2nd International Conference and Exhibition on

Satellite & Space Missions

July 21-23, 2016 Berlin, Germany

Flexible Description of Earth Observation Data Processing

CGIS
Computer Graphics and Interactive Systems

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

Dorian Gorgan
Computer Science Department,
Technical University of Cluj-Napoca
dorian.gorgan@cgis.utcluj.ro

Overview

- Data, knowledge, information
- Spatial data processing
- Flexible description – WorDeL Language
- BigEarth Platform
- BigEarth based applications

CGIS Research Team

- Computer Graphics and Interactive Systems Research Laboratory (http://cgis.utcluj.ro)
- Computer Science Department
- Technical University of Cluj-Napoca (http://www.utcluj.ro)

CGIS
Computer Graphics and Interactive Systems

MINISTRY OF EDUCATION AND RESEARCH
TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

- CGIS expertise in:
  - computer graphics, user interaction techniques, graphics modelling and simulation
  - distributed interactive application development methodologies, Cloud, Grid, and GPU applications development, and HPC based processing and visualization
  - Earth Observation and Earth Science applications development

Achievements

- Projects

  IASON, enviroGRIDS, mEducator, SEE-GRID-CSI

  BIGEARTH, PECSA, GiSHEO, MedioGRID

  COST projects: KEYSTONE, ComplexHPC

- Software tools and platforms

  BIGEARTH platform, WorDeL language, gSWAT platform and application, gSWATSim, GreenLand, ESIP, gProcess, eGLE, GreenView, eTrace







BigEarth Project



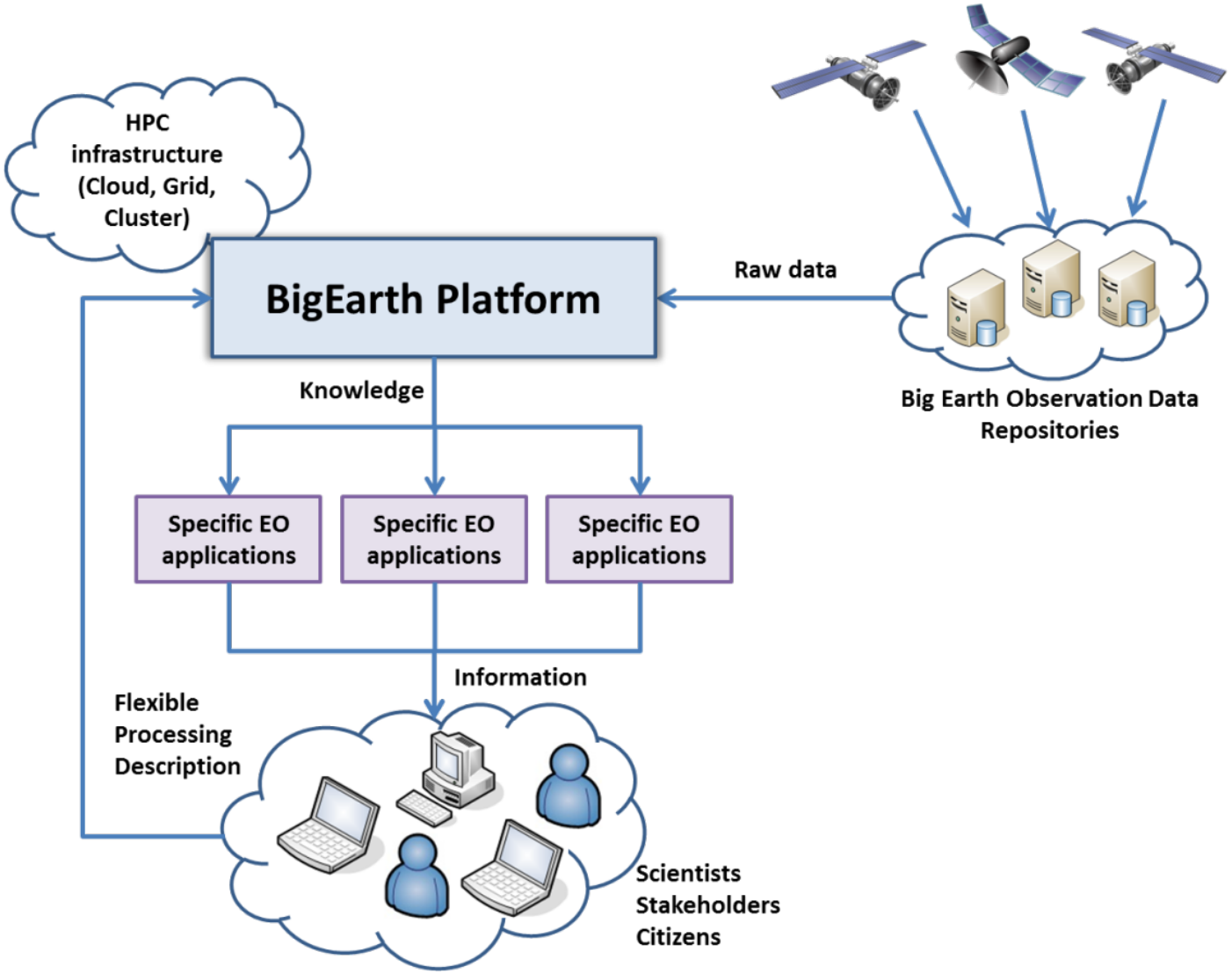
- Funded by ROSA (Romanian Space Agency) and ESA (European SA), http://cgis.utcluj.ro/bigearth/
- Description
  - Methodologies to develop and execute analytics on Big Earth Data
  - Provide flexible and interactive description of HPC processing
  - Adaptive HPC based computation
- Focus on
  - Access massive data, knowledge, and information
  - User access to simple and complex processing algorithms
  - Processing scheduling and execution performance
- HPC platforms
  - Cloud, GPU Clusters



Spatial Data and Big Data

- Huge data globally available
  - GB -> TB -> ZB -> … 2020, 40 ZB (~14 x 2012)
- 3% marked/annotated, 0.5% analyzed
- Big Data – volume, variety, velocity, variability, validity
- Earth Observation Data (EO Data)
- High costs of data management
- Increase data value by using instead of just storing
- Acquisition -> Encoding -> Use and Publish
- Data -> Knowledge -> Information (-> data -> information…)
- High Performance Computation resources + Analytics

Raw Data, Knowledge, Information

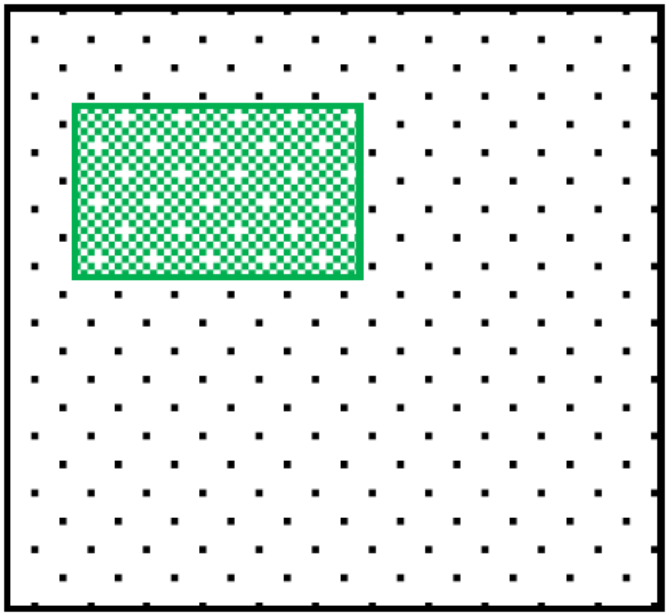


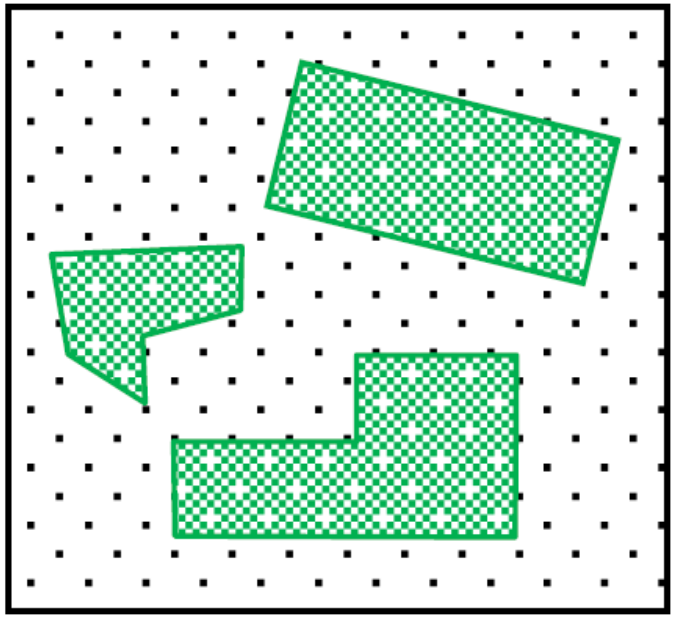
Satellite Image Data Processing Types

- Custom processing within the same input image
  - Earth data processing for retrieving information about specific geographic areas that are enclosed within the same image extent
  - Entire image vs. selected areas (bounding box, shapefiles)
- Indexed items processing
  - Information is defined based on structured models, such as matrices or arrays, where each element has a known position
- Unindexed items processing

Processing within Custom's Area

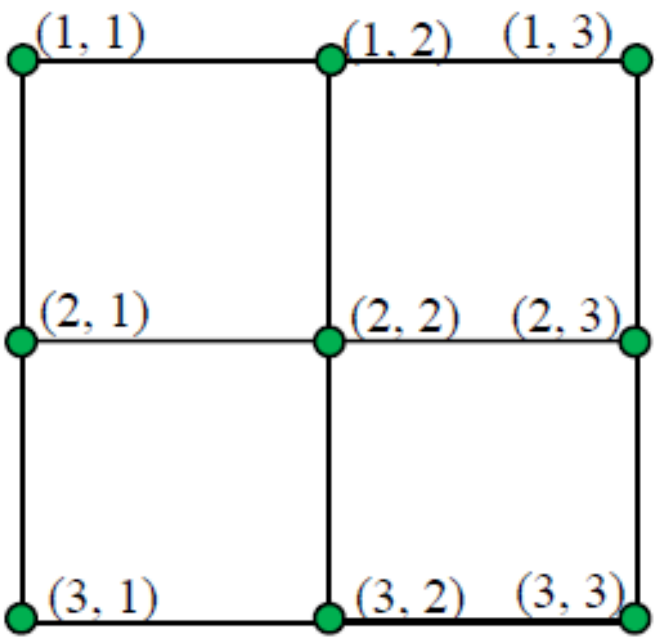
- Specify one area of interest vs. multiple areas of interest
  - E.g. Compare the vegetation growth over years in these regions, by computing the Normalized Difference Vegetation Index (NDVI) for each particular area
  - E.g. Change tracking, Urban evolution over years.

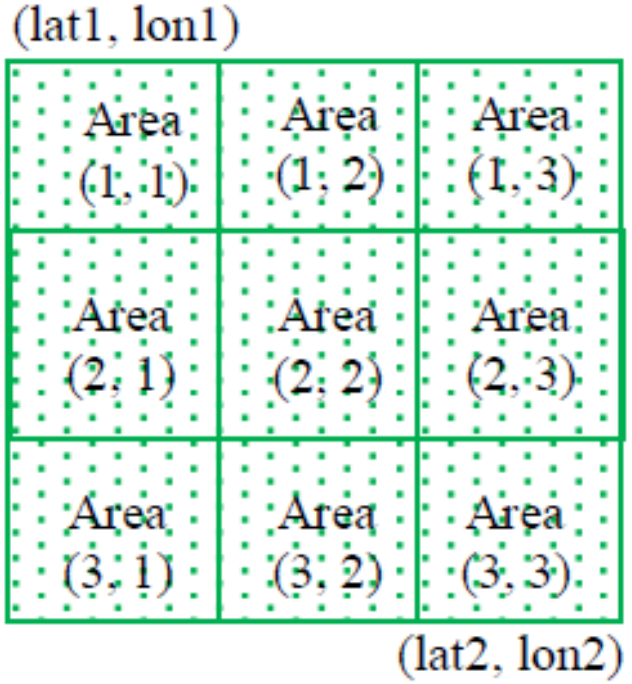




Indexed Items Processing

- Order a set of items by indexing them (e.g. list of satellite images, satellite data)
- Decompose the satellite image within equal inner areas
- Sub-areas are processed individually to extract information

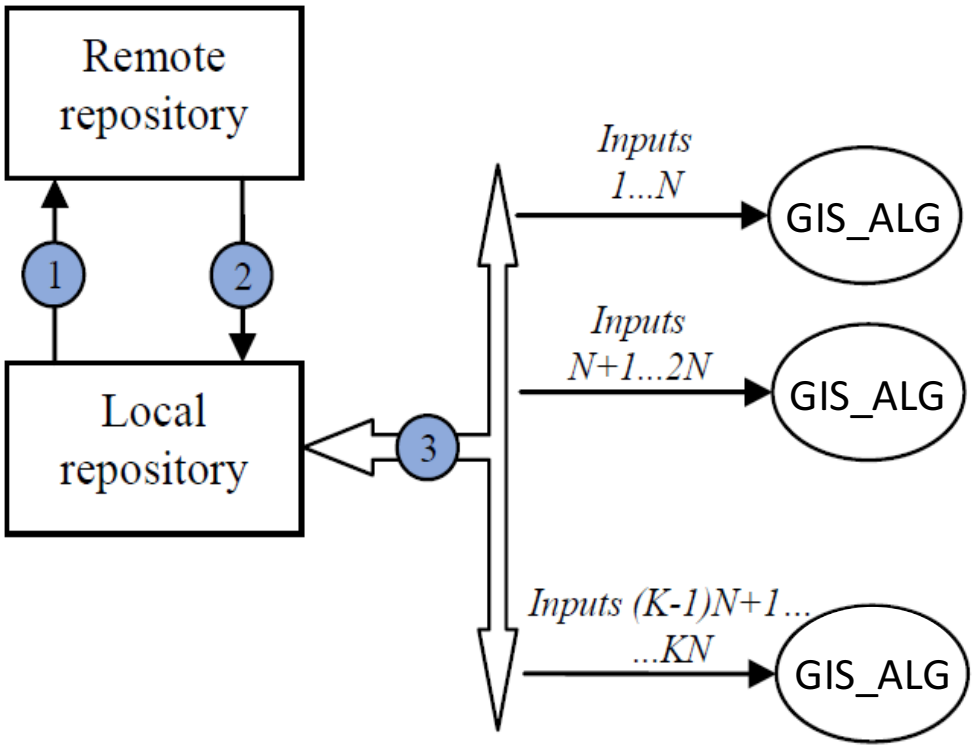




Unindexed Items Processing

- Items with unknown order are computed on the fly

  E.g. Different processing types that collect information from remote data feeders providing access to unindexed resources

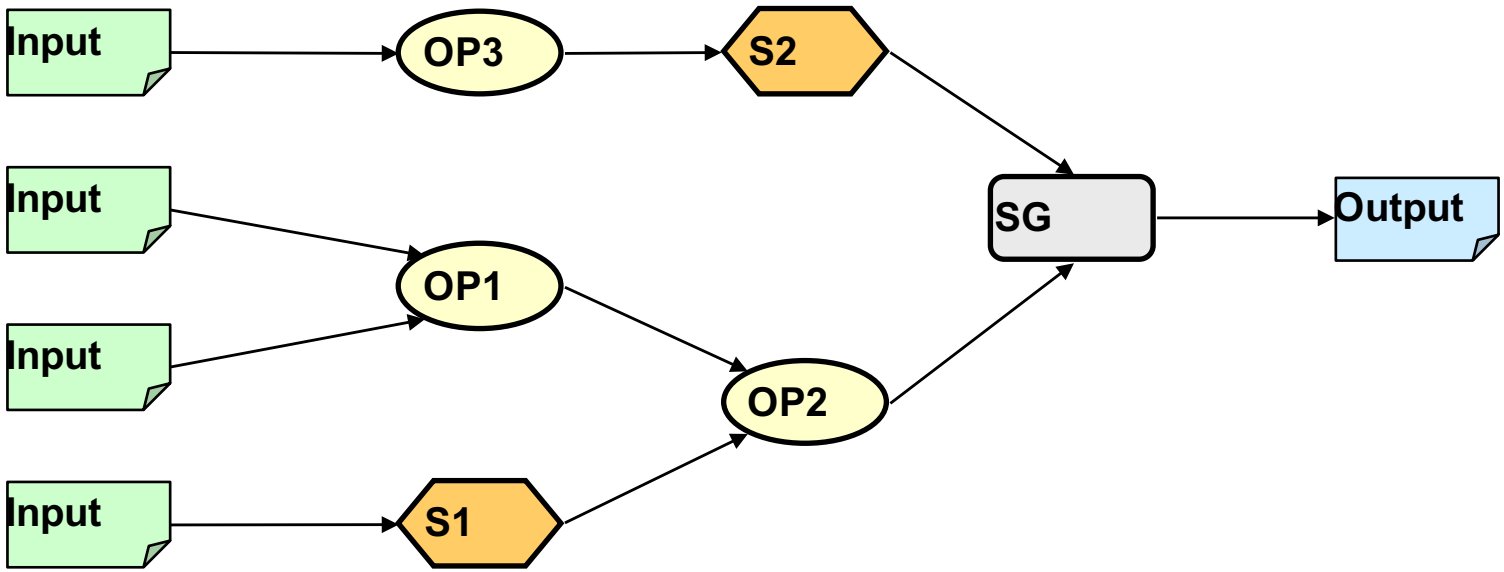


Hypergraphs – Process Description Graph

- The hypergraph is a complex graph that combines operators (OP), services (S) and subgraphs (SG). The subgraph could be a simple or complex graph.

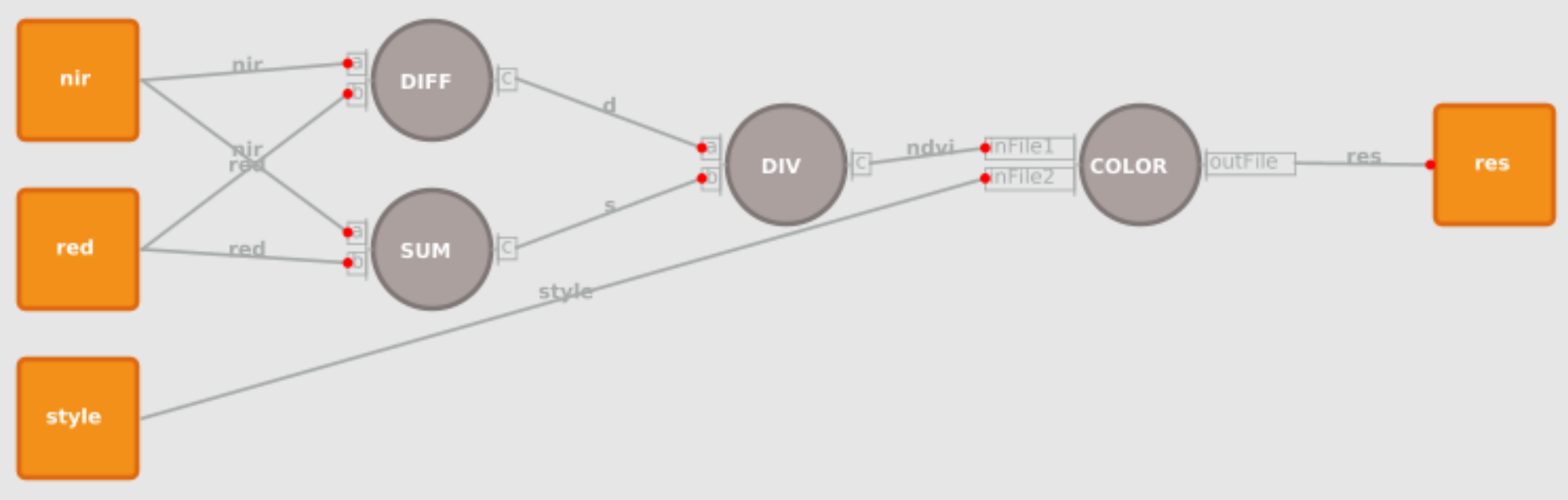
  Inputs: Spectral bands in satellite images – Landsat-8, Sentinel-2, MODIS, QuickBird, etc

  OP examples: Add, Subtract, Blur, Sharpen, EdgeDetection, HistogramEq, PseudoColoring, Erosion, etc.

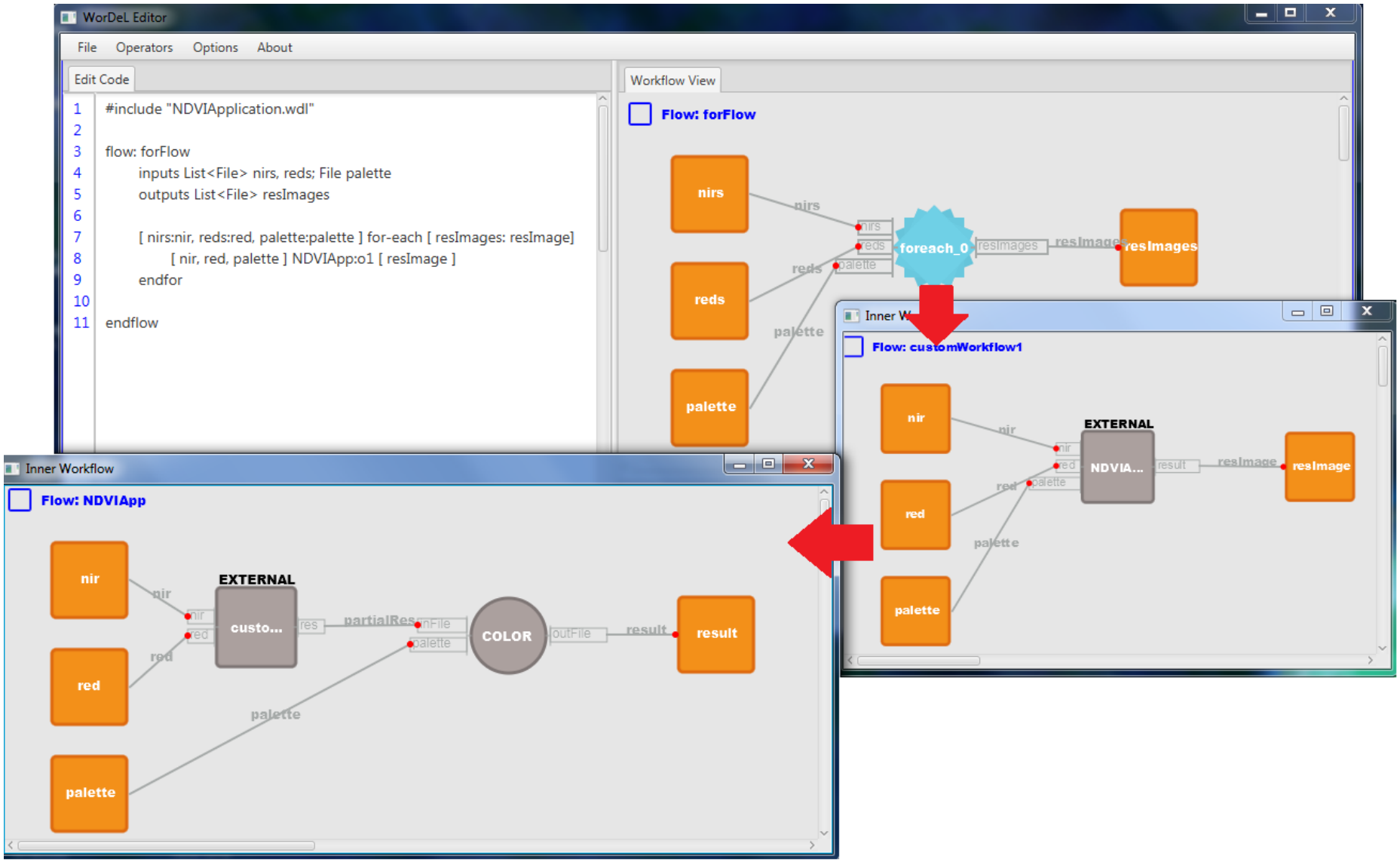


Workflow Description

- Network of interconnected operators
- Process (algorithm) description
- Hyperstructure



Workflow Description



WorDeL Language

- Workflow Description Language (WorDeL)
- Describes the processing algorithms
  - Compact format
  - Intuitive
  - Allows the identification of parallelizable algorithm sections
  - Flexibility – easy to use, increase the efficiency

WorDeL Program Structure

- Include section – includes external sections/ files
- Definition section – workflow definition
- Process section – processing tasks

```
1  include flow1.wdl
2
3  flow: flow2
4     Inputs: a:String;
5     Outputs: b:String;
6   [ a ] flow1: o1 [ c ]
7   [ c ] op1: o2 [ b ]
8  end
9
10 process: p1
11   var: inString = "test_string";
12   [ inString ] flow2: instance [ result.dat ]
13 end
```

Include section

Interface definition

Workflow body

Workflow Description section

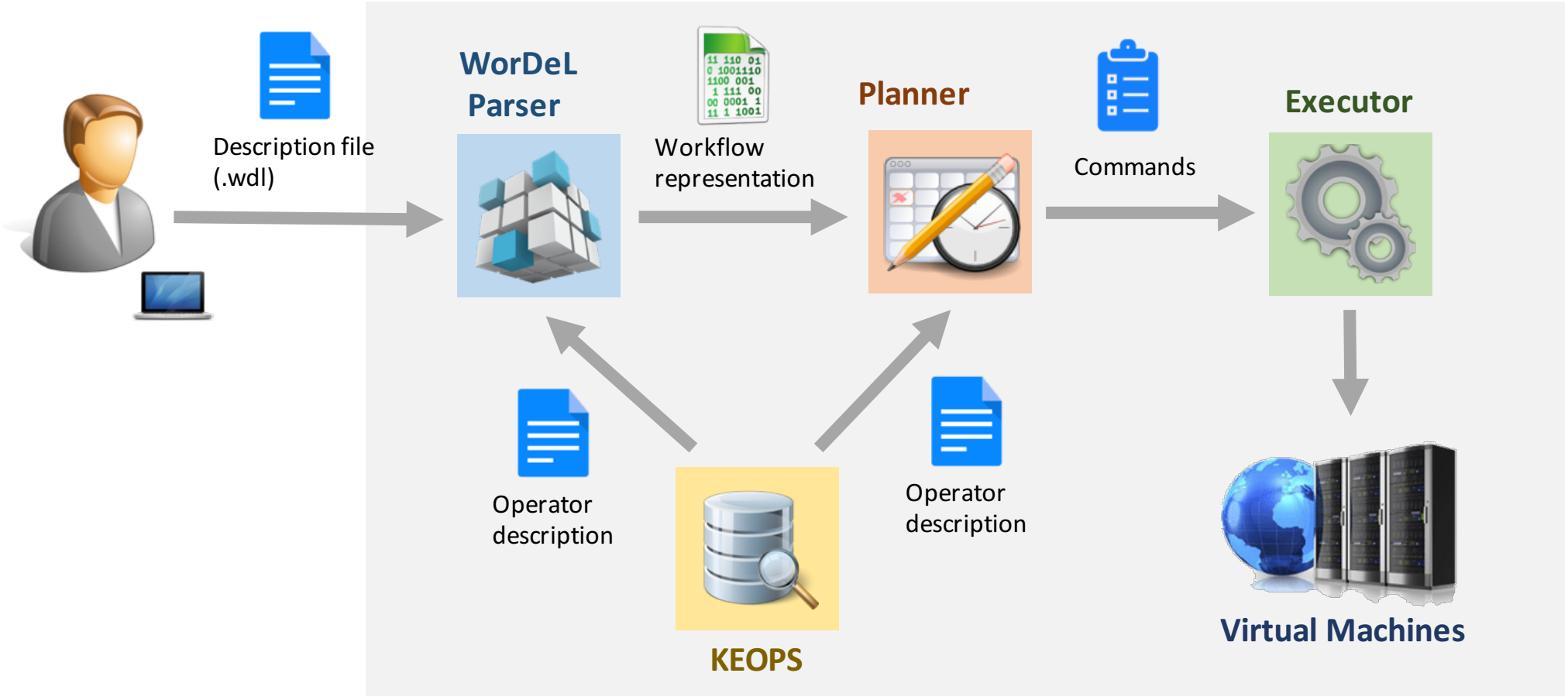
Processing section

BigEarth Platform

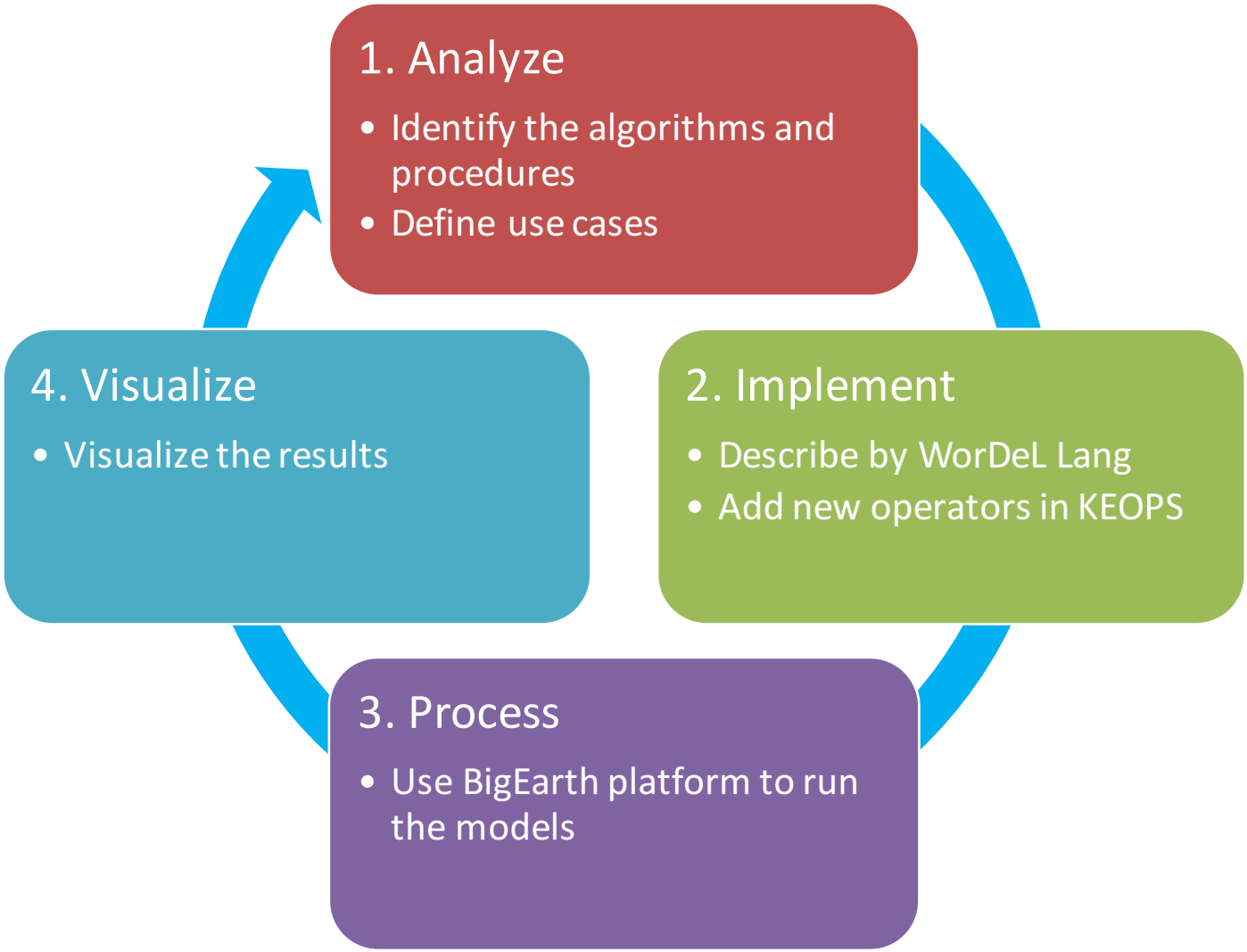
- Offered as a Virtual Machine
- Advantages:
  - No installation
  - No configuration (for the basic version)
  - Could be launched on various Cloud infrastructure (private or public)



Integration into BigEarth Platform



BigEarth Based Methodology



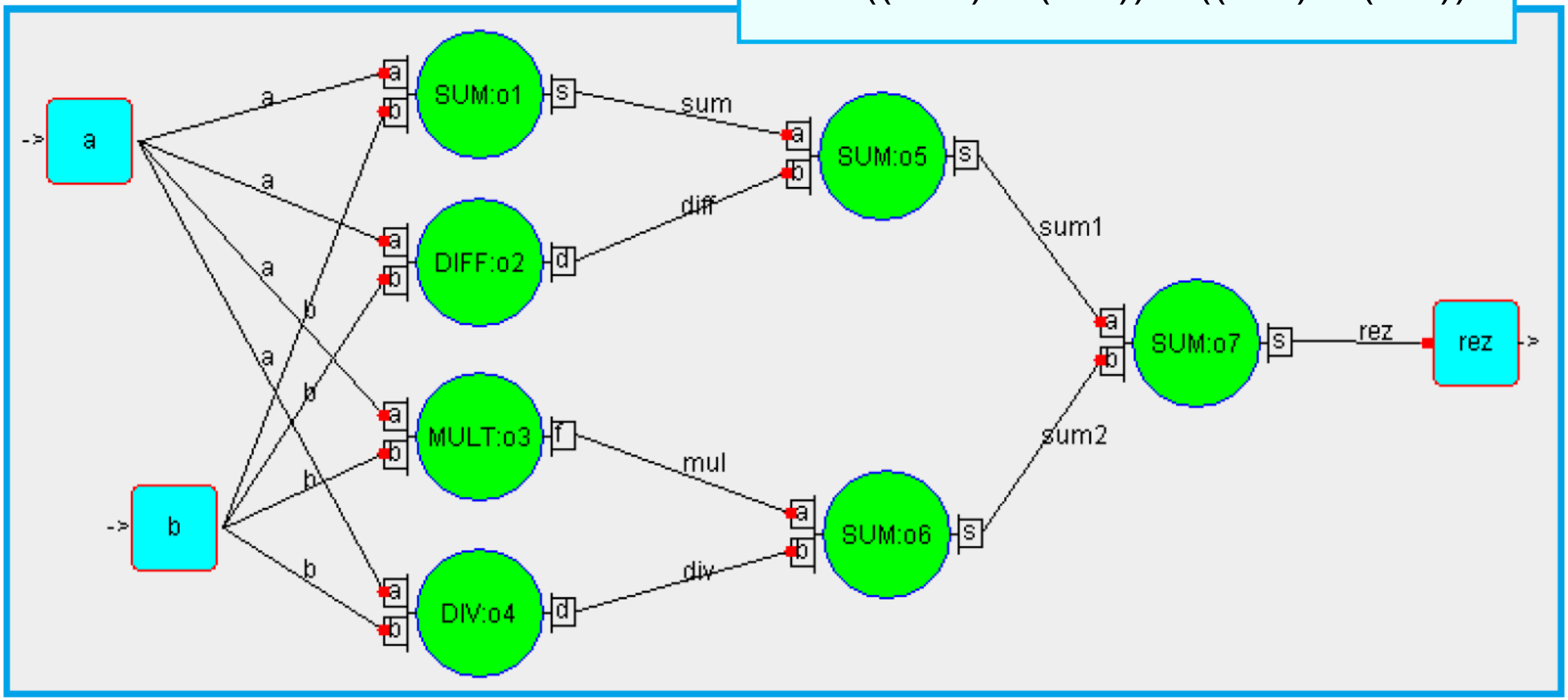
Experiments on WorDeL

- Integrating WorDeL within the BigEarth platform
- Computation of: E = (a+b) + (a-b) + a*b + a/b
- Experimental considerations:
  - a, b, E: integers
  - Computation time estimated as:

    “+” (2 sec), “–” (2 sec), “*” (3 sec), “/” (4 sec)
- Objectives:
  - Functionality
  - Computation time
  - Parallelism

Experiments: Process Workflow

E = ((a+b) + (a-b)) + ((a*b) + (a/b))

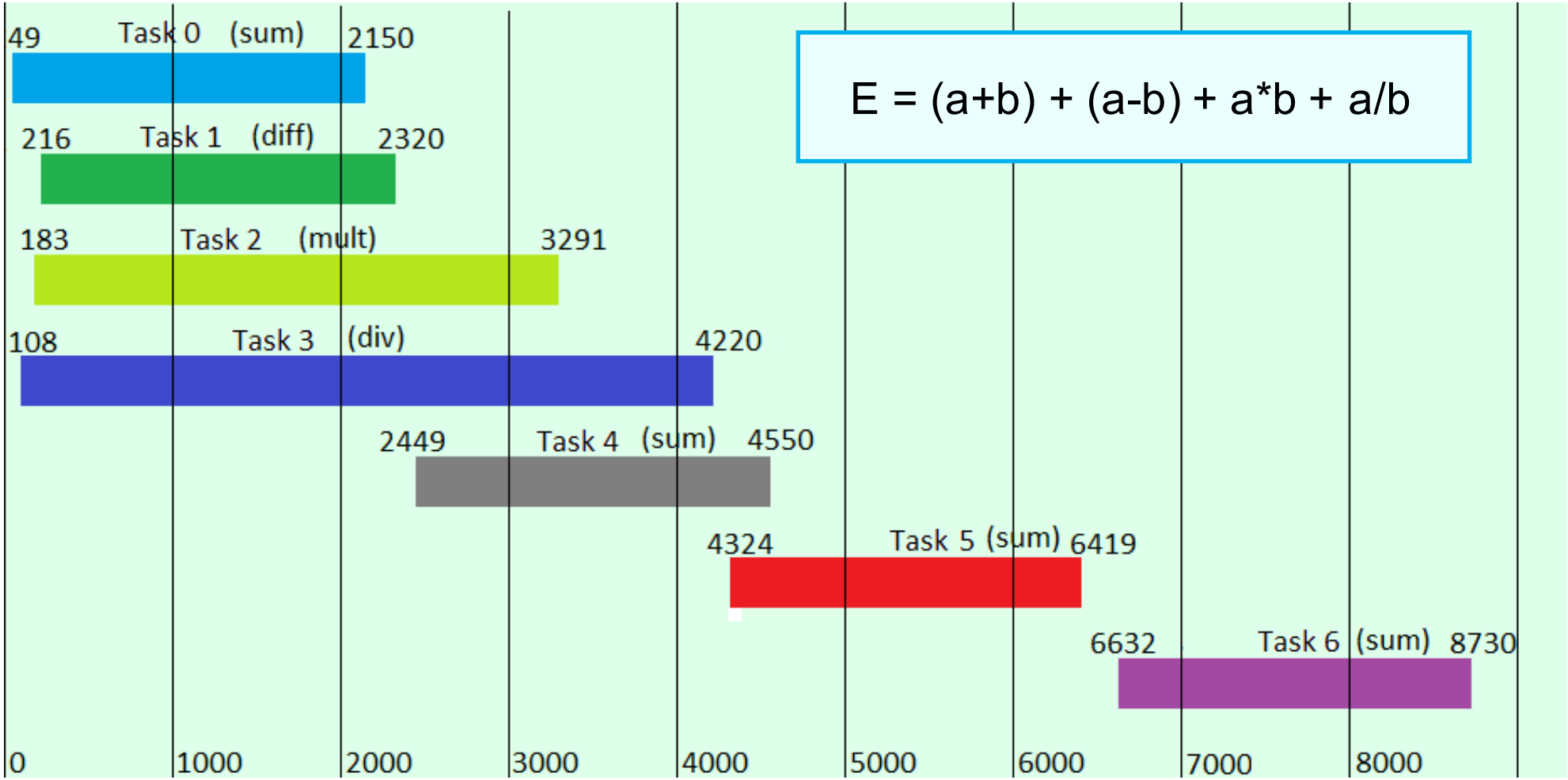


Experiments: WorDeL Description

E = (a+b) + (a-b) + a*b + a/b

```
1  flow: flow1
2    input: a,b:Integer
3    output: rez :Integer
4
5          [ a, b ] SUM : o1 [ sum ]
6          [ a, b ] DIFF : o2 [ diff ]
7          [ a, b ] MULT : o3 [ mul ]
8          [ a, b ] DIV : o4 [ div ]
9
10         [ sum, diff ] SUM : o5 [ sum1 ]
11         [ mul, div ] SUM : o6 [ sum2 ]
12         [ sum1, sum2 ] SUM : o7 [ rez ]
13 end
14 simulation: s1
15     [ valueA, valueB ] flow1 : instance1 [ result ]
16 end
```

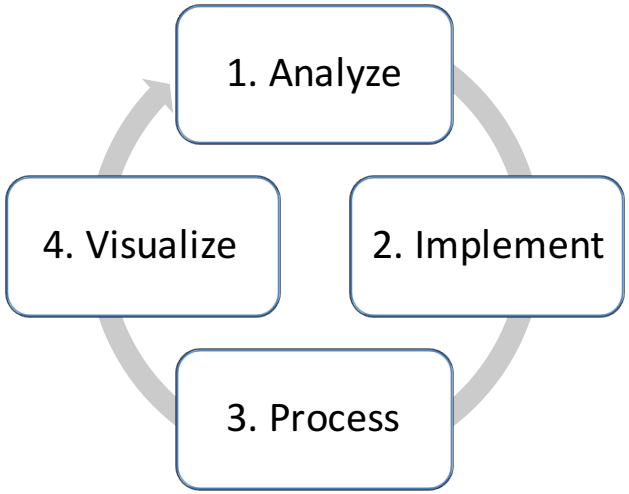
Experiments: WorDeL Description



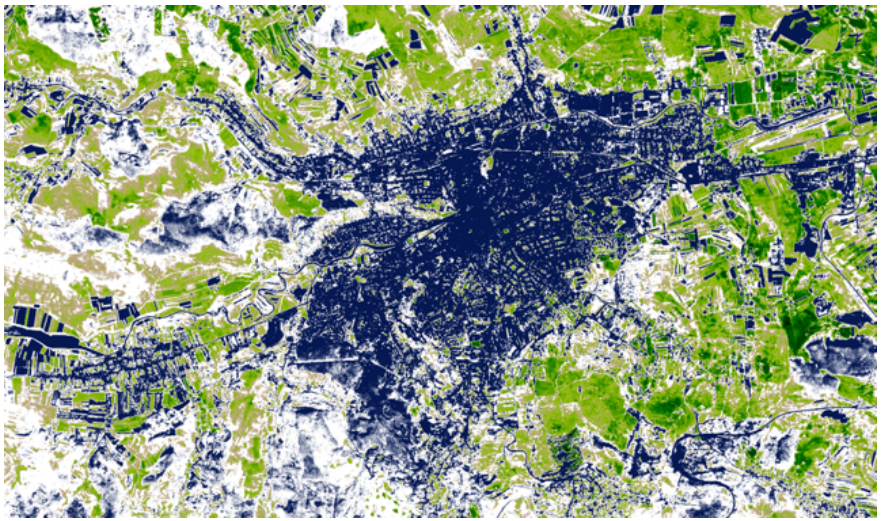
8,730 ms of the parallel execution vs 17,716 ms of the sequential execution.

Green Areas of Cities

- Identify the greenness of cities
- Solution:
  - Based on processing satellite images (e.g. Sentinel 2 products)
  - Define an NDVI processing workflow in WorDeL and apply it to various regions

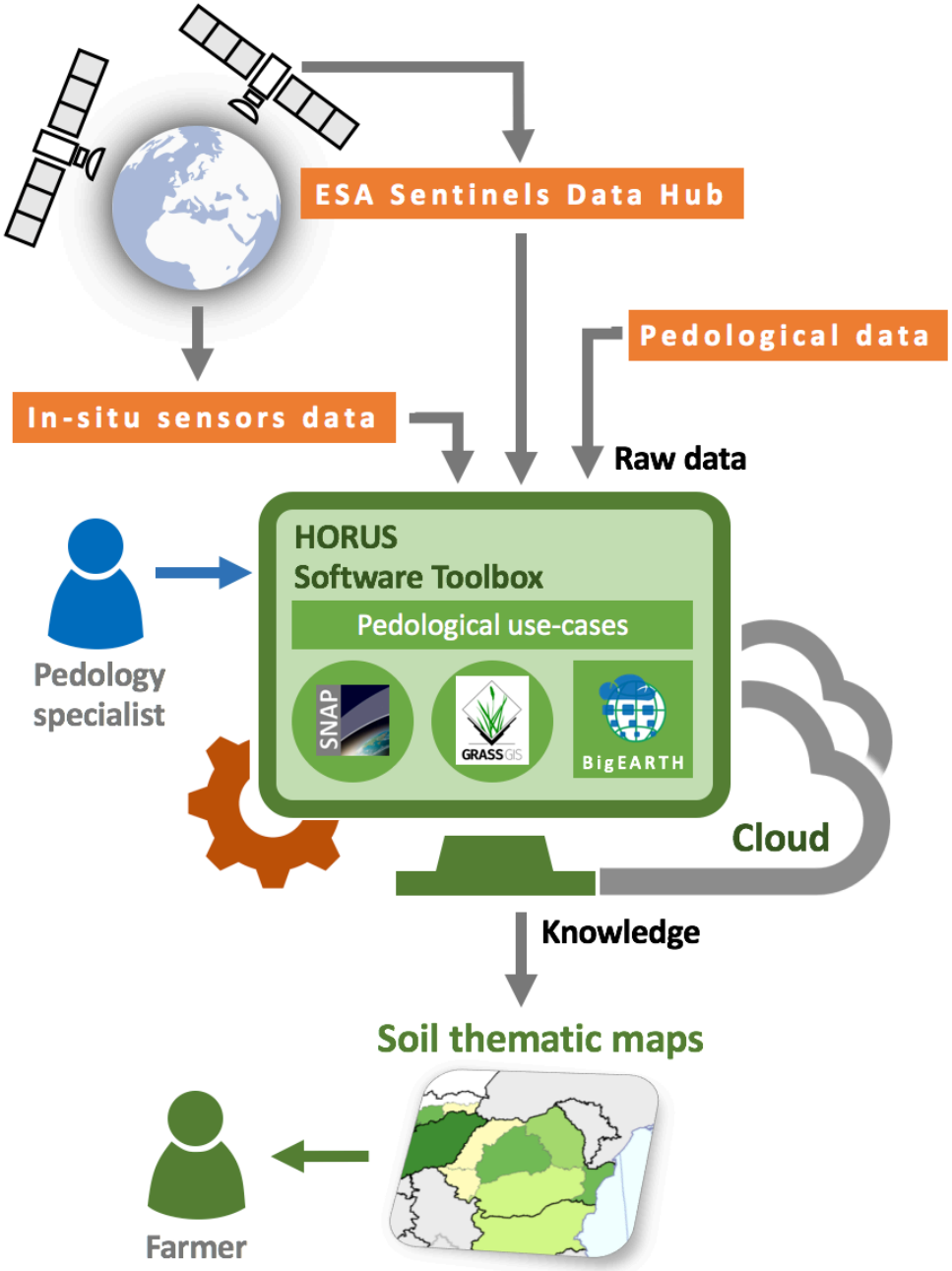






Soil Management

- Using Copernicus products in pedology studies for Transylvanian area
- Development of tools for pedology studies based on Earth Observation data
- Building a community for sustainable development



Acknowledgments

- European Space Agency

- Romanian Space Agency

- BIGEARTH Project

  http://cgis.utcluj.ro/projects/bigearth

2nd International Conference and Exhibition on

Satellite & Space Missions

July 21-23, 2016 Berlin, Germany

Many thanks for your attention!

CGIS
Computer Graphics and Interactive Systems

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

Dorian Gorgan
Computer Science Department,
Technical University of Cluj-Napoca
dorian.gorgan@cgis.utcluj.ro