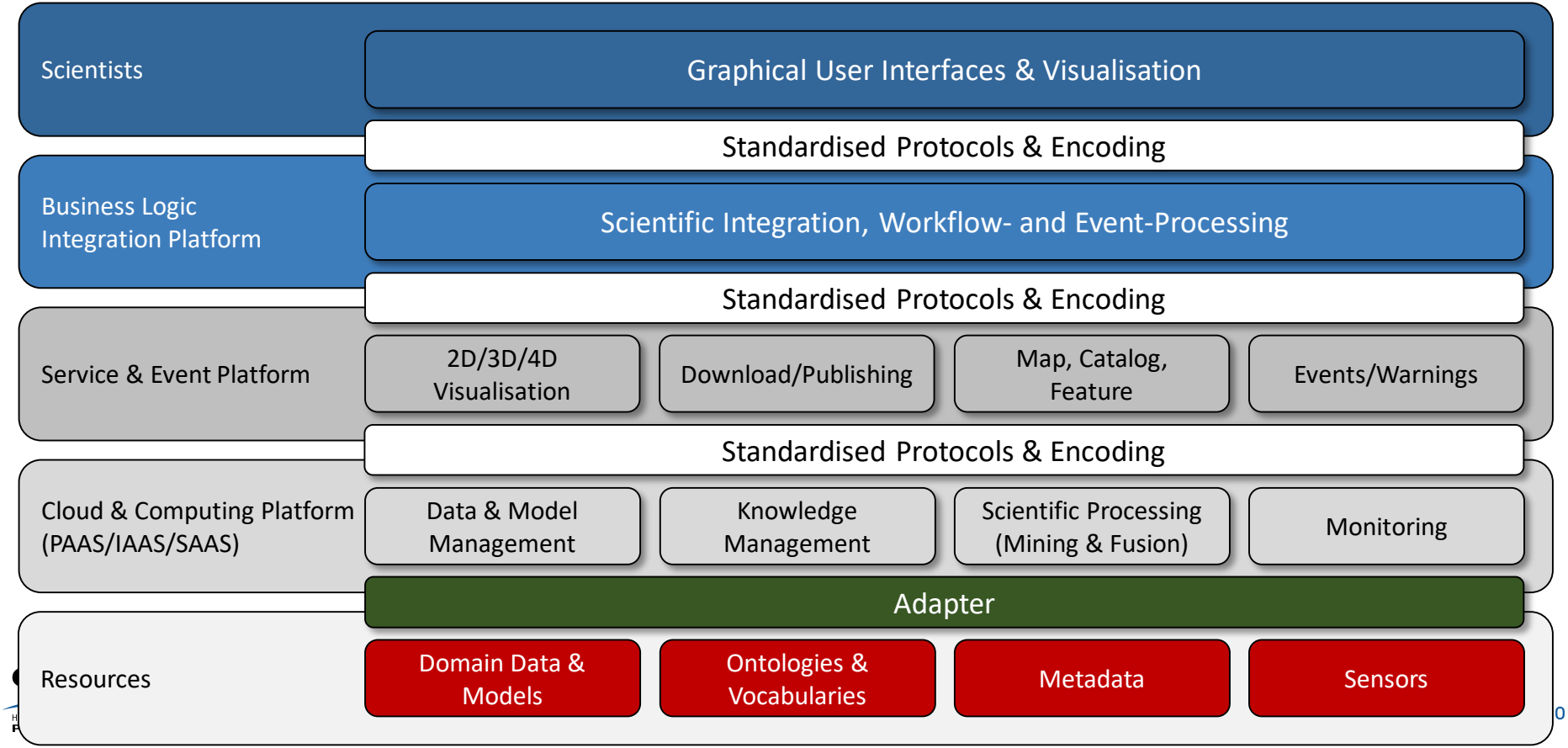


Diagram based on the Business Process Model Notation (BPMN) Adapted from Presentation EGU 2018, ESSI2.6

Reference Architecture for Scientific Infrastructures



SOA Principles & REST API in the TRIDEC Cloud

- REST: GET data
  - Embed resources like pre-processed maps, events, and notifications
    - Integration in – and automation of workflows
  - Embed for further use in any other system
    - Processing workflow composition and automation
- REST: POST data
  - Publish simulation & modelling results, and latest processed events
    - Distributed Event Detection and Processing
  - Publish (historic) archives of earthquakes and sea level measurements
    - Integration in – and automation of workflows
- SOA: Complex Event Processing and Composable Web Services
  - Get (be notified) and embed data automatically
    - Distributed Event Detection and Processing
  - Publish and Monitor (access) sensor events and simulation results
    - Sensor Web Technology, Integration into workflows
  - Start simulation or analysis processes (instantly) via REST as well
    - Distributed and web-based geo-processing

Thank you for your attention

Rainer Häner (rainer.haener@gfz-potsdam.de)

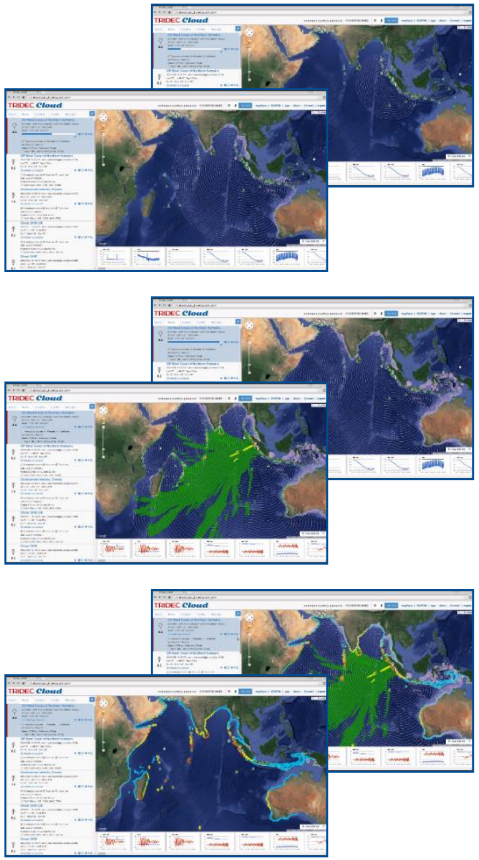
Helmholtz Centre Potsdam GFZ (German Research Centre for Geosciences)

eScience Centre

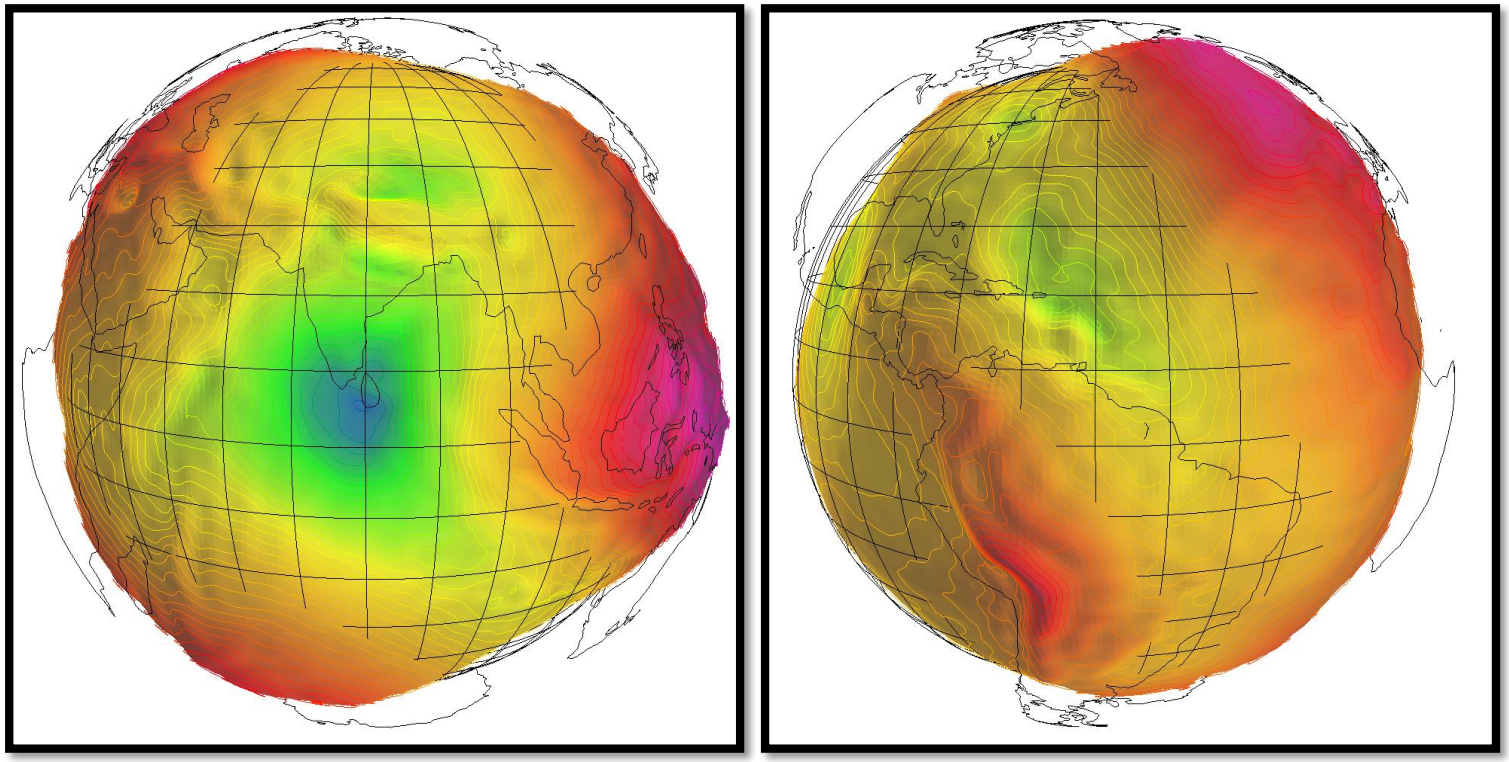
https://www.epos-ip.org/tcs/geological-information-and-modeling/data-services/wp15-services-and-architecture

Simulation

- Forecasts of the potential propagation of a tsunami
  - Tsunami travel times (TTT),
  - Estimated times of arrivals (ETA),
  - Wave jets and estimated wave heights (EWH/SSH), and
  - Mareograms
- Visualization of isochrones and tsunami-jets
- Analysis of time series
- Processing of tsunami forecast points (TFP) and tsunami forecast zones (CFZ) for further use, e.g. for generating warning messages



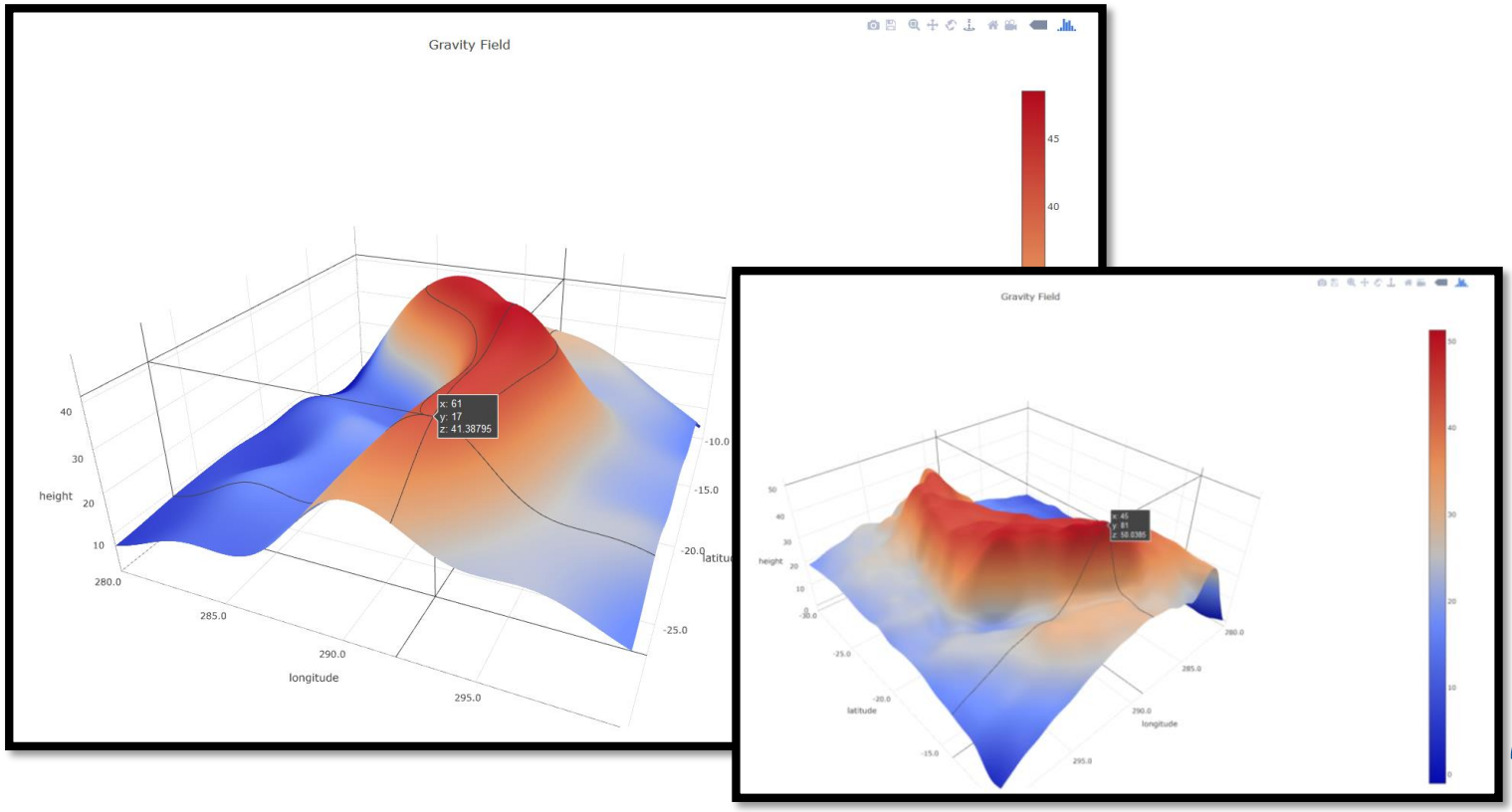
Interactive Visualisation of the Gravity Field



GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

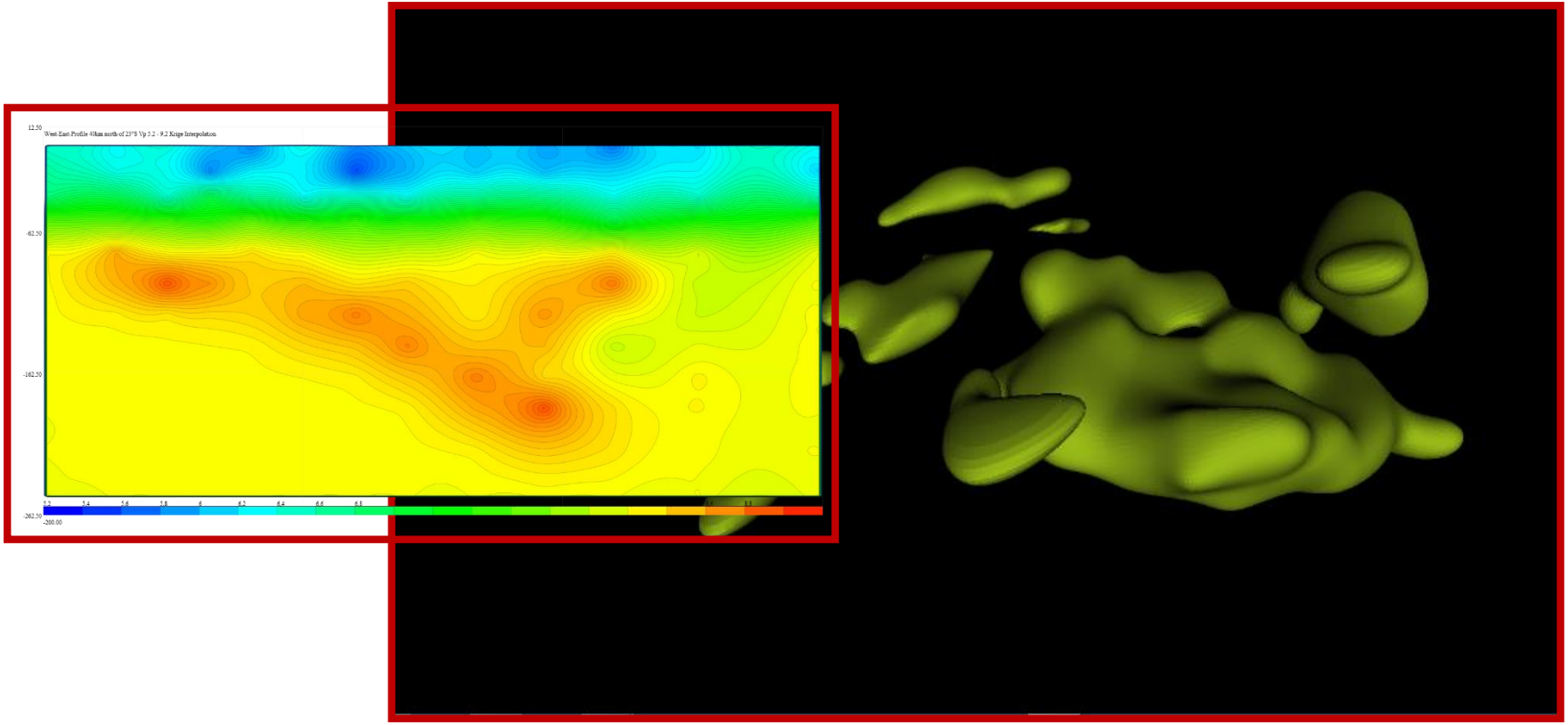
Gravity Field (Height Anomaly)



GFZ Helmholtz Centre Potsdam

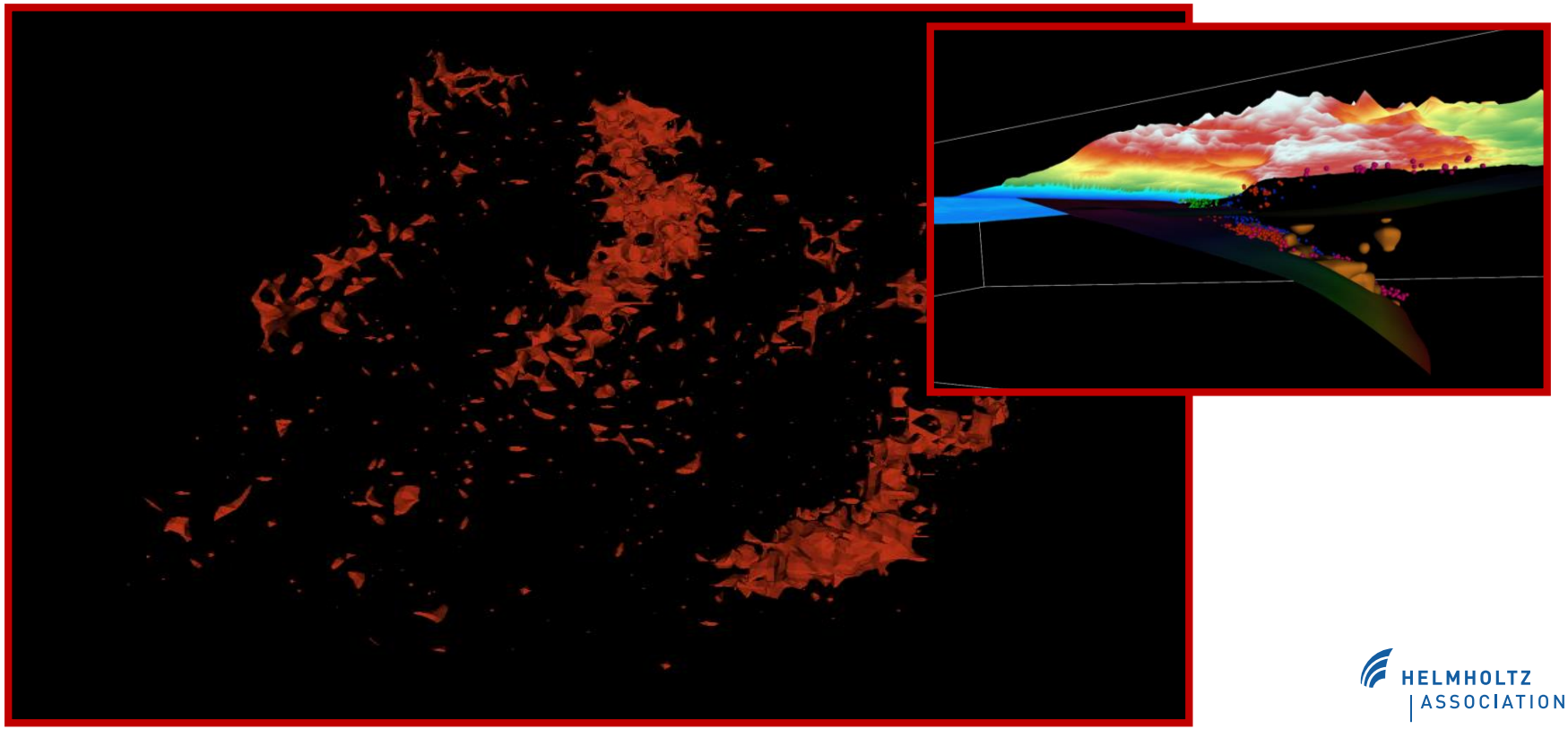
HELMHOLTZ ASSOCIATION

Approximated Geological Structure, Chile

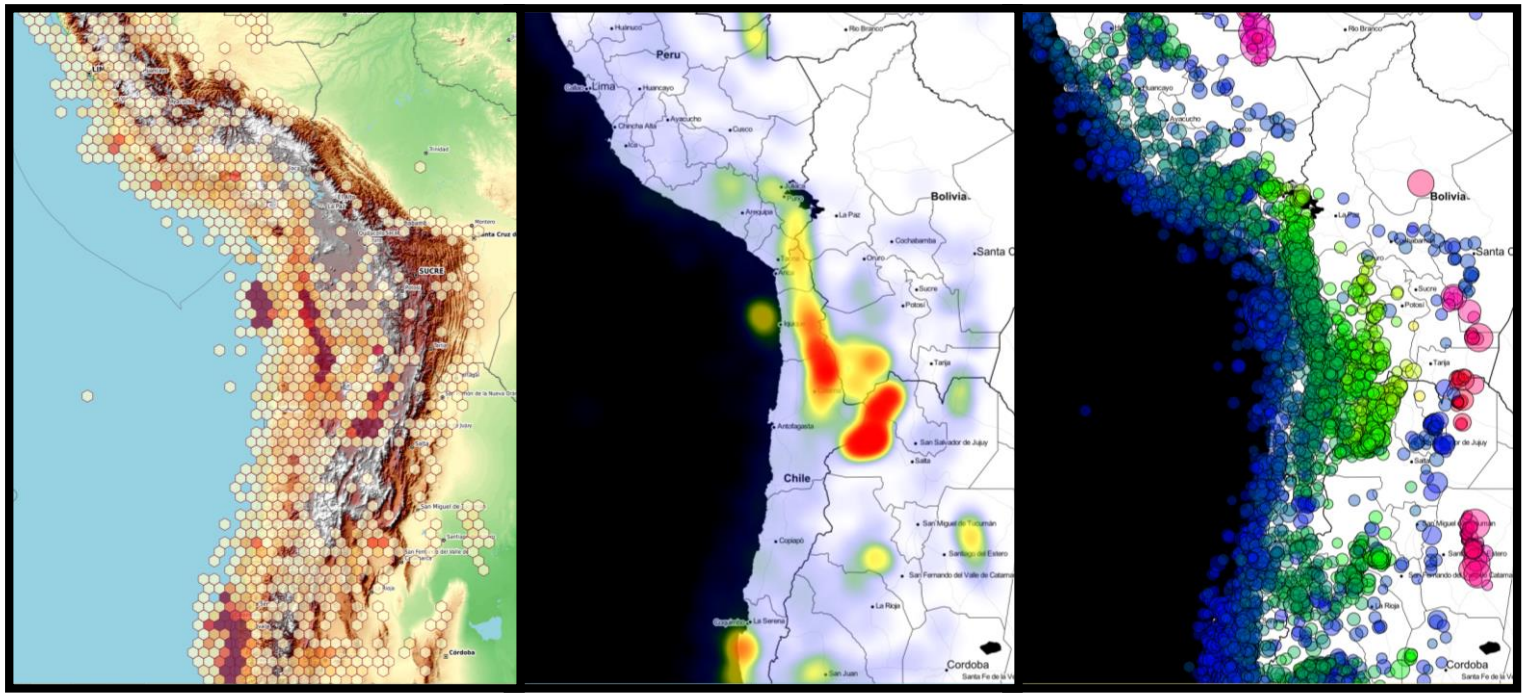


GFZ
Helmholtz Centre
POTSDAM

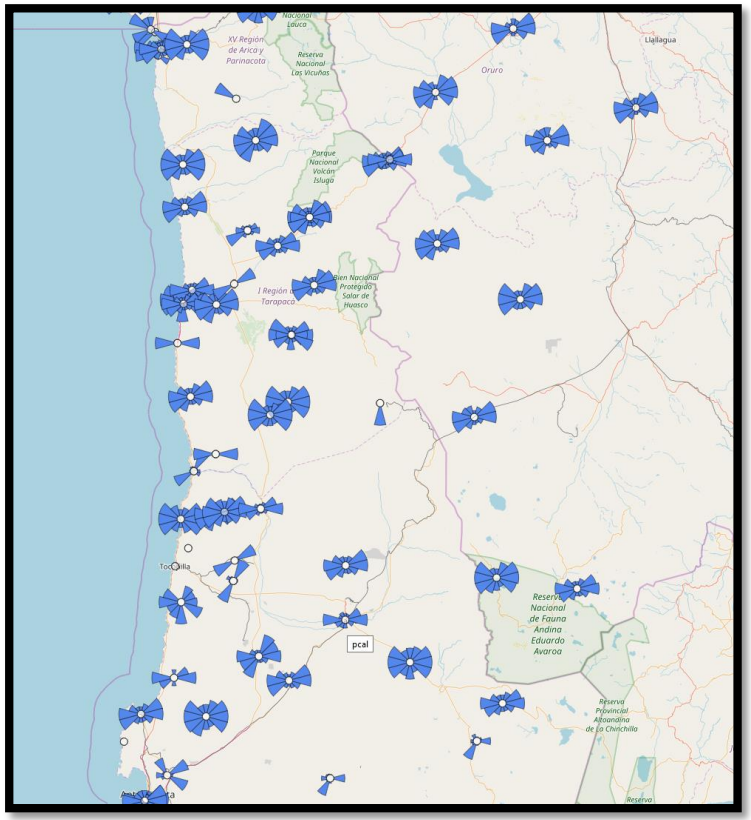
Earthquakes: Spatial Cluster Analysis, Chile

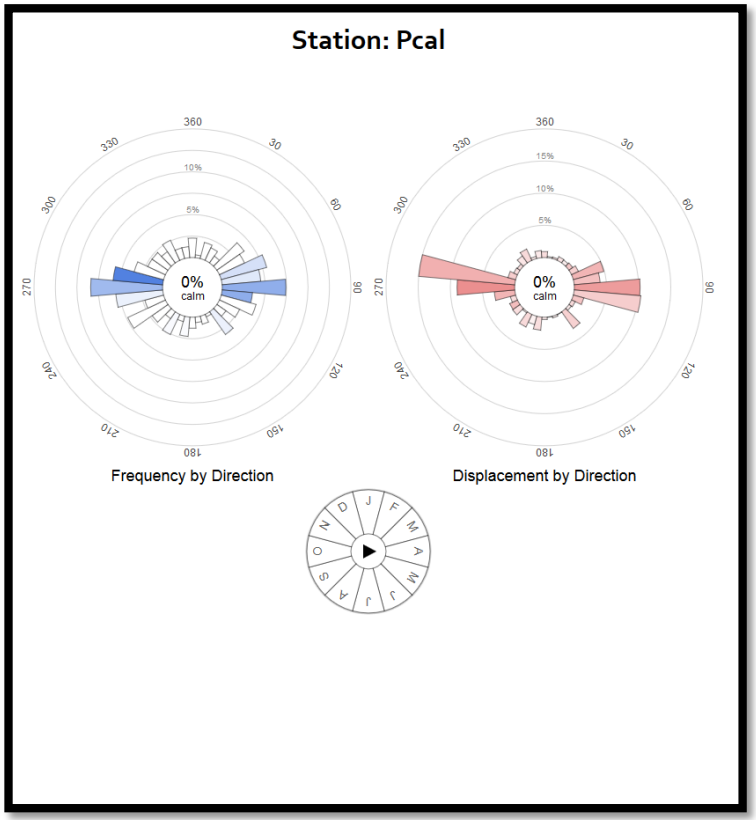


Earthquakes: Hexagonal Binning, Heatmap, Features

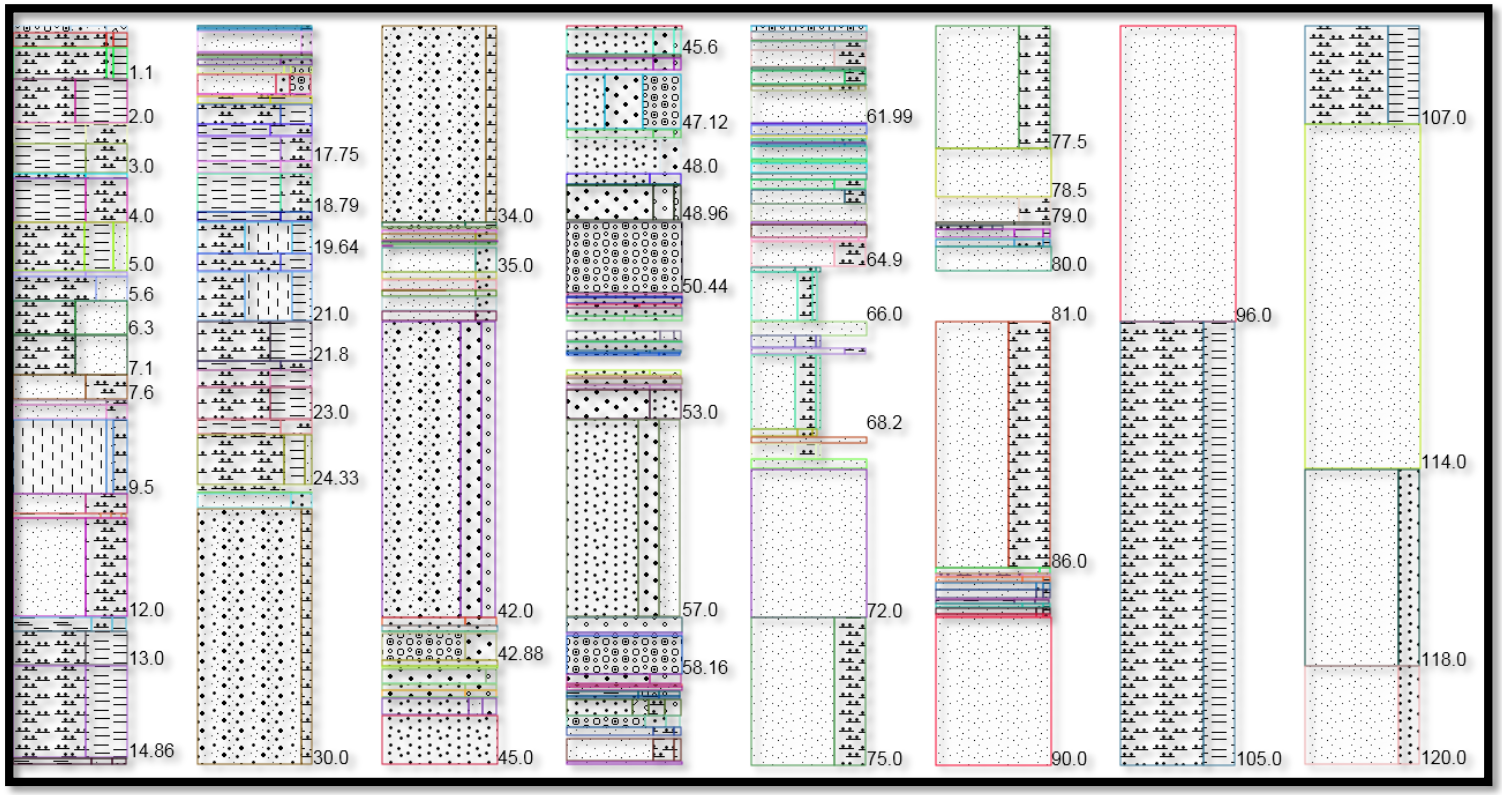


GNSS: Displacement Vector ("Wind-Rose") based on CGPS





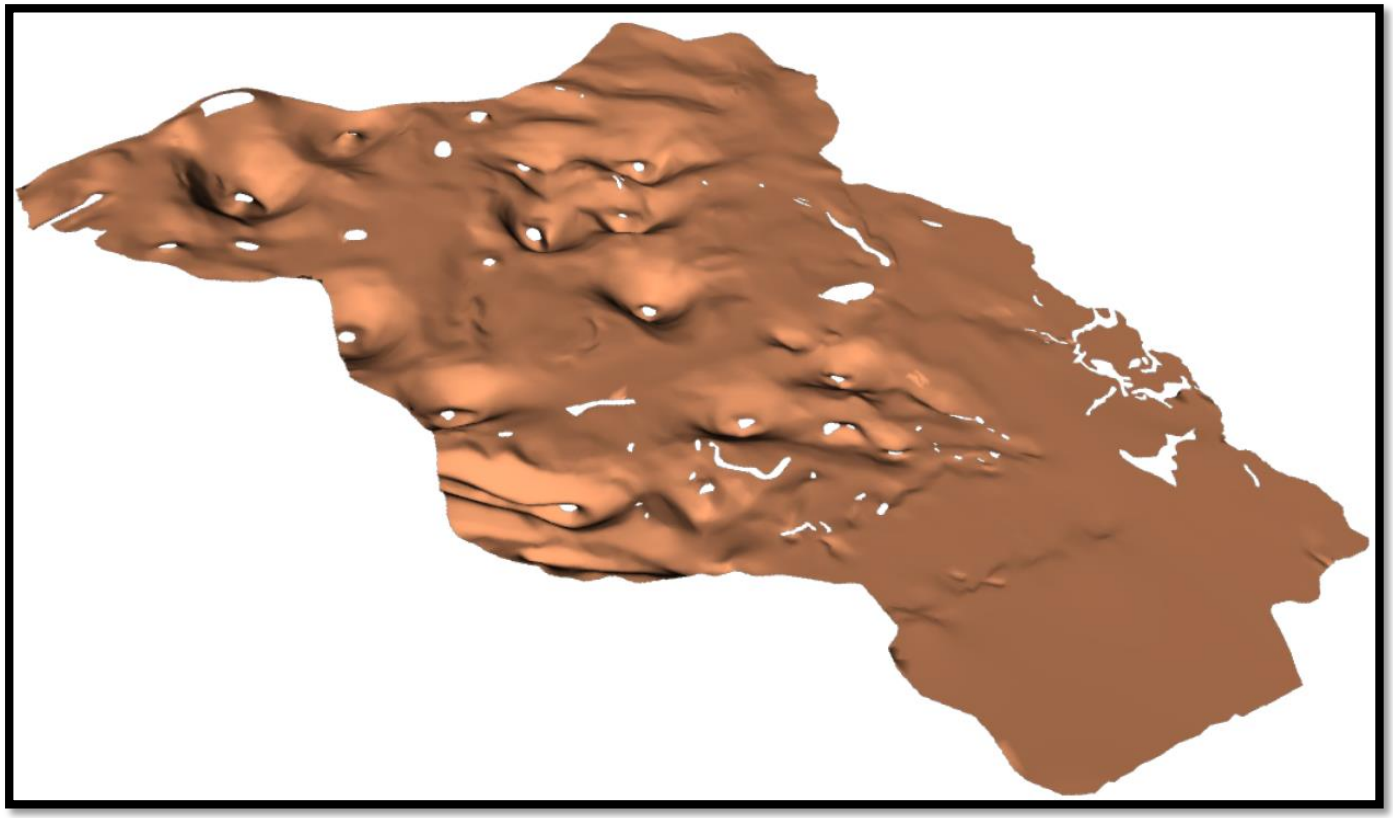
Subsurface Management: Lithology-Log



GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

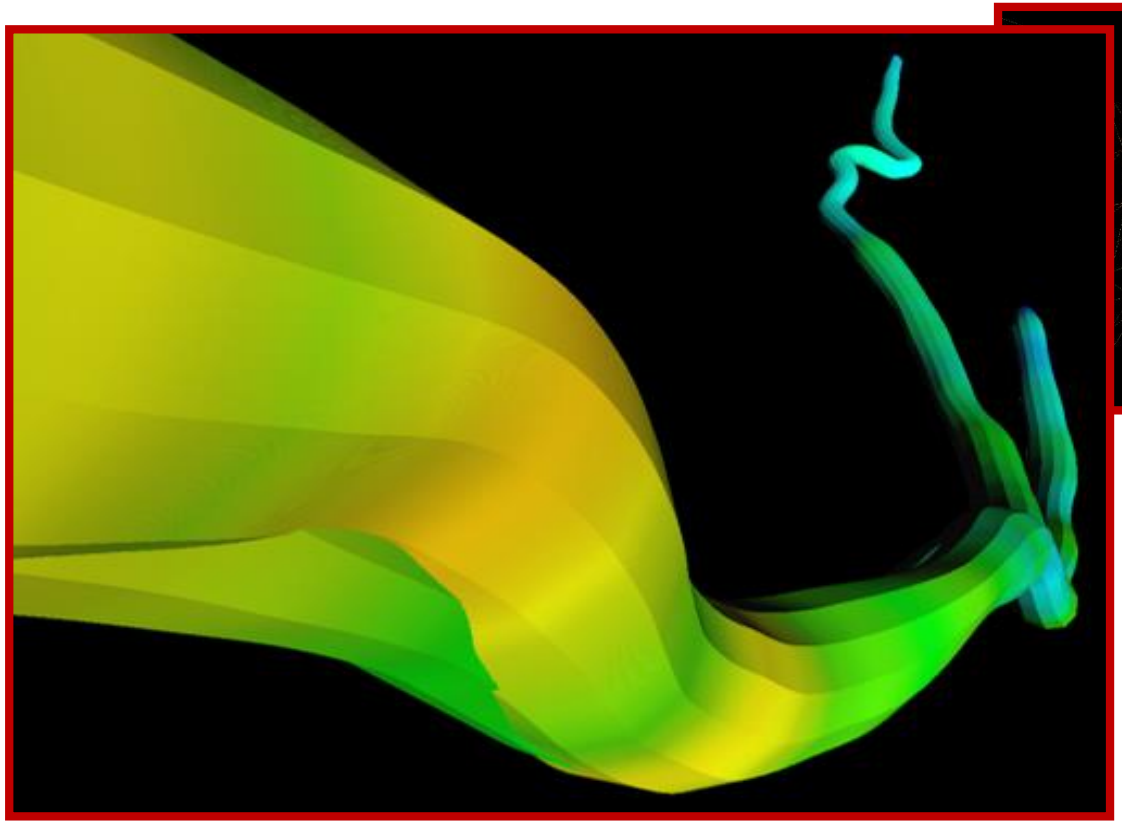
Subsurface Management: Tertiary Basis Brandenburg

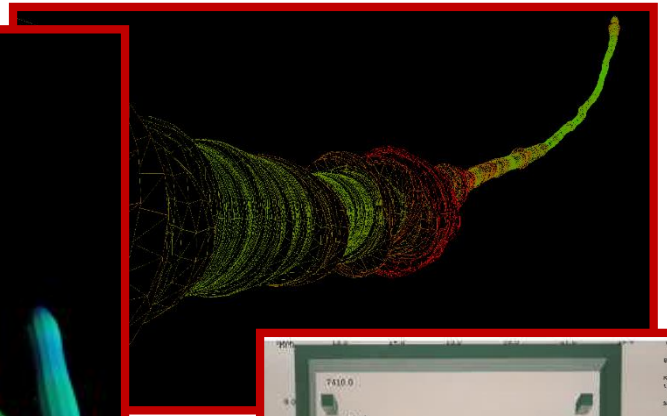


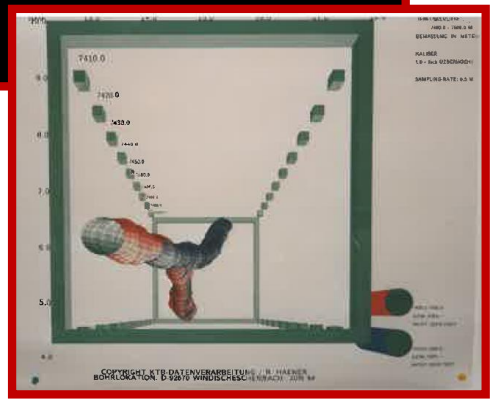
GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION

Subsurface Management: Borehole Geometry







GFZ Helmholtz Centre Potsdam

HELMHOLTZ ASSOCIATION