DISTRIBUTED SERVICES FOR ENABLING THE AUTOMATED PROCESSING OF BIG EARTH OBSERVATION DATA

Sebastian Drost[1,2], Christian Danowski-Buhren[1], Arne Vogt[1], Matthes Rieke[2]

1: Bochum University of Applied Sciences, Department of Geodesy
2: 52°North GmbH

Geospatial Sensing | Virtual 2020

Münster, August 31 – September 02

Motivation

- Increased sediment and material inputs into watercourses and dams caused by:
  - Intensification of agriculture (e.g. more frequent fertilization)
  - Increased number of extreme weather events (e.g. heavy rain, drought) as a result of climate change

  Fertilization   Heavy rainfall   Soil erosion

- ➢ New challenges for water authorities and suppliers regarding the quality control of drinking water
  - Need for an improved water monitoring program
  - Development of strategies to reduce pollutant inputs into water bodies

Project Overview

- **WaCoDiS** – Copernicus-based services for monitoring material inputs in watercourses and dams

- Funded by the Federal Ministry of Transport and Digital Infrastructure as part of the mFUND programme

- Wupper region in North-Rhine Westfalia, Germany

- Wupperverband as the responsible water authority
  - Operation of river dams and sewage treatment plants
  - Managing sensor network for in-situ data
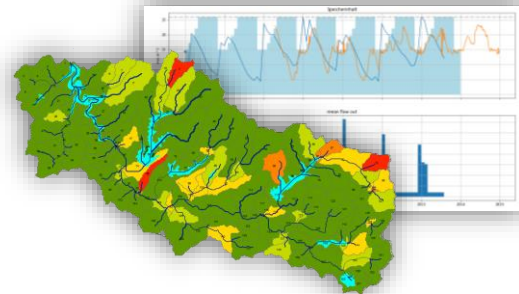
- Project Partners:

Project Goals

- Development of an improved water monitoring program that comprises:
  - Advanced environmental monitoring by providing analytics services and EO products
    - E.g. Analysis of the interannual variability of land-cover, soil moisture and nutrient balance
  - Combining in-situ and remote-sensing data in order to optimize hydrological models for
    - Sediment transport and erosion, nutrient pollution, …
  - Integration of CODE-DE infrastructure for processing Big Earth Observation Data
- ➢ Geolocating and quantifying material outputs from agricultural areas
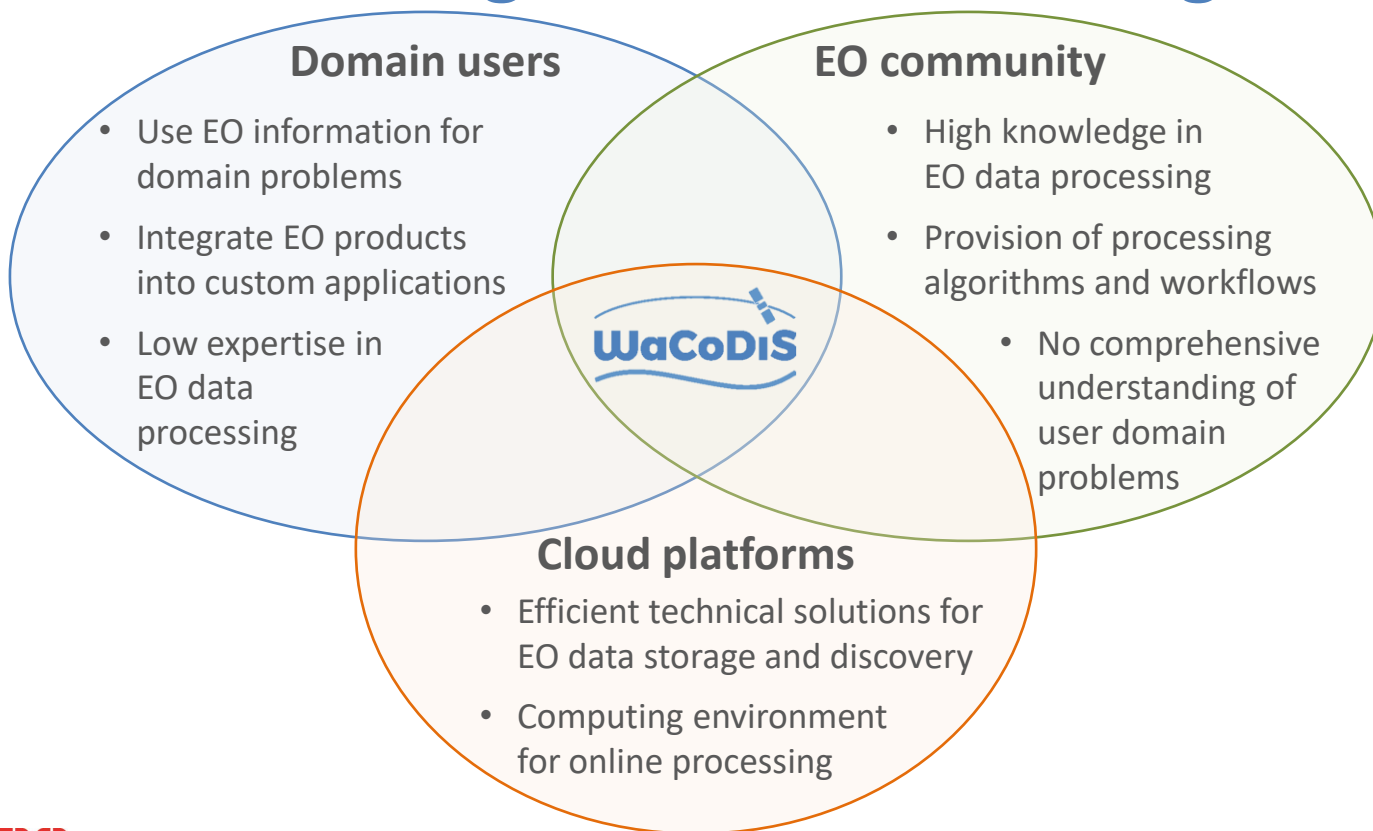
Detailed land classification

Change of water-land borders

Sediment transport modelling

Focus: Big EO Data Processing

Domain users

- Use EO information for domain problems

- Integrate EO products into custom applications

- Low expertise in EO data processing

EO community

- High knowledge in EO data processing

- Provision of processing algorithms and workflows

- No comprehensive understanding of user domain problems

WaCoDiS

Cloud platforms

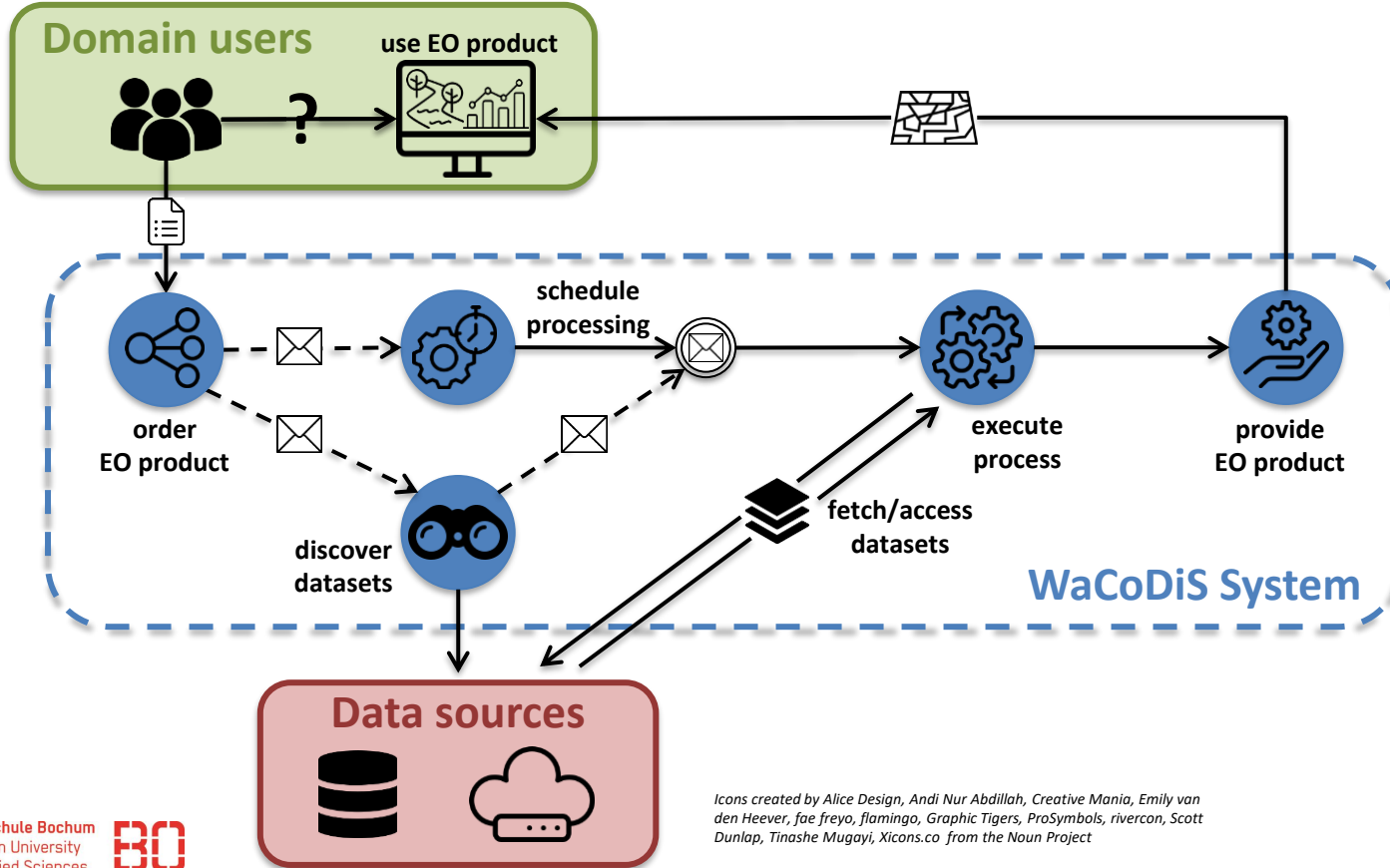- Efficient technical solutions for EO data storage and discovery

- Computing environment for online processing

Challenges in Big EO Data Processing

1. Sharing expert knowledge in EO data processing (EO processing workflows, algorithms and analysis information) with domain users

2. Facilitation of accessibility and usability of cloud computing platforms for non programmers

3. Increasing the interoperability and portability of EO processing services across different cloud providers

4. Support for reproducibility of processing results to increase the trust in EO based information

Automated Processing Workflow

Icons created by Alice Design, Andi Nur Abdillah, Creative Mania, Emily van den Heever, fae freyo, flamingo, Graphic Tigers, ProSymbols, rivercon, Scott Dunlap, Tinashe Mugayi, Xicons.co from the Noun Project

System Architecture

Microservice Design

- Single services will handle the process workflow tasks
- Loose coupling of collaborating services
- Container deployment

Publish/Subscribe Pattern

- Message-based communication via a message broker
- Asynchronous event handling (e.g. "job created")
- Process triggering as soon as required data is available

Data Centre Observation

- Dedicated components observe different data centres
  - Sensor Web, CODE-DE, DWD, …
- Metadata exchange of required datasets
 - Interested components subscribe for data availabilities
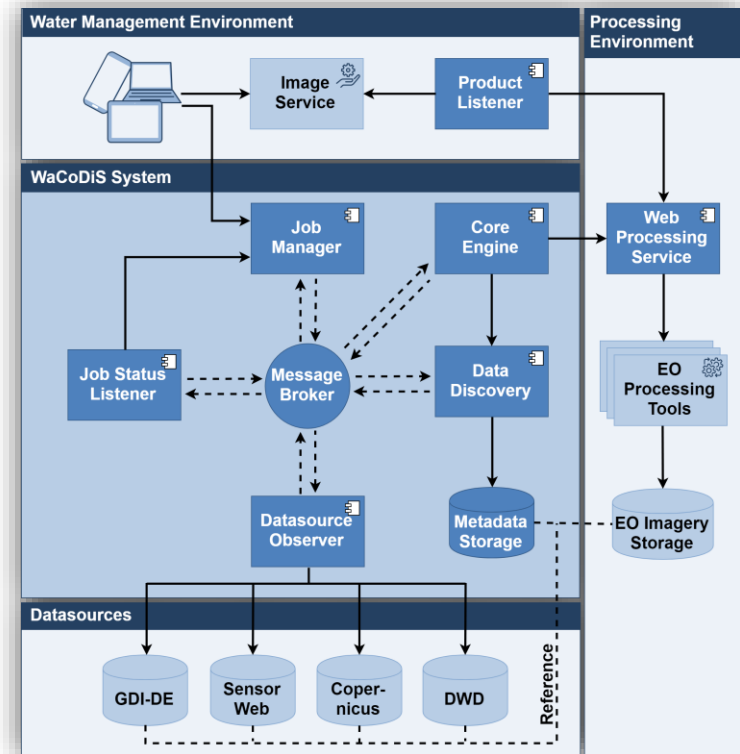
Domain Model

- **Execution triggers**
  - Event-based: *e.g. on data availability*
  - Pattern-based (crontab syntax): *0 0 1 * **

- **Spatio-temporal coverage**
  - Area of interest as GeoJSON bounding box
  - Temporal coverage as backwarts duration

- **Input data subset definitions**
  - Criterias relate to specific datasets
    - e.g. *mission ID* and *product level* for Sentinel-2 data

- ➤ API for message exchanges
  - Provision of required information for system components
 - Enables the extension by new components

```json
{
  "name": "interannual-soil-moisture-detection",
  "execution": {
    "pattern": "0 0 1 * *"
  },
  "temporalCoverage": {
    "duration": "PT1M"
  },
  "areaOfInterest": {
    "extent": [7.000, 52.000, 7.100, 52.100]
  },
  "processingTool": "soil-moisture-detection",
  "inputs": [
    {
      "sourceType": "CopernicusSubsetDefinition",
      "id": "LCC-INPUT-01",
      "satellite": "sentinel-2",
      "productLevel": "LEVEL2A",
      "maximumCloudCoverage": "75"
    }
  ]
}
```

Hochschule Bochum
Bochum University
of Applied Sciences

Approaching Interoperability

- Encapsulation of EO data processing by OGC Web Processing Service (WPS) 2.0
    - Standardized interface for synchronous and asynchronous processing tasks
    - Allows the definition of inputs, outputs and processing parameters

- WPS is the main entrypoint to all supported EO tools
    - Each WPS process encapsulates execution details for certain tools
    - WaCoDiS System interacts with WPS interface to trigger tool executions

- Standardized interface allows the lightweight introduction of additional tools that operate in the cloud
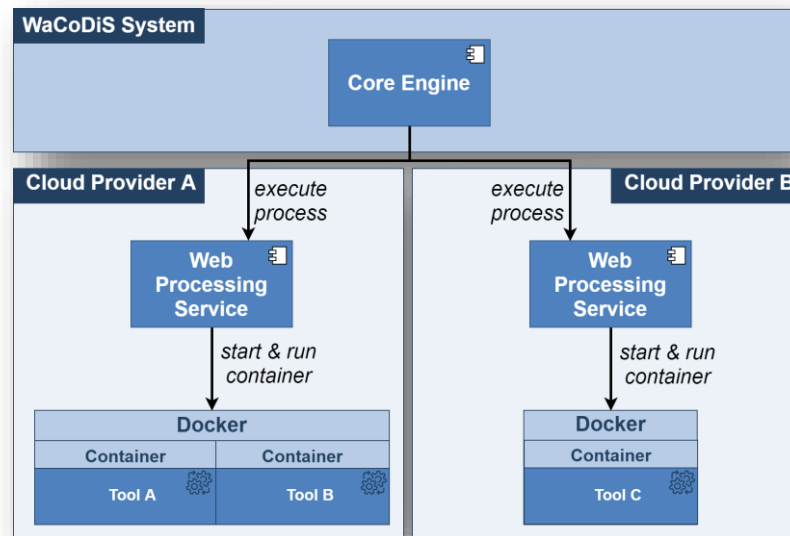
EO Product Ingestion
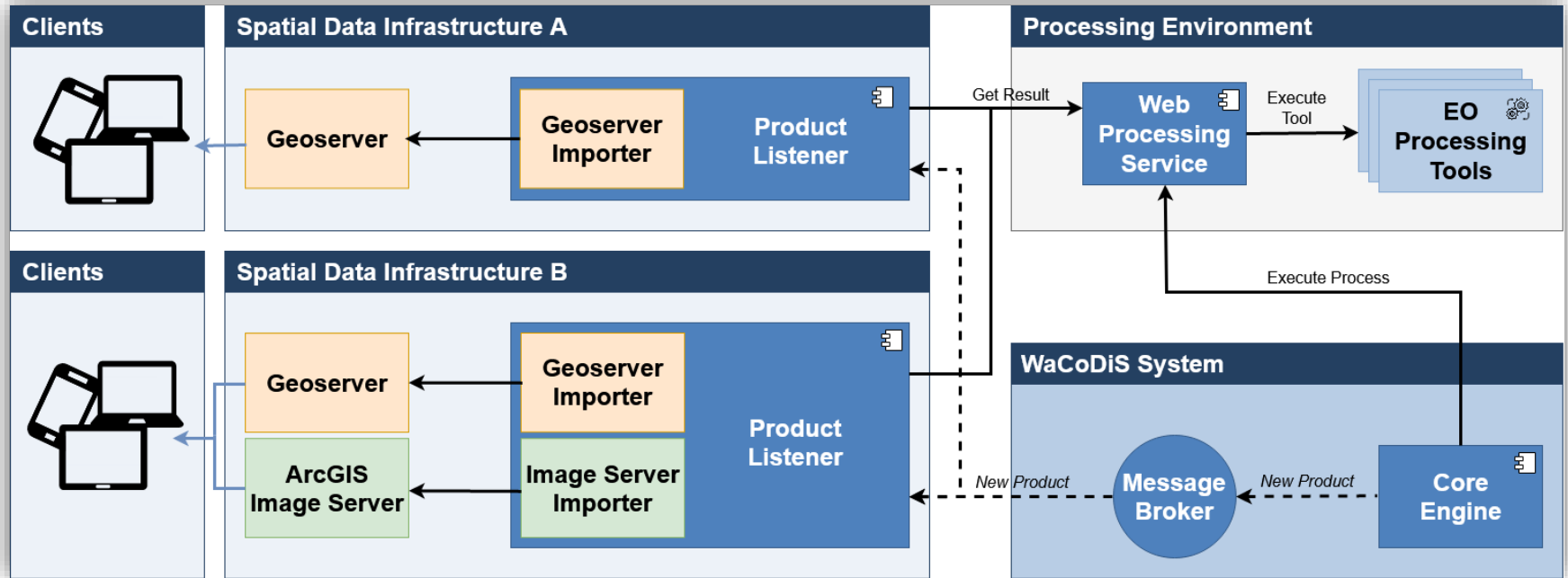
Reproducibility of Processing Results

- Domain users have to trust in automated generated EO information in order to reuse EO products for own applications in a reliable way
  - E.g. calculated land cover classes must be valid within a certain area to reuse it for hydrologic modeling and simulation

- Required meta information for EO products
  - Spatio-temporal coverage and creation timestamp
  - Information about the services that provide the product
    - E.g. ArcGIS Image Service information
  - Description of the underlying algorithm that produced the result
    - E.g. WPS process description (possible via *DescribeProcess* request)
  - Input datasets that were used for processing
  - Accuracy of the resulting product (especially for classification results)

Reproducibility of Processing Results

- WaCoDiS System generated metadata for a land cover classification
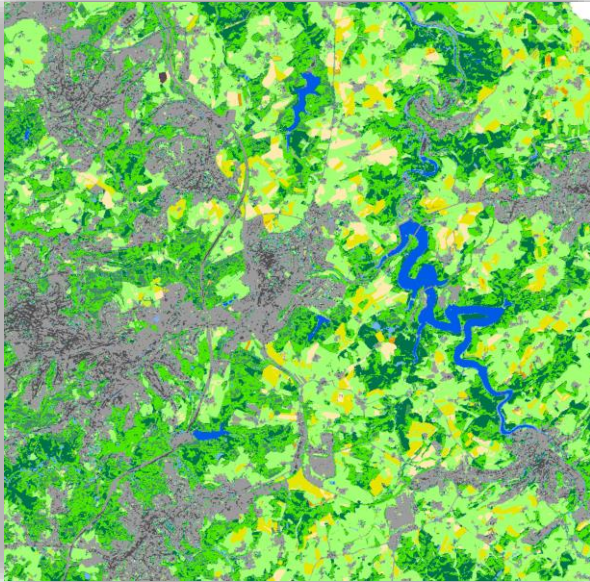
```
{
  "areaOfInterest": {
    "extent": [6.927, 51.279, 7.447, 51.337]
  },
  "timeFrame": {
    "startTime": "2020-08-12T10:30:31.024Z",
    "endTime": "2020-08-12T10:30:31.024Z"
  },
  "created": "2020-08-19T10:00:50.888Z",
  "productType": "land-cover-classification",
  "inputReferences": ["T56q5nMBSGSeK10aP4_T"],
  "envelopeService": "https://demo.wacodis.org/data-access/envelopes",
  "process": "de.hsbo.wacodis.land_cover_classification",
  "serviceDefinition": {
    "backendType": "ArcGISImageServerBackend",
    "productCollection": "EO:WACODIS:LAND_COVER_CLASSIFICATION",
    "baseUrl": "https://gis.wacodis.org:6443/arcgis",
    "serviceTypes": ["ImageServer"]
  }
}
```

Hochschule Bochum
Bochum University
of Applied Sciences

Deployment Patterns

CODE-DE Deployment (Concept)

- Use of the CODE-DE processing environment via virtual machines
  - Direct access to Sentinel and Landsat collections (mounted into VMs via S3 interface)
  - Computation resources limited by user category
    - Project setup: 8 vCPU, 64 GB RAM, 2 TB block storage, 1 TB file storage, 1 public IP
- VM 1: Hosting the WPS with external access
  - Triggering containerized EO tools via Docker REST API
- VM 2: Running predefined processing scripts
  - Subset of Sentinel Toolbox operators
- VM 3: Running custom developed algorithms
  - Creating EO products for water monitoring
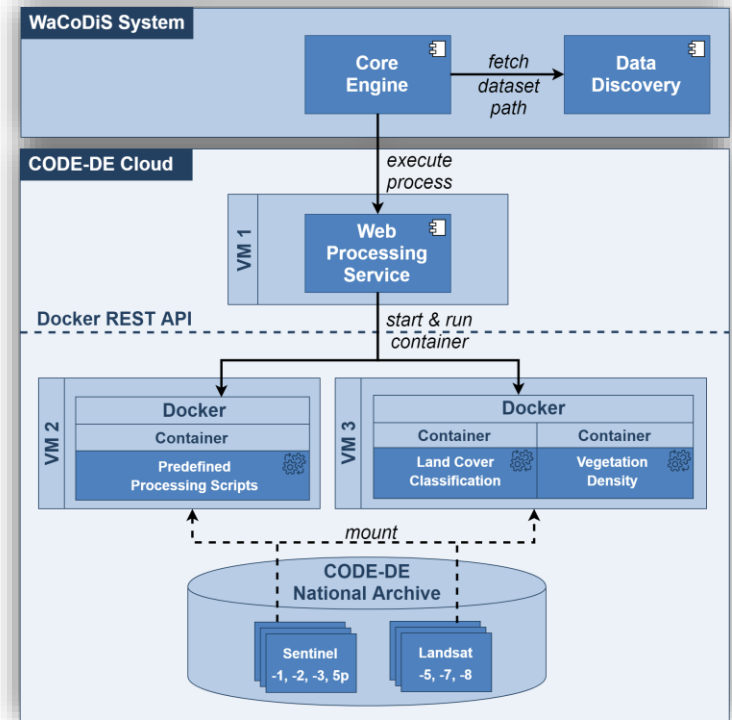
Conclusion

- The proposed WaCoDiS System facilitates the use of Earth Observation information for water monitoring tasks
    - Users order a product -> WaCoDiS handles processing and provision tasks automatically
- No vendor lock-in due to standardized WPS interface for wrapping EO tools
- Detailed processing metadata contribute to EO product reproducibility

Outlook

- System validation as part of an pre-operational deployment, considering different test scenarios defined by the Wupperverband
- Evaluation of the CODE-DE deployment concept
- Taking into account accuracy information for EO products (metadata inclusion)

ANY QUESTIONS OR COMMENTS?

Sebastian Drost

sebastian.drost@hs-bochum.de

Geospatial Sensing | Virtual 2020

Münster, August 31 – September 02

https://github.com/wacodis