

Big Geospatial Data Processing and Analysis Developments in the IQmulus project

Dániel Kristóf, Angéla Olasz

Institute of Geodesy, Cartography and Remote Sensing (FÖMI), Budapest, Hungary

Roberto Giachetta, Binh Nguyen Thai

Eötvös Loránd University, Faculty of Informatics (ELTE-IK), Budapest, Hungary

■ Current situation

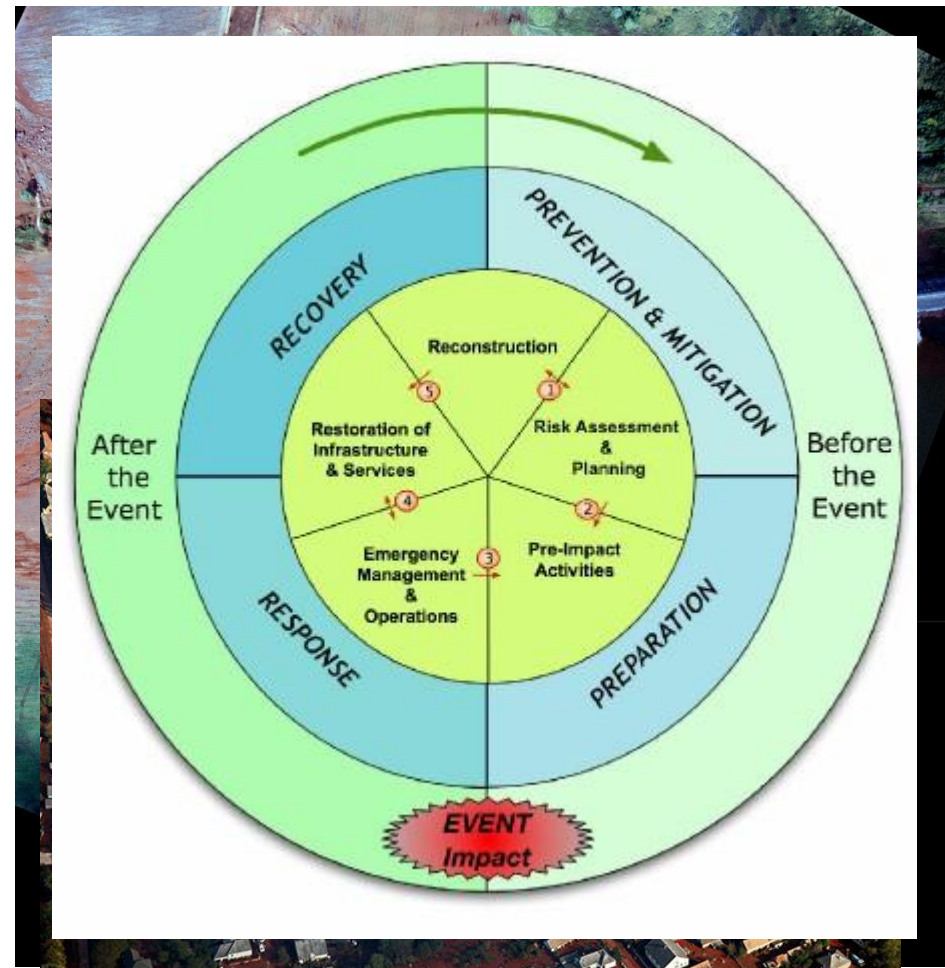
- Information in large geospatial datasets often exists but only accessed for damage assessment, „what went wrong?” analysis

■ Main objective

- To enable (optimized) use of large, heterogeneous geo-spatial data sets for better decision making through a high-volume fusion and analysis platform

■ Two test scenarios:

- Marine Spatial Planning
- Land applications for Rapid Response and Territorial Management

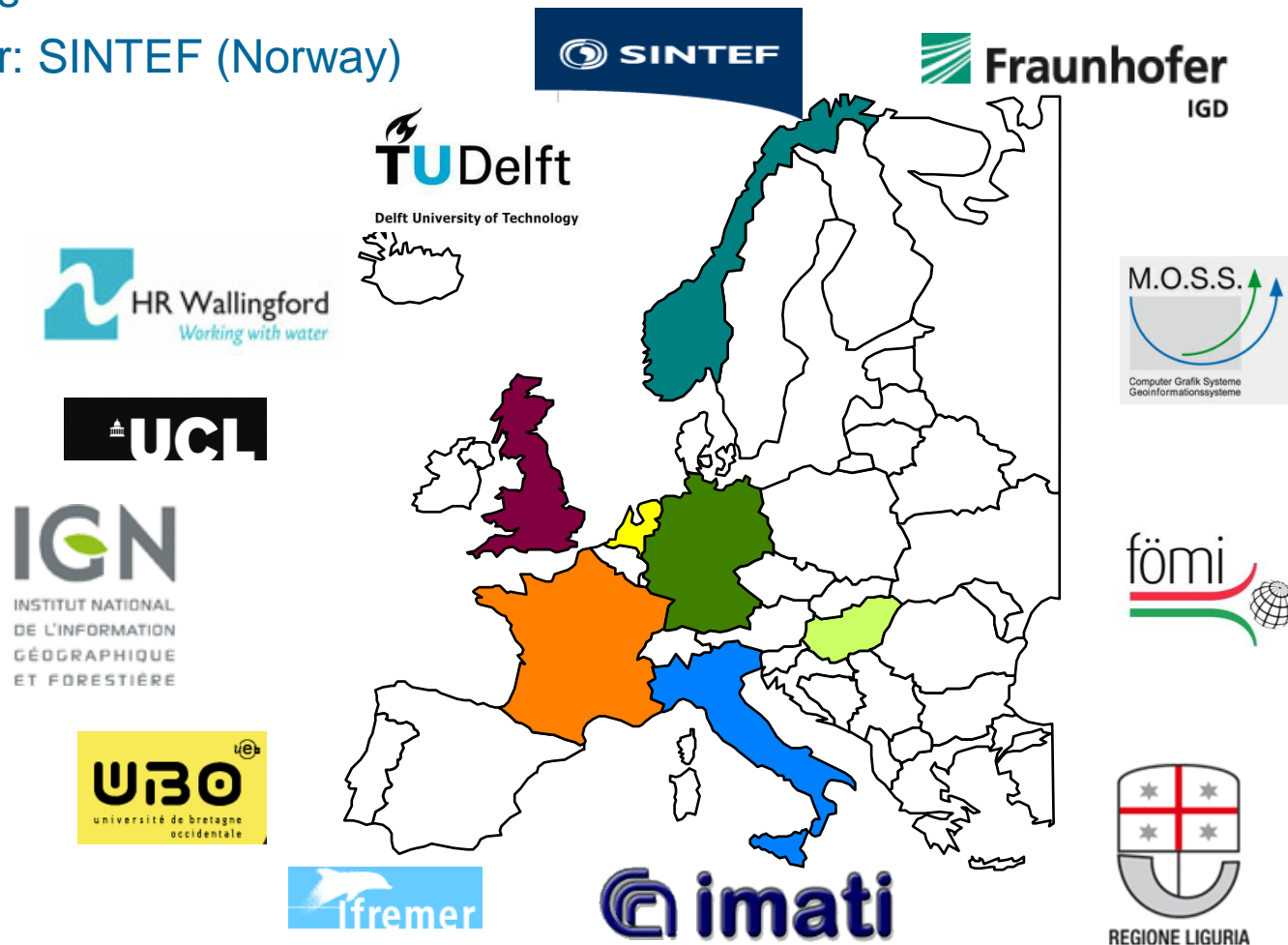


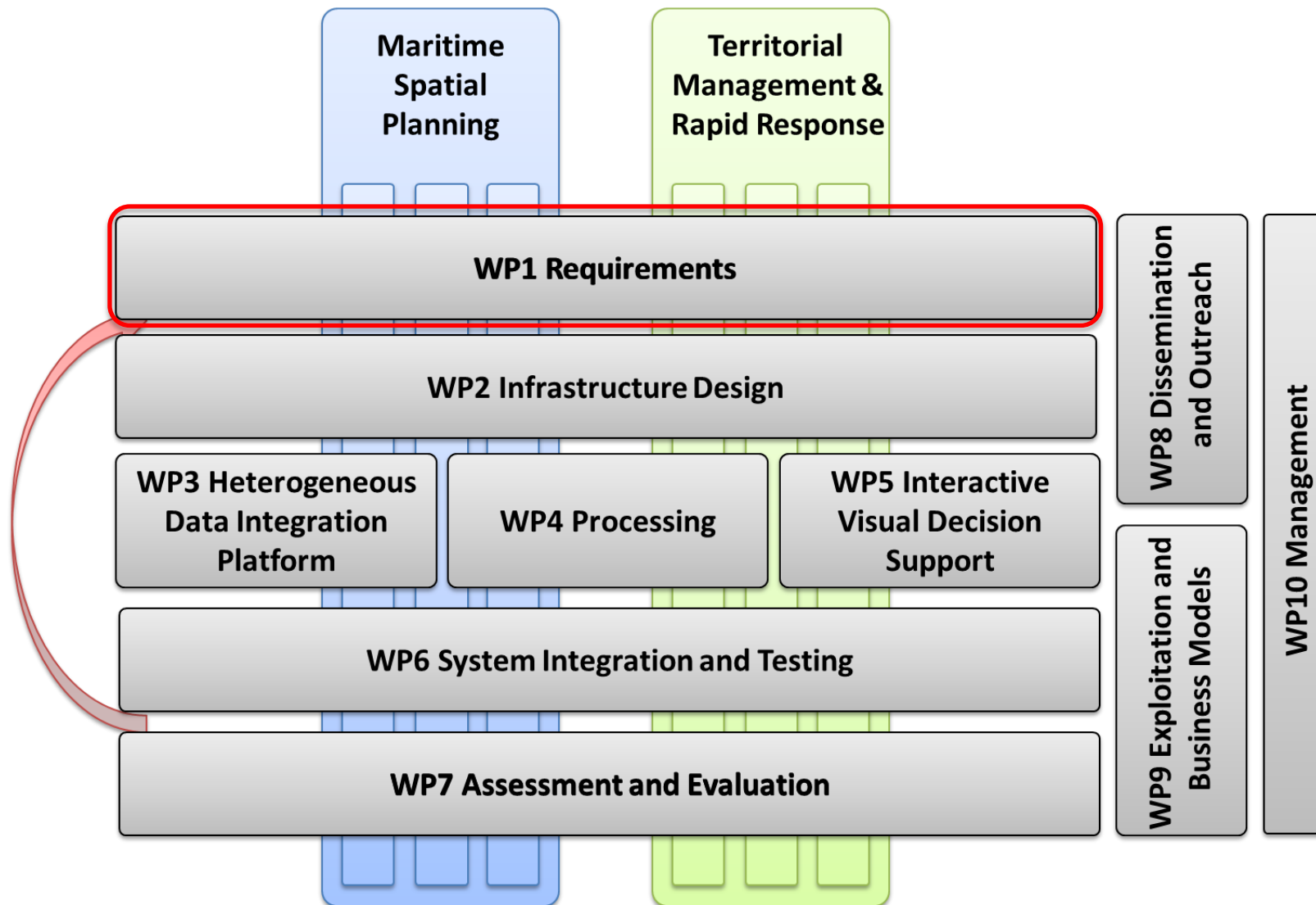
Ajka, October 2010

- Full title: „A High-Volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages and Volumetric Data Sets”
- A 4-year Integrating Project (IP) in the area of Intelligent Information Management within ICT 2011.4.4 Challenge 4: Technologies for Digital Content and Languages
- Started on November 1, 2012, and will finish October 31, 2016
- Total budget: 10 147 459€, EU contribution: 8 100 000€

■ 12 European partners from 7 countries

- university teams, applied research institutes, SMEs, national and regional organizations
- coordinator: SINTEF (Norway)





Issues - WP1 - Requirements

iqmulus.igd.fraunhofer.de/redmine/projects/wp1-requirements/issues?utf8=✓&set_filter=1&f[]=status_id&op[status_id]=o&f[]=&c[]=tracker&c[]=status

Home My page Projects Help

Logged in as dkristof My account Sign out

IQmulus » WP1 - Requirements

Search: » WP1 - Requirements

Overview Activity Roadmap **Issues** New issue Wiki Settings

Issues

▼ Filters

☒ Status Add filter

► Options

✓ Apply ↻ Clear 💾 Save

✓ #	Tracker	Status	Priority	Subject	Assignee	Updated	Target version
1629	Workflow	New	Urgent	US2 Individual tree extraction from urban LMMS data		10/02/2014 09:42 AM	
1628	Workflow	New	Urgent	US1 Detection of buildings for monitoring and cadastral updating		09/30/2014 10:19 PM	
1627	Workflow	New	Urgent	MS4 Measuring submarine dune migration		09/30/2014 10:12 PM	
1626	Workflow	New	Urgent	MS3 Simple feature extraction		09/30/2014 10:08 PM	
1625	Workflow	New	Urgent	MS2 Distance of survey data from a surface (Error checking)		09/30/2014 10:06 PM	
1624	Workflow	New	Urgent	MS1 Elevation model from point cloud data		09/30/2014 10:05 PM	
1623	Showcase	New	Urgent	Urban Showcase 1.2.2_SC3		09/30/2014 01:39 PM	
1622	Workflow	New	Urgent	LS5 Comparison of simulated floods/landslides with observed data		09/30/2014 09:58 PM	
1621	Workflow	New	Urgent	LS3 Flood and waterlogging detection		10/07/2014 09:46 AM	
1620	Workflow	New	Urgent	LS4 Detection and characterization of landslides		09/30/2014 09:50 PM	
1619	Showcase	New	Urgent	Marine Showcase 1.2.2_SC1		09/30/2014 01:23 PM	
1618	Workflow	New	Urgent	LS2 Analysis of precipitation data		09/30/2014 09:57 PM	
1617	Workflow	New	Urgent	LS1 Multiresolution model for land monitoring		10/07/2014 02:03 PM	
1616	Showcase	New	Urgent	Integrated Land Showcase 1.2.2_SC2		09/29/2014 10:34 PM	
1615	Requirement (IQmulus)	New	Normal	TR Implement adequate parallel algorithms		09/23/2014 01:57 PM	
1614	Requirement (IQmulus)	New	Low	TR Implement User Management for System Access		09/23/2014 01:57 PM	
1613	Use case (IQmulus)	New	High	1.2.1_22_UC1		09/23/2014 01:57 PM	
1612	Use case (IQmulus)	New	High	1.1_54_UC1		09/23/2014 01:57 PM	

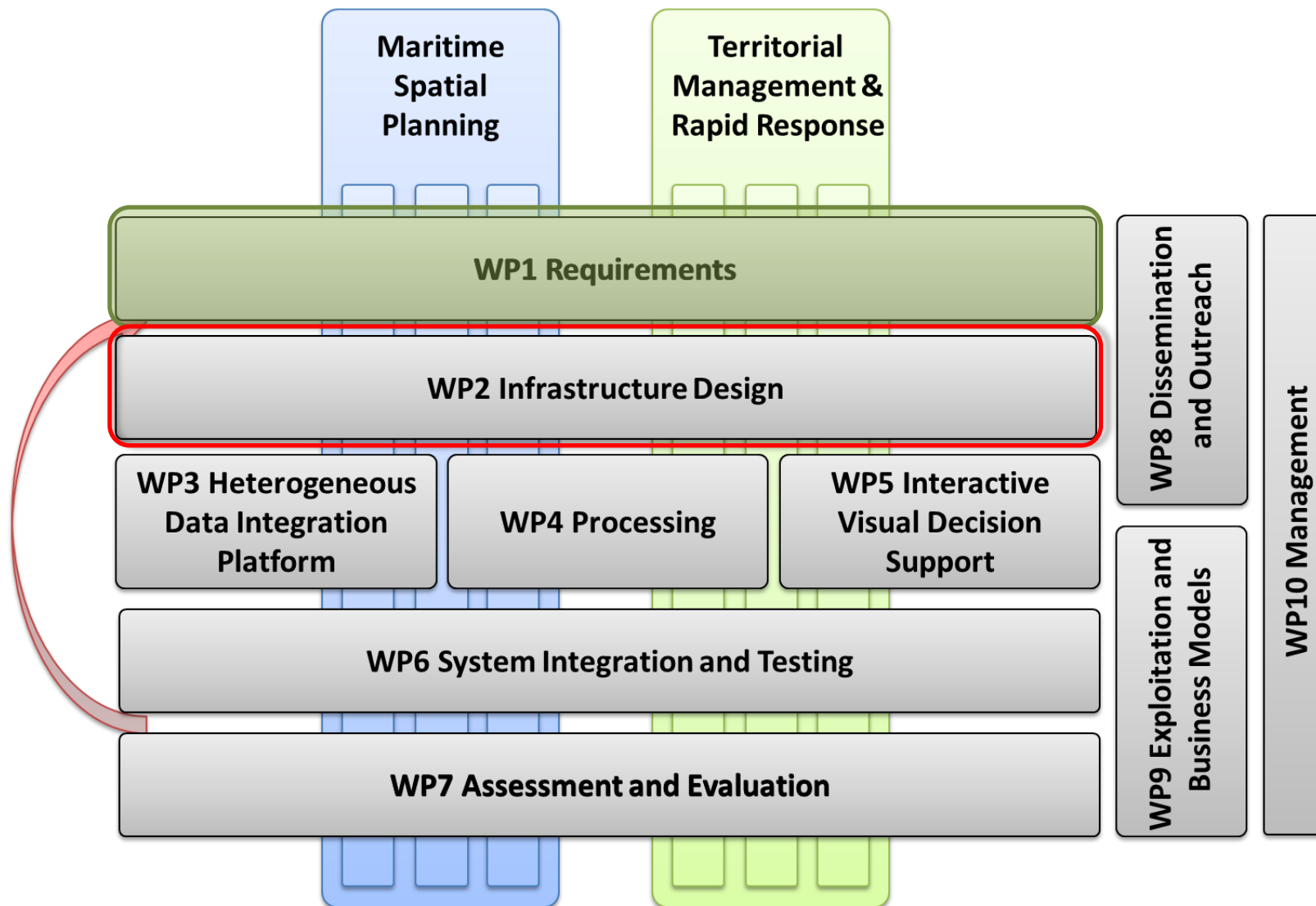
Issues

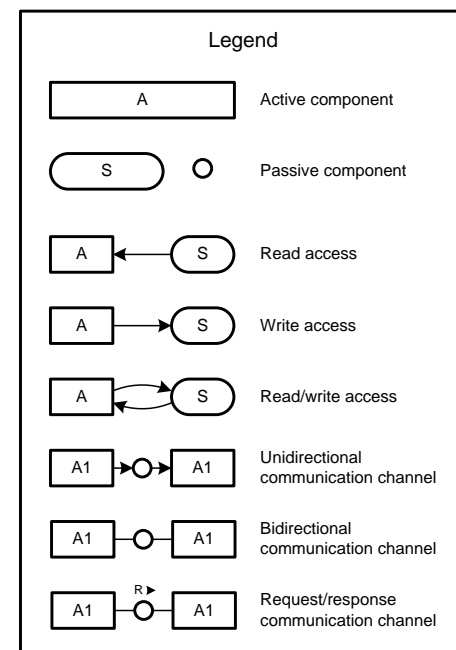
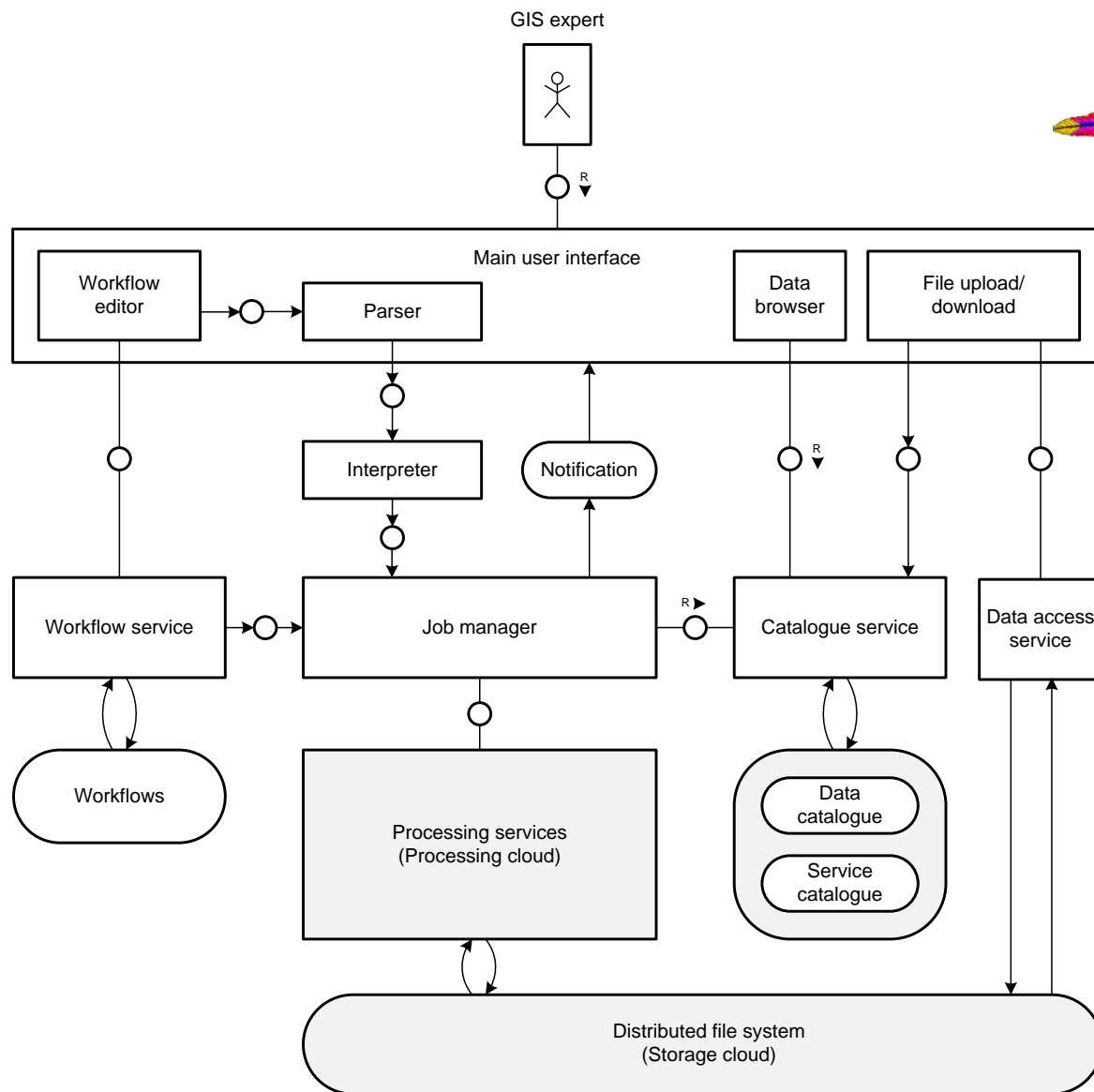
[View all issues](#)

[Summary](#)

Custom queries

Meine offenen Tickets nach Zielversion with_category





IQmulus Workflow Editor

 Save
  Cut
  Copy
  Paste

Workflows

- Common
 - Select trees
 - Change detection
- Private
 - Coregister DTM
 - Interpolation

Cue card

You are currently working on data from **CityModel**. Enter a **command** to manipulate the data. Enter **end** to close **CityModel**.

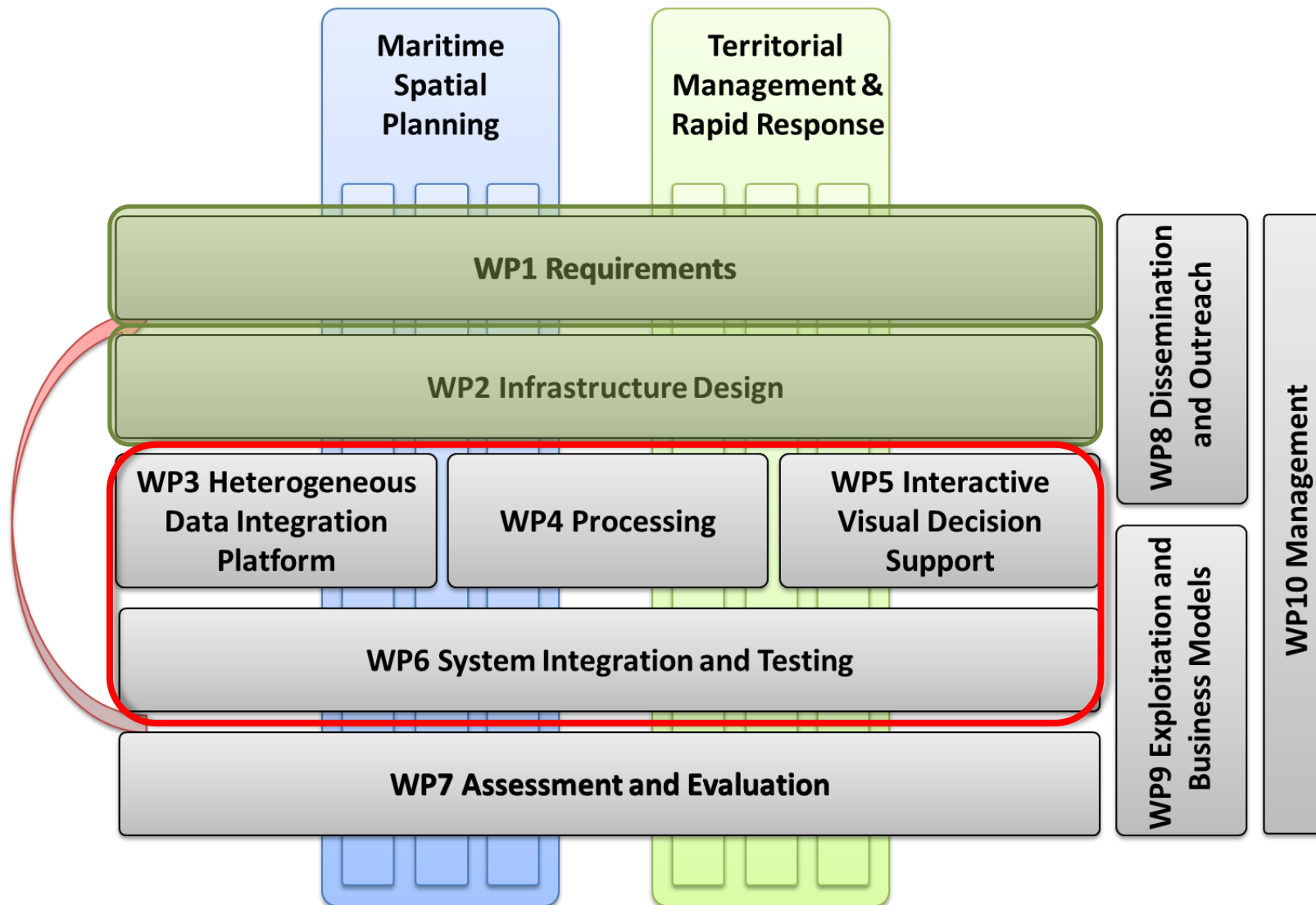
Possible commands are:

- exclude**
Remove certain objects from loaded data
- select**
Select certain objects from loaded data and exclude others
- add to**
Add loaded data to a given data set
- end**
Closes the current data set

* Select trees ✕

```

1 with [Area] do
2   exclude NonStaticObjects
3   select added Trees and deformed FacadeElements
4   add to [CityModel]
5 end
6
7 with [CityModel] do
8   exclude Antennas
9   c
10  v calibrate
11 end
12  colorize
   calculate
  
```



Functional

pointcloud-2-pointcloud distance
surface normal estimation
simplification
critical point extraction
curvature estimation

.....

Generic with respect to the context

Domain-specific

analysis, classification, feature
extraction, surface generation,
change detection ..

*The context defines the “meaning”
and target of the analysis*

Issues - WP4 - Processing S...
+

iqmulus.igd.fraunhofer.de/redmine/projects/wp4-processing-services/issues
Google

Home My page Projects Help
Logged in as dkristof My account Sign out

IQmulus » WP4 - Processing Services

Search:
» WP4 - Processing Services

Overview Activity Roadmap **Issues** New issue Wiki Settings

Issues

Filters
☒ Status open
Add filter

Options

Apply Clear Save

#	Tracker	Status	Priority	Subject	Assignee	Updated	Target version
1660	Bug	Solved	Immediately	Compilation error for Services 70-76-80	Roberto Giachetta	11/11/2014 03:59 PM	
1657	Service (WP4)	Assigned	Normal	Convert cumulated precipitation to regular grid		10/08/2014 03:10 PM	
1656	Service (WP4)	Assigned	Normal	Convert multi-resolution for hydrology to regular grid		10/08/2014 02:05 PM	
1655	Service (WP4)	New	Normal	Compute cumulated precipitation over multi-resolution for hydrology		10/08/2014 02:00 PM	
1654	Service (WP4)	New	Normal	Measuring surface displacement		10/08/2014 02:55 PM	
1653	Service (WP4)	Assigned	Normal	Convolution filtering		10/08/2014 02:54 PM	D4.2.2
1652	Service (WP4)	Assigned	Normal	Raster thresholding		10/08/2014 02:53 PM	D4.3.2
1651	Service (WP4)	Assigned	Normal	Raster Resampling		10/08/2014 03:07 PM	
1650	Service (WP4)	Assigned	Normal	Tracking of critical points		10/08/2014 01:37 PM	
1649	Service (WP4)	Assigned	Normal	Critical Points		10/08/2014 02:39 PM	D4.3.2
1648	Service (WP4)	Assigned	Normal	Approximation of rainfall data by LR-splines		10/08/2014 01:31 PM	D4.4.2
1647	Service (WP4)	Assigned	Normal	Approximation of rainfall data by radial basis functions (RBF)		10/08/2014 01:25 PM	D4.4.2
1646	Service (WP4)	Assigned	Normal	Approximation rainfall data by Kriging		10/08/2014 10:44 AM	D4.4.2
1645	Service (WP4)	Assigned	Normal	Multi-resolution triangulation for hydrology		10/08/2014 10:33 AM	D4.4.2
1644	Service (WP4)	Assigned	Normal	Detection of flow lines and drainage basins from triangle meshes		10/08/2014 10:29 AM	D4.3.2
1643	Service (WP4)	Assigned	Normal	Constrained triangulation		11/05/2014 10:46 AM	D4.4.2
1642	Service (WP4)	Assigned	Normal	Sub-sampling (thinning) of Point Cloud		10/08/2014 10:21 AM	D4.4.2

Issues

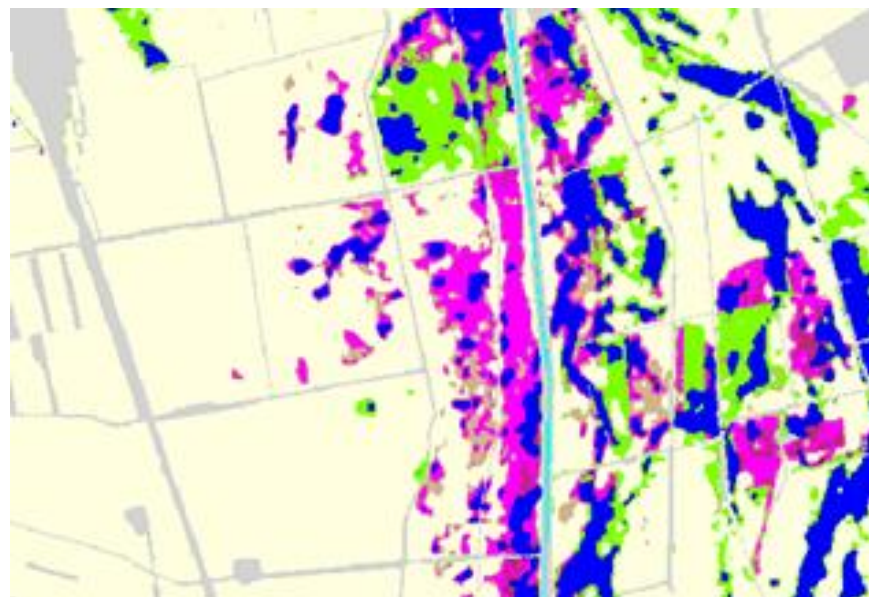
[View all issues](#)
[Summary](#)

Custom queries

[by responsible](#)
[Meine offenen Tickets nach Zielversion with_category](#)

User Story:

„I want to quickly delineate and categorize flooded areas based on satellite/aerial imagery and combine it with existing spatial datasets to provide decision makers with information and maps for damage assessment.“



Source: Consolidated user requirements, Deliverable D1.2.2

„As-is” situation:

Current workflow implemented using proprietary software

- Preprocessing of satellite images
 - Geometric transformation, reprojection
 - Cloud and cloud shadow filtering ➡ cloud mask
 - Calibration ➡ ToA reflectance calculation
 - Calculation of spectral indices ➡ NDVI, NDSI, NDWI ➡ NDxi
- Processing – thematic classification via thresholding
 - Inputs: ToA reflectance calibrated image, spectral indices, cloud mask, mask of natural waters (rivers and lakes), mask of non-eligible areas (from LPIS).
- Result: thematic image with 9 categories

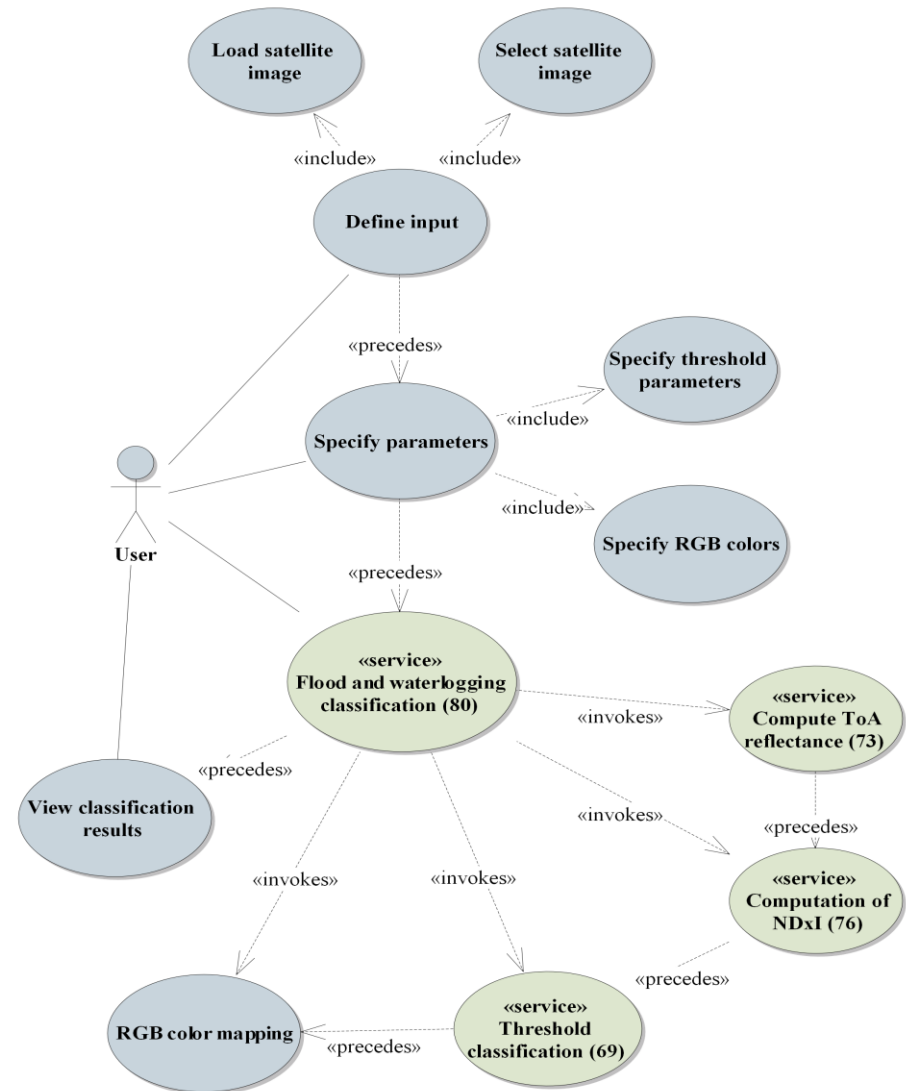
Row	Histogram	Class_Names	Color
0	19912179		
1	0	Natural waters	
2	806203	waterlog	
3	141084	seriously affected soil	
4	536413	moderately affected soil	
5	475654	weakly affected soil	
6	1205572	vegetation in waterlog	
7	22697039	dry areas	
9	0	clouds	
11	17325767	not supported areas	

Motivations:

- *Operational cooperation with civil protection and water authorities*
 - *EU-wide mapping of water bodies from HR satellite imagery*
 - *country-wide delineation of lakes from orthophotos*
-
- Faster, more automated processing to enable more iterations to improve results
 - Processing on huge areas and/or high-resolution data
 - Better use of available processing power and human resources
 - More interactivity to determine appropriate parameters (thresholds)
 - Smarter algorithms to extract relevant information more efficiently

Services developed in cooperation with ELTE-IK, based on the **AEGIS open-source spatio-temporal framework** developed there

- **Preprocessing of raster data:** Calculates TOA reflectance. Automatic parameter setting for SPOT-5 based on metadata parser
- **Calculation of Spectral Indices:** to calculate NDxI indices for subsequent classification. Band combinations are set automatically based on image metadata
- **Thematic classification of raster data:** is used for the creation of thematic rasters with a pre-defined set of categories, based on thresholds
- **Flood and waterlogging classification:** complex wrapper service that enables workflow nesting



- High-level DSL:

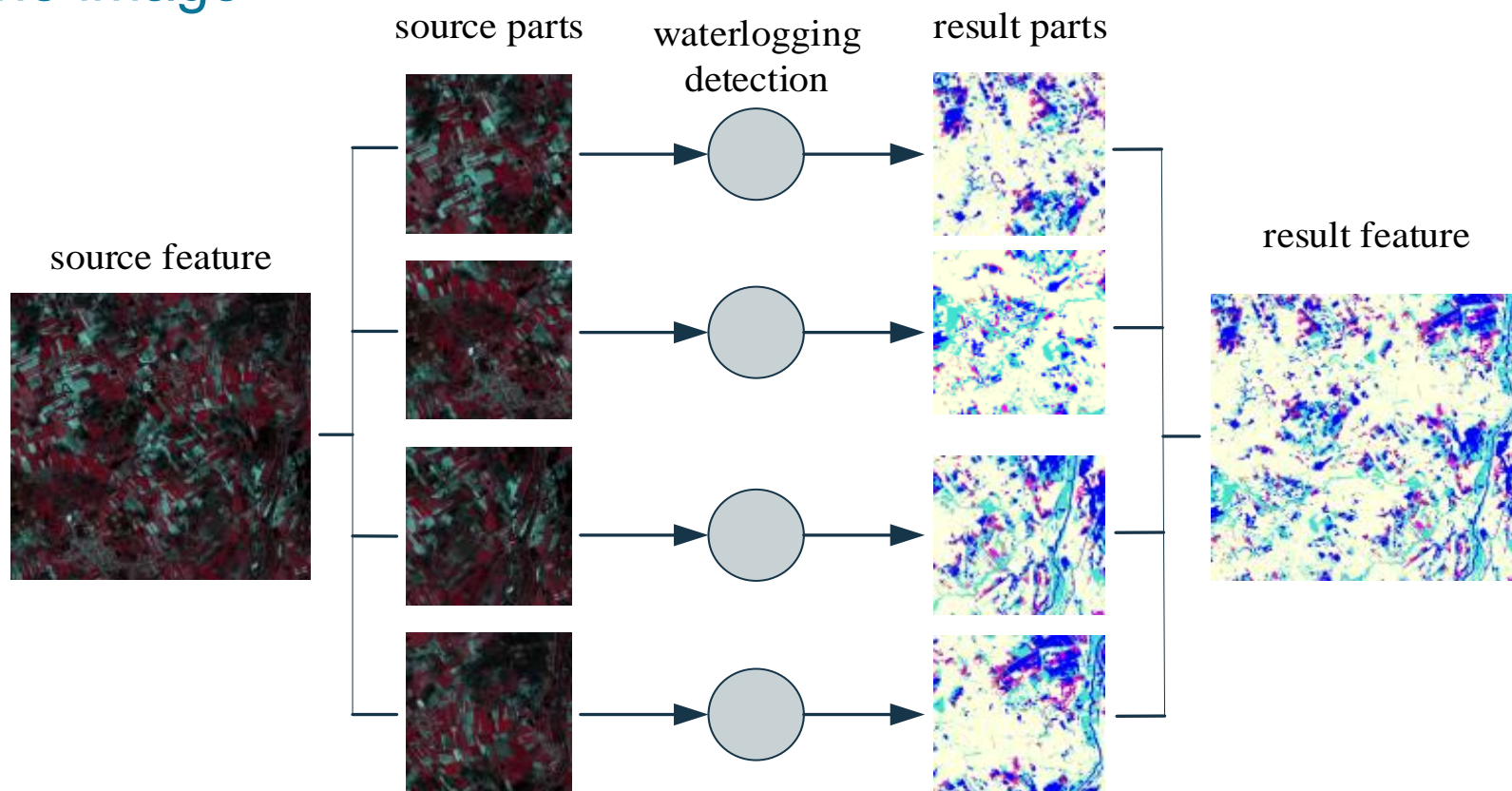
```
with [Area] of recent SatelliteImage do
  generate ThematicMap
  colorize Waterlog [WaterlogColor]
  visualize
end
```

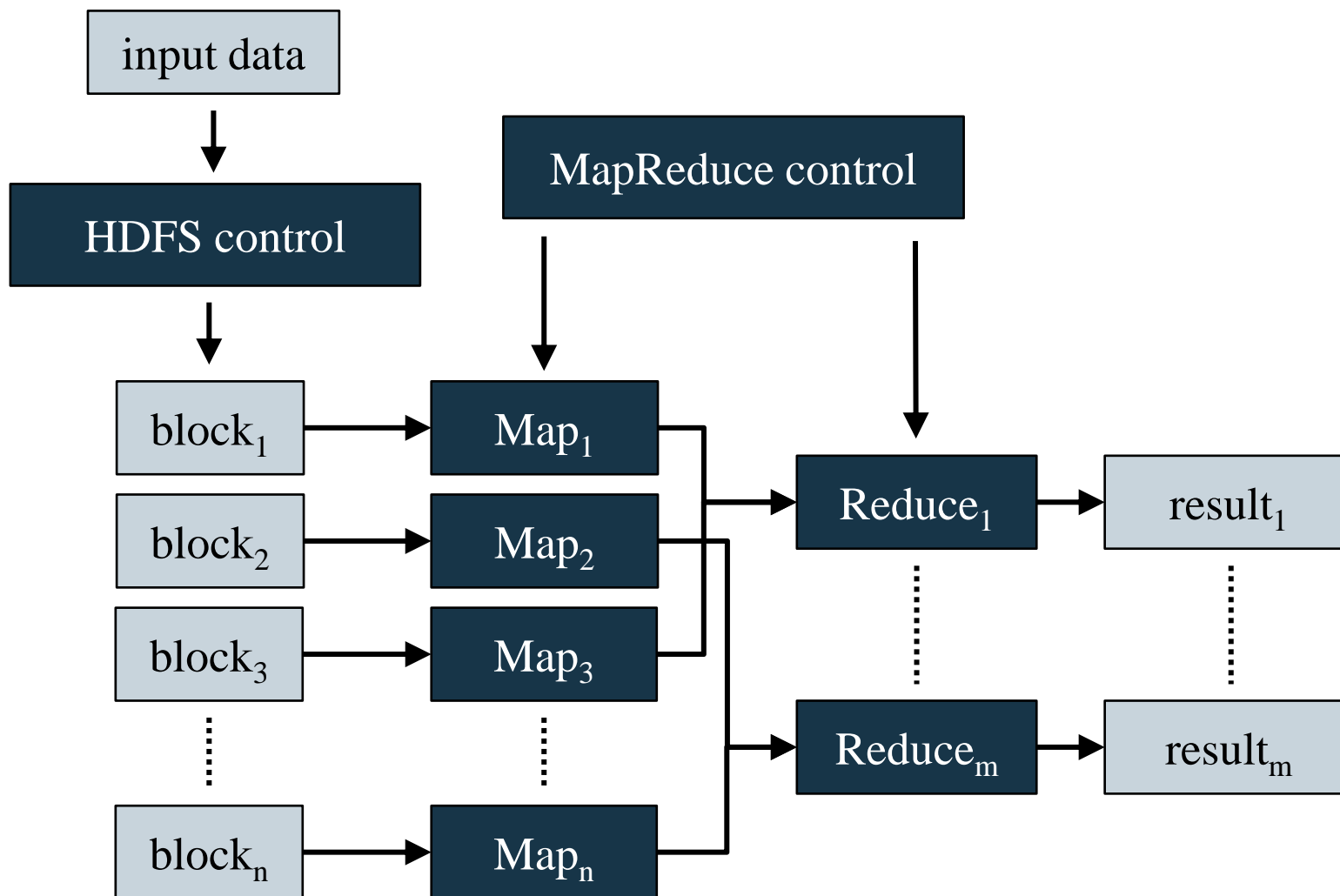
- Low-level DSL:

```
with [Area] of recent SatelliteImage do
  apply RasterDataPreprocessing
  apply SpectralIndicesComputation
    using si: [SpectralIndices] and wl: [WaveLengthInfo]
    as indices
  apply ThematicClassification
    using indices: indices
  apply Coloring
    using category: Waterlog and color: [WaterlogColor]
  visualize
end
```

- Parsed and transformed into XML for process chaining

- The detection process can be performed pixel-wise, thus the operation can be performed in parallel on any part of the image





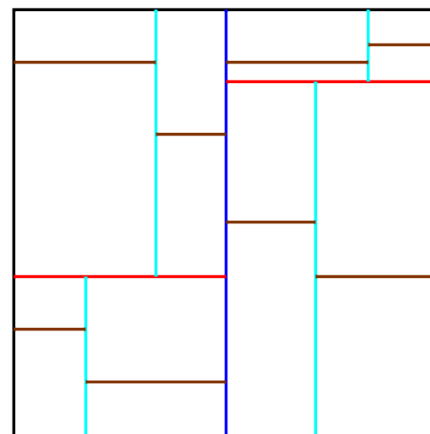
- An architecture to accommodate services of heterogeneous origins and development background
- Domain-Specific Languages (DSLs) to provide users with efficient means of workflow formulation
- Both generic and specific processing services to enable distributed processing and to fulfill user requirements in selected use cases
- New solutions based on a distributed, service-oriented processing model already enable faster processing in those cases





- Evaluate the integrated prototype
- Optimize performance by architectural and algorithmic enhancements (e.g. parallelization)
- Provide smarter solutions in use cases already covered
- Cover user needs in more and more use cases
 - Develop additional processing services
 - Construct workflows in cases already covered
- Data partitioning (see next slide)

- Partitioning, storage and indexing of geospatial data
 - Spatio-temporal characteristics to be preserved – crucial in analysis, e.g. feature extraction
 - Binary formats become unreadable with default HDFS partitioning
 - Different analysis techniques require different partitioning strategies
 - HDFS is not tailored for this
 - Current workaround: quad-tree pre-partitioning and indexing by a custom utility respecting HDFS cluster size
 - To be further investigated!





Thank you! Questions?

Dániel Kristóf

kristof.daniel@fomi.hu

www.fomi.hu

www.iqmulus.eu

Join us on LinkedIn – look for the „IQmulus” group