Enabling Grids for E-sciencE

Challenges in Earth Science Application Development for Grid Architectures

Dorian Gorgan
Technical University of Cluj-Napoca
dorian.gorgan@cs.utcluj.ro

www.eu-egee.org

SEE-GRID-SCI
SEE-GRID eInfrastructure for regional eScience

e-infrastructure

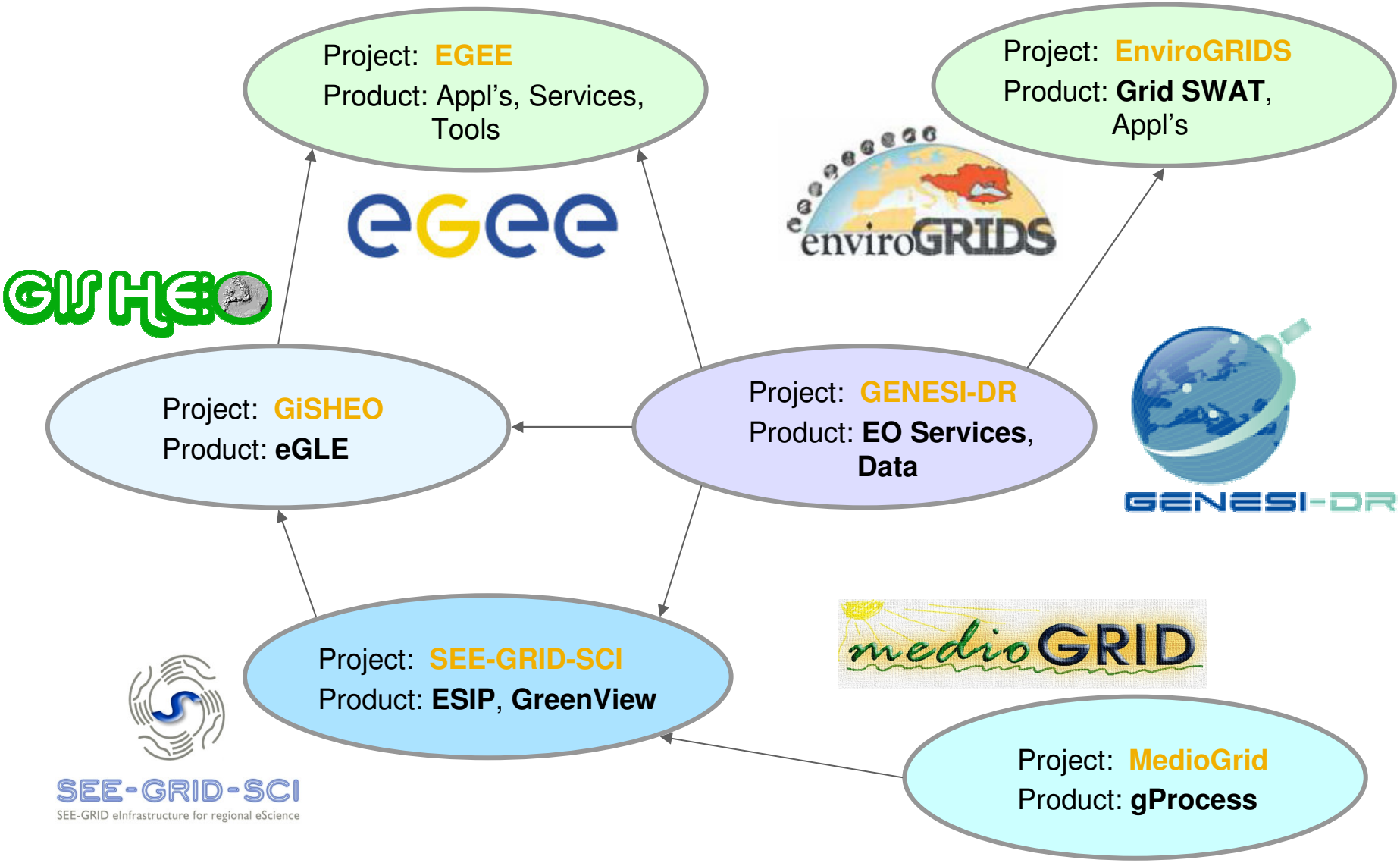
Outline

- Objectives and Issues
- SEE-GRID-SCI Project
- ESIP and gProcess platforms description
- EO Application Development Methodology
- GreenView application
- GiSHEO Project
- eGLE applications
- EnviroGRIDS Project
- Conclusions

Objectives and Issues

- Explore huge spatial data (i.e. satellite images) to supply information on the earth surface, weather, clime, geographic areas, pollution, and natural phenomena
- Support many variables based processing - satellite image type (e.g. MODIS, Landsat), geographic area, soil composition, vegetation cover, season, and context (e.g. clouds)
- Develop tools and components to support the development of Grid oriented EO (Earth Observation) applications
- Develop and experiment the EO Application Development Methodology
- Flexible description, instantiation, scheduling and optimal execution of the Grid processing
- Compatibility of EO projects: SEE-GRID-SCI, GENESI-DR, GiSHEO, EGEE, and EnviroGRIDS
- Explore and experiment data availability and accessibility to GENESI-DR from remote Grid applications in terms of *accessing policies*, *types*, *formats*, *services*, *data replication*, *transfer performance*, *granularity*, *consistency*, and *efficiency*

Earth Observation Related Projects



ESIP Platform

- **ESIP - Environment Oriented Satellite Image Processing Platform**

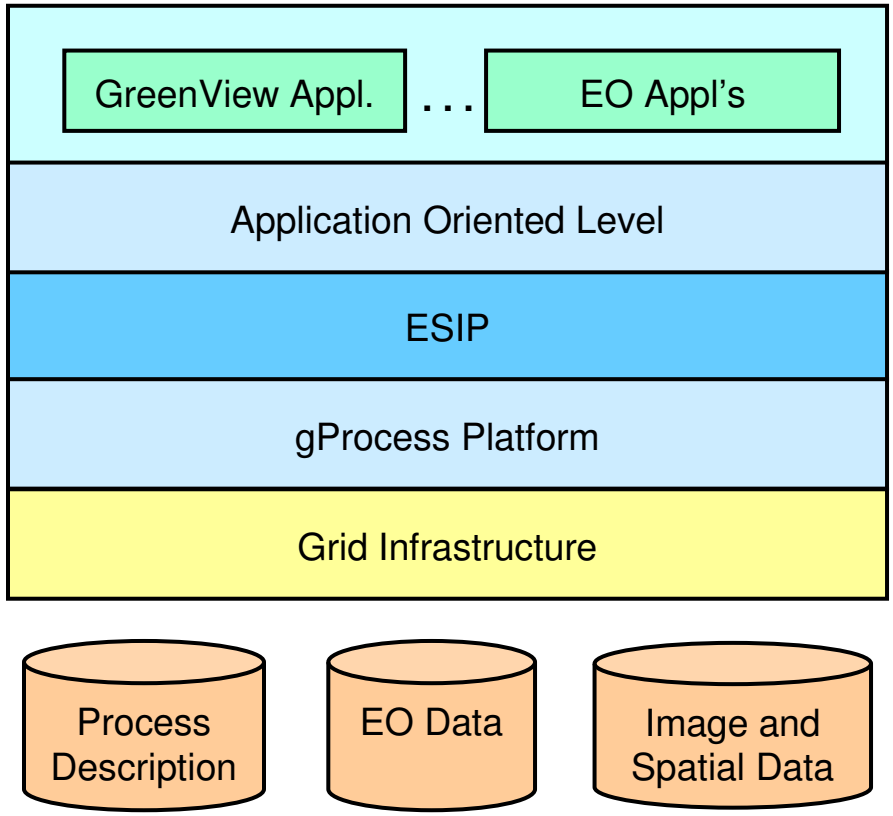
- The Environment oriented Satellite Data Processing Platform (**ESIP**) is supported through the **SEE-GRID-SCI** (**SEE-GRID eInfrastructure for regional eScience**) FP7 project (2008-2010), funded by the European Commission, http://www.see-grid-sci.eu/

- **Partners**
  - ESIP Platform development:
    1. Technical University of Cluj-Napoca (UTCN), ESIP coordinator
    2. National Center for Information Technology (NCIT) Bucharest
    3. West University of Timisoara (UVT)
    4. ICI Bucharest, national coordinator of the SEE-GRID-SCI project

ESIP Features

- **ESIP** layers on gProcess platform that is a set of satellite image oriented operators, services, and particular algorithms.
- The **gProcess** platform is a collection of Grid services and tools providing the following basic functionality:
  - Visual manipulation based interactive description of the Grid based satellite image processing by pattern workflow (i.e. DAG)
  - Development of hypergraphs as a composition of basic operators, services, and subgraphs
  - Pattern workflow instantiation for particular satellite image
  - Satellite data management, access and visualization
  - Workflow based Grid execution
  - Process execution control and visualization
  - Optimal execution for appropriate mapping of the processing over the Grid resources

Functional Levels



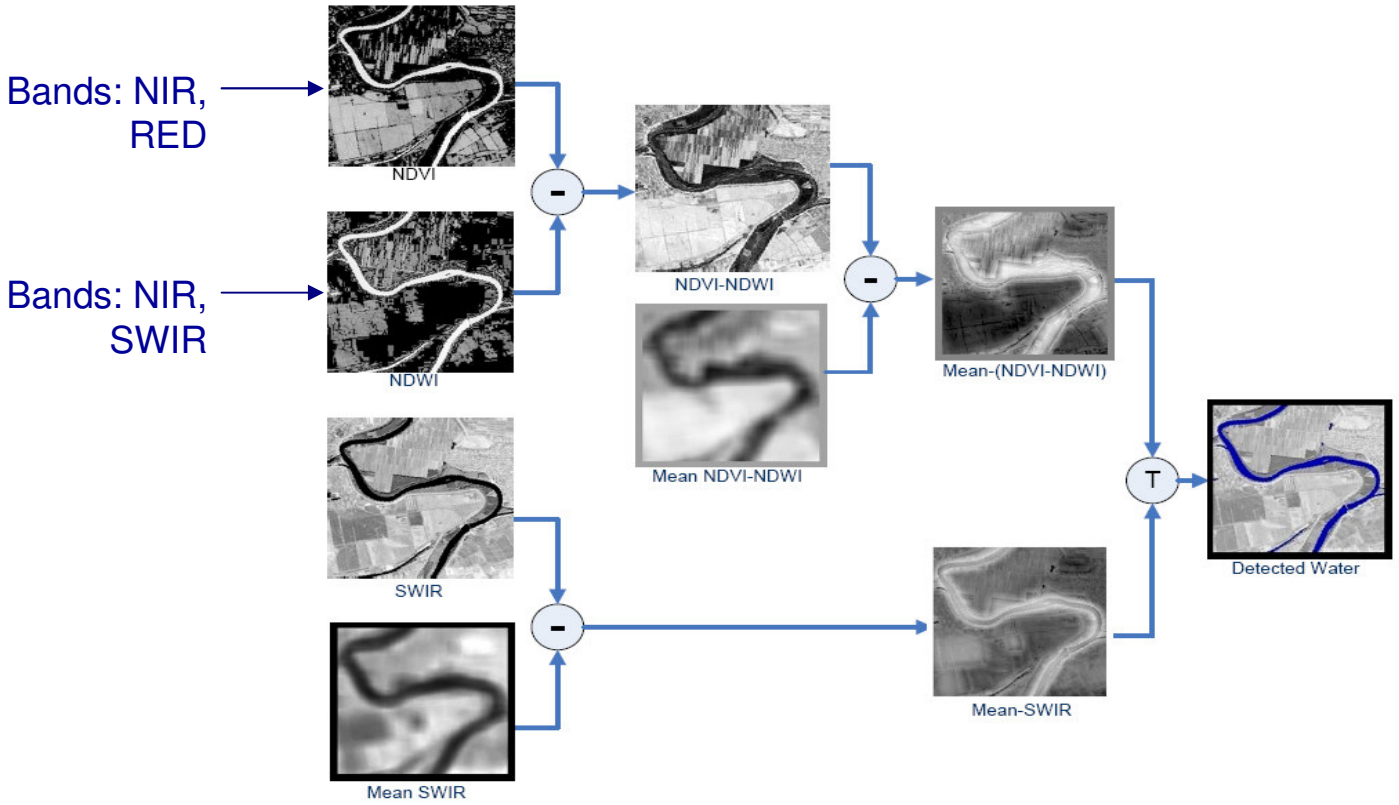
- **Data**: satellite images (e.g. MODIS, Landsat, Ikonos and QuickBird); environmental data; diagrammatic process description.
- **Grid Infrastructure**: middleware (Globus Toolkit 4, gLite).
- **gProcess Platform**: service oriented architecture; collection of services providing creation, execution, and scheduling of the jobs; job resource management; access to image database; file transfer; data replication; workflow based execution; image processing.
- **ESIP**: adds satellite image oriented operators, services, and particular algorithms.
- **Application Oriented Level**: set of tools and components that support the development and the execution of the workflow based distributed processing.
- **EO Oriented Applications**: pilot applications through which the user may access data, and particular processing and visualization.

Gond's Water Detection Algorithm

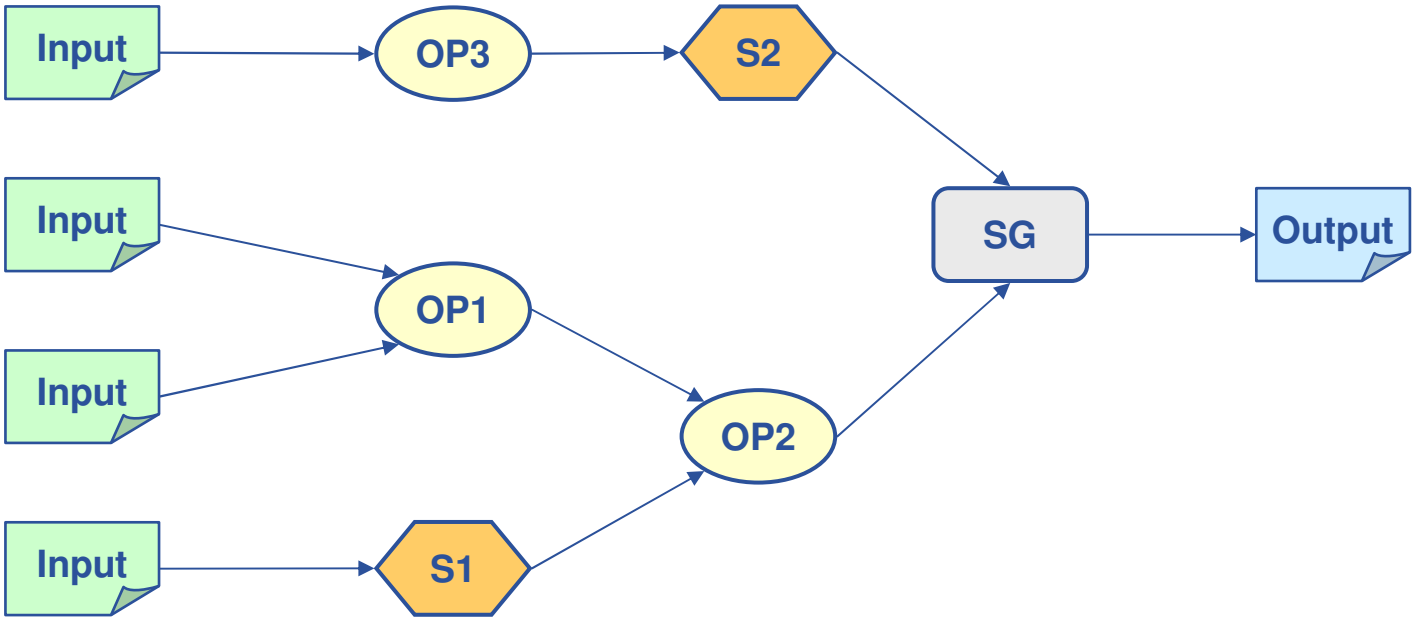
- **Gond's water detection algorithm**

  Gond V., Bartholom E., Ouattara F., Nonguierma A. and Bado I. Surveillance et cartographie des plans d'eau et des zones humides et inondables en rgions arides avec l'instrument VEGETATION embarque sur Spot 4, International Journal of Remote Sensing, 2004, 25,5. pp. 987- 1004.

- **Landsat satellite image: SWIR (5), Red (3) and NIR (4) spectral bands**



Hypergraphs

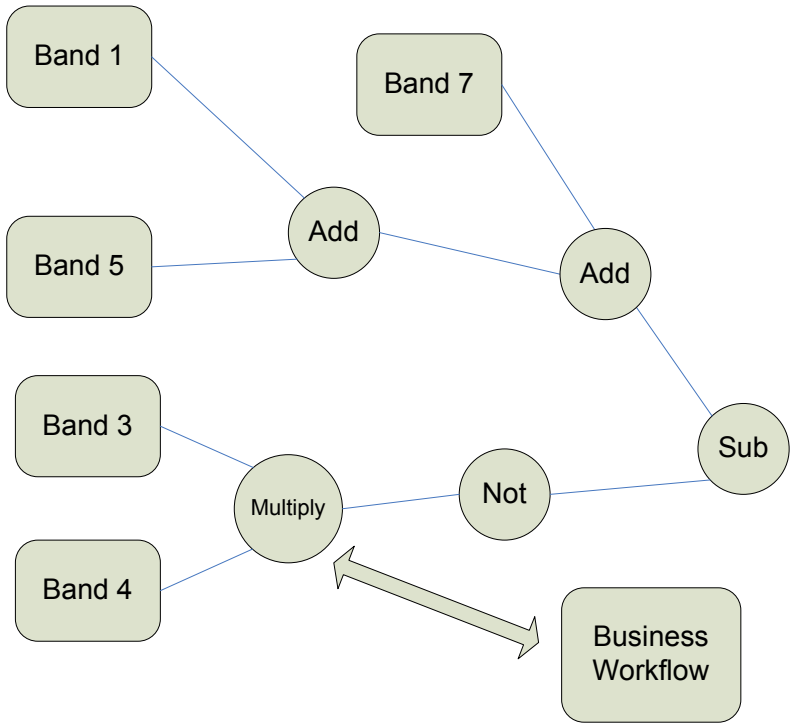


The hypergraph is a complex graph that combines operators (OP), services (S) and subgraphs (SG). The subgraph could be a simple or complex graph.

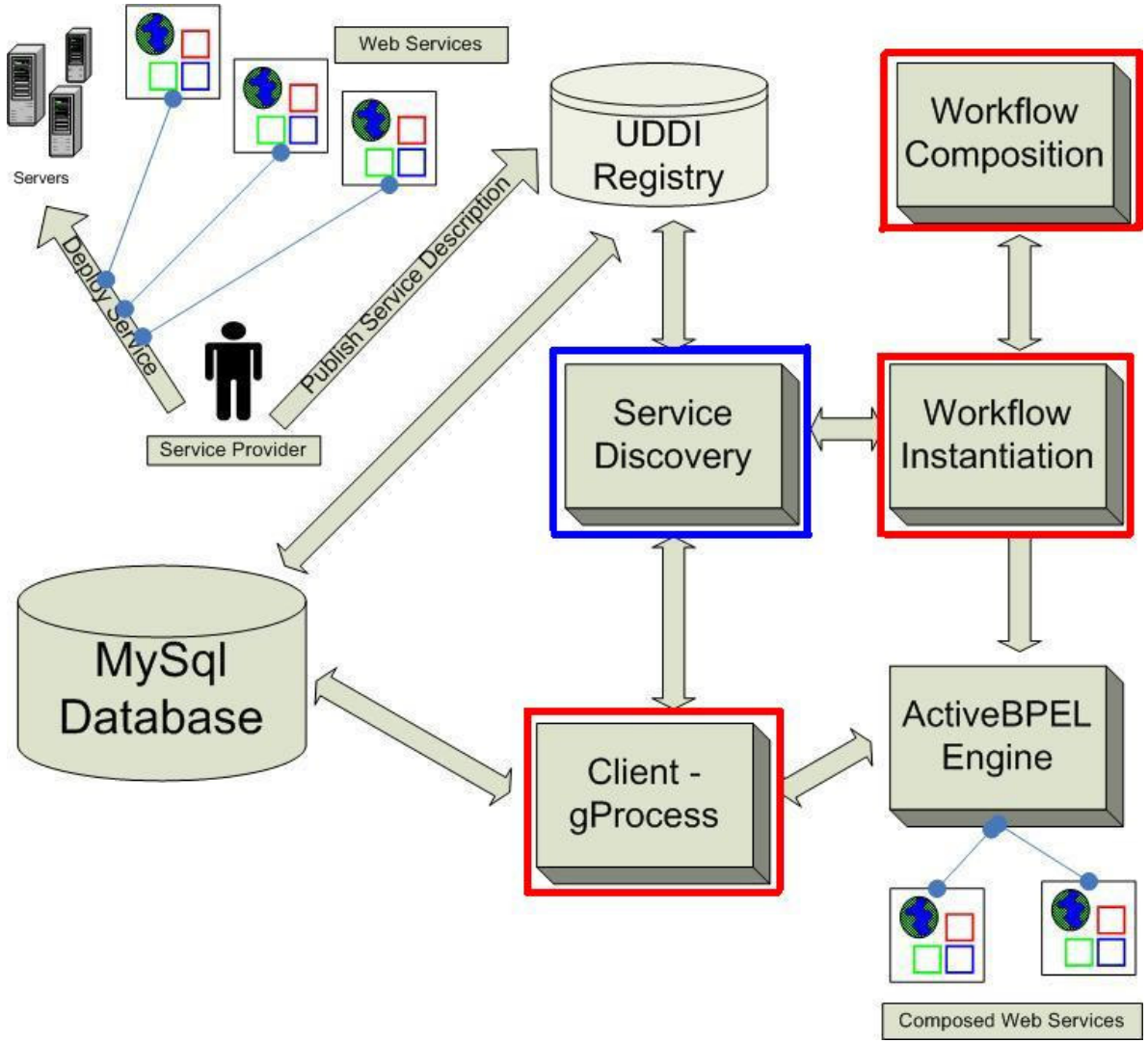
Inputs: Spectral bands in satellite images – MODIS, Landsat, QuickBird, etc
OP examples: Add, Subtract, Blur, Sharpen, EdgeDetection, HistogramEq, PseudoColoring, Erosion, etc.

Web Service Composition

- Design and implement a business workflow architectural model which allows the extension of the workflow description and execution tools in the Grid environment with Web service based computation nodes
  - Automatic discovery of Web services
  - Automatic binding of Web services
  - Semantic annotation
  - Composition of Web services based on templates
  - Automatic invocation of composed Web services from Grid environment



Service Composition Related Architecture



EO App Development Methodology

1. **Algorithm identification and analysis**
2. **Data model definition**
3. **Identify atomic parts of the computation/algorithms**
   - Parallel and serial processing
   - Atomic computation Implementation as services, procedures, distinct applications
4. **Algorithm implementation**
5. **gProcess based process description**
   - Workflow based description (i.e. PDG)
   - Workflow instantiation (i.e. iPDG)
   - Execution over Grid
6. **Develop the application interface**
   - Build up the GUI and link the interactive components to the Grid application entities

GreenView Application

- **GreenView** - Refinement of surface and vegetation parameters in SEE region based on satellite images

Research is supported by **SEE-GRID-SCI** (FP7) project, funded by the European Commission through the contract nr RI-211338.

- Investigates the change of the vegetation distribution in the Carpathian Basin and its climate-related causes. The study could be extended to study the impact of urban environment on the vegetation.

- GreenView application development partners:
  1. UTCN, NCIT, UVT, and ICI (Romania)
  2. Eötvös Loránd University (ELU) from Budapest (Hungary)
  3. Research and Educational Networking Association of Moldova (RENAM)

Acknowledgments:
- Climate change data have been retrieved from the PRUDENCE data archive, funded by the EU through contract EVK2-CT2001-00132.
- MODIS data have been produced and distributed by NASA through the EOS Data Gateway system.
- Biome-BGC version 4.1.1 was provided by the Numerical Terradynamic Simulation Group (NTSG) at the University of Montana. NTSG assumes no responsibility for the proper use of Biome-BGC by others.

GreenView Pilot Application

- Functionality:
  - Data interpolation
  - Model calibration
- Uses the ESIP Platform
- Inputs
  - satellite image types (MODIS, Landsat, Aster etc.)
  - Meteorological data
  - Field measurements

    Eddy covariance measurements for a particular geographical area

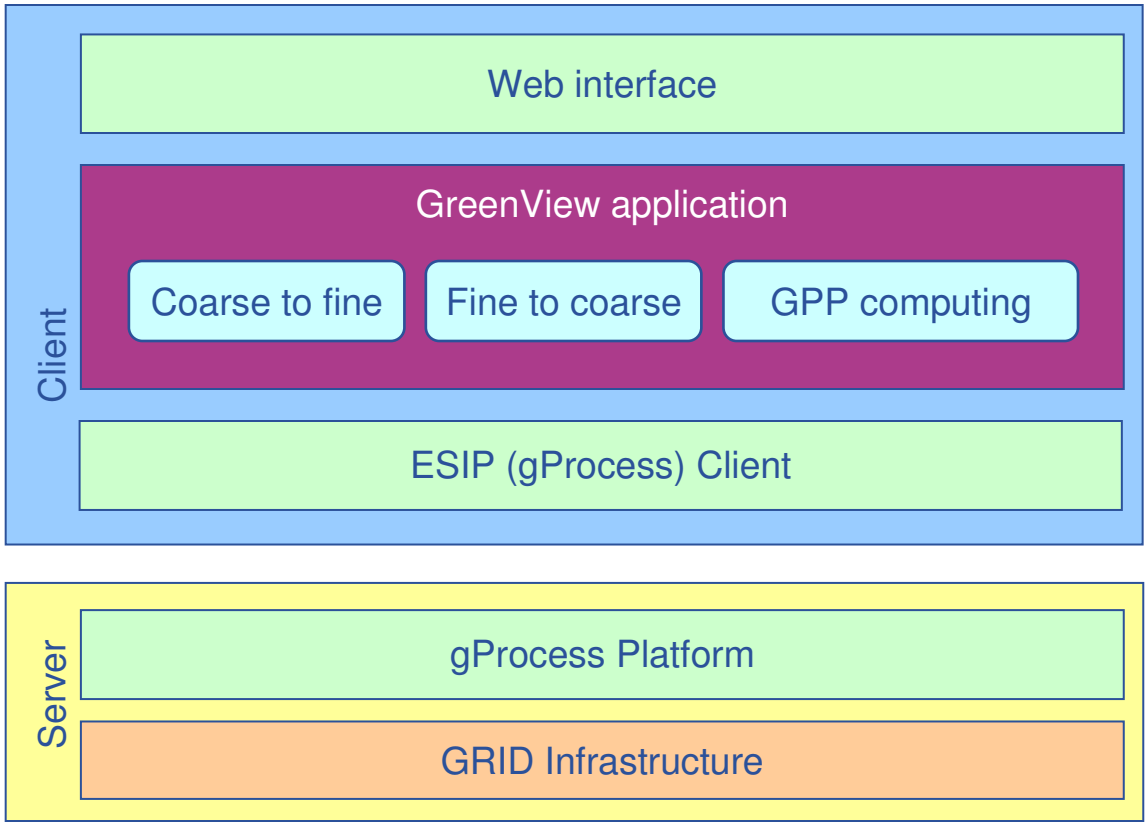
Example:

- The data model is defined by a MODIS satellite image representing a certain area of the Earth and temperature data values recorded in specific points inside the same area.
- The resolution of the MODIS image is 1Km^2 and the resolution of the temperature measurements is about 150Km^2

GreenView Application

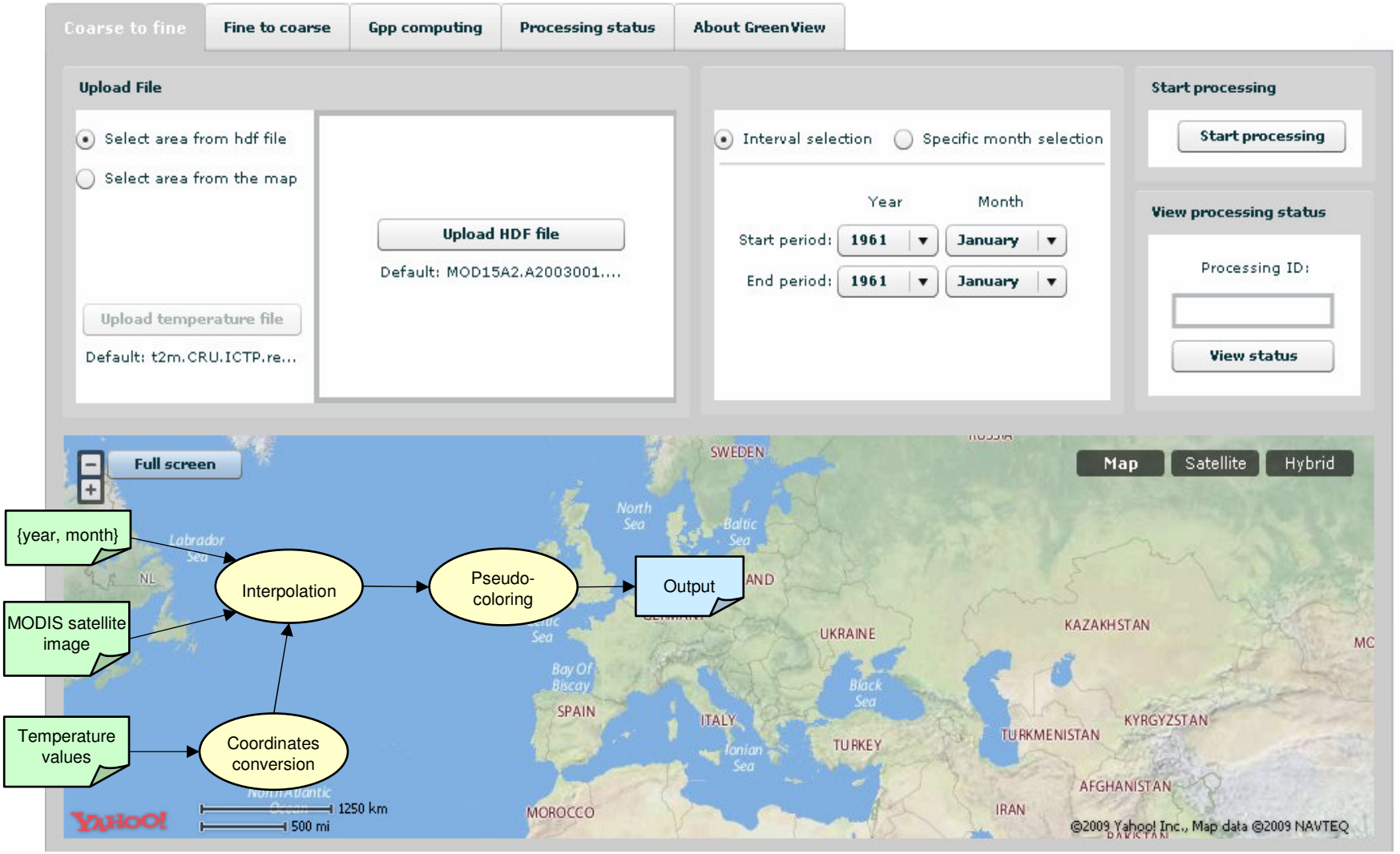
- Data interpolation
  - Computes data to an arbitrarily chosen grid defined by the user
    - fine to coarse resolution
    - coarse to fine resolution
  - Provides data series for a period chosen by the user, in time steps of the original data or averaged for a certain period
- Model calibration
  - GPP (Gross Primary Production) the total amount of carbon taken up by vegetation via photosynthesis
  - Determination of GPP
    - Field measurements (eddy covariance technique) – accurate, but limited
    - MOD 17 product (algorithm is based on the BIOME-BGC ecosystem model)
  - Calibration and sensitivity analysis – based on BIOME BGC model
    - Period: 1997–2008, Monte Carlo method
    - Requires high computation resources → Grid computation
  - Post processing
    - Nonlinear inversion method

GreenView Related Architecture

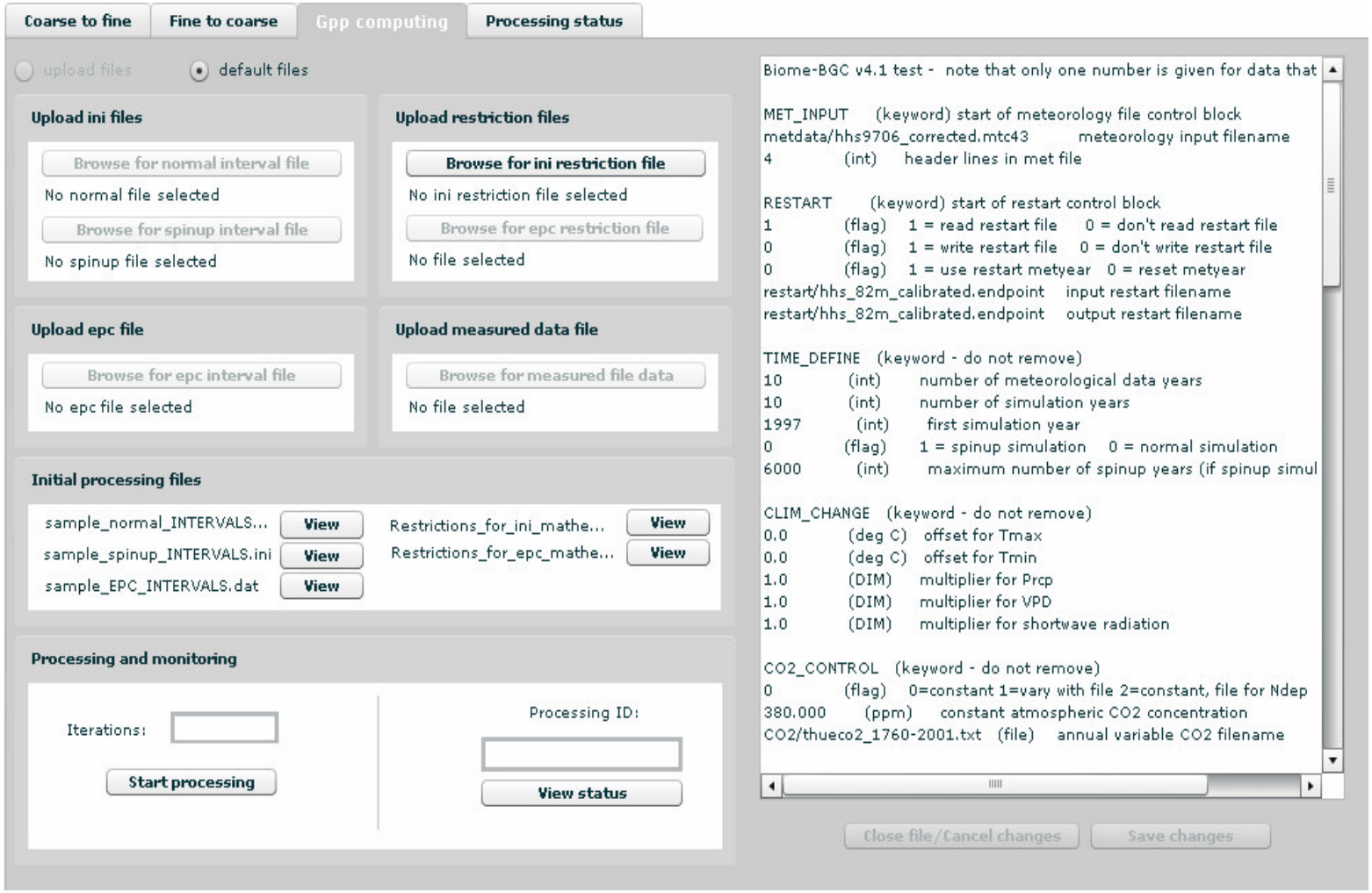


GreenView - Coarse To Fine Interpolation

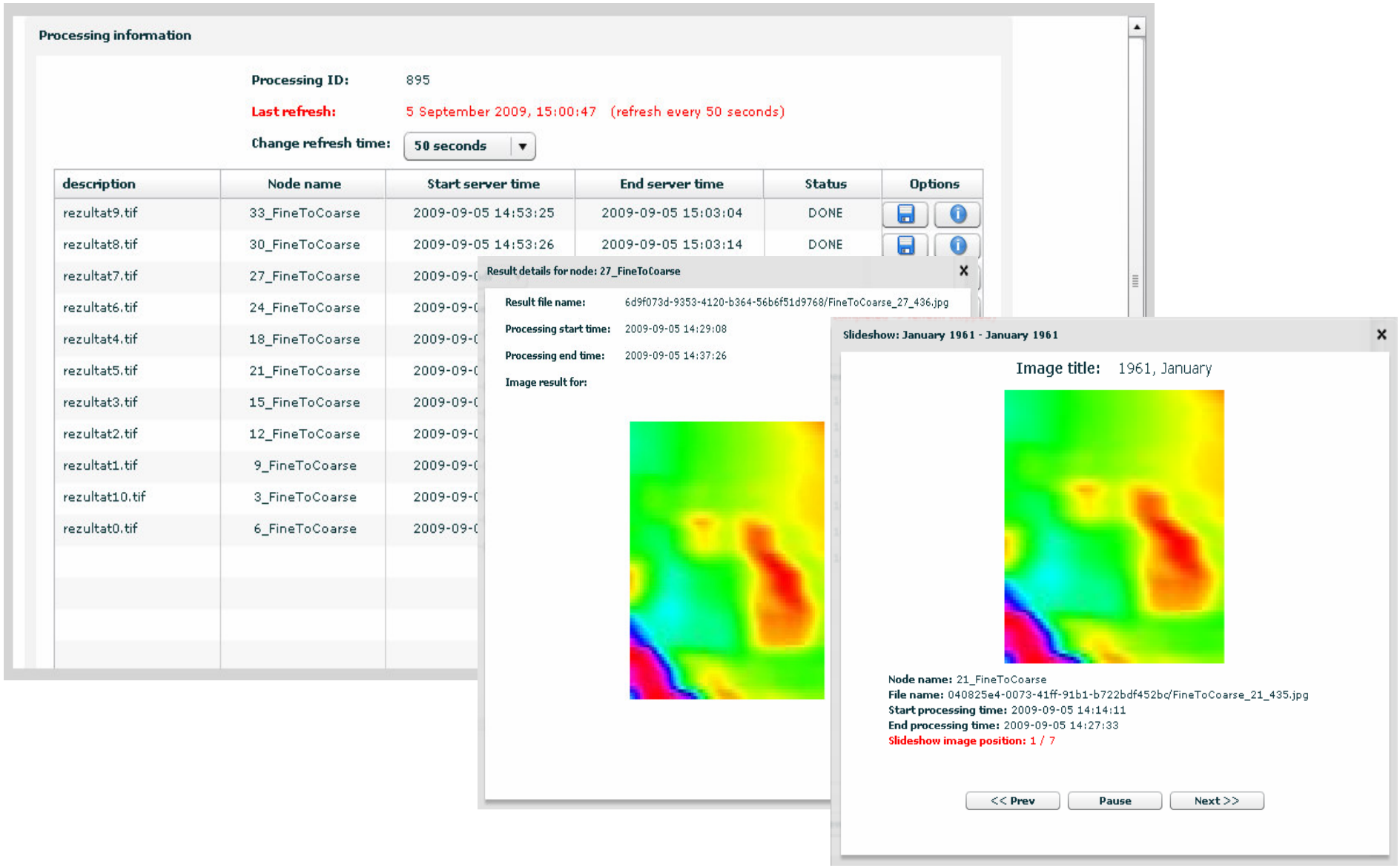
GreenView is available on Internet: http://gisheo01.mediogrid.utcluj.ro:8095/interpolation_v2.1/



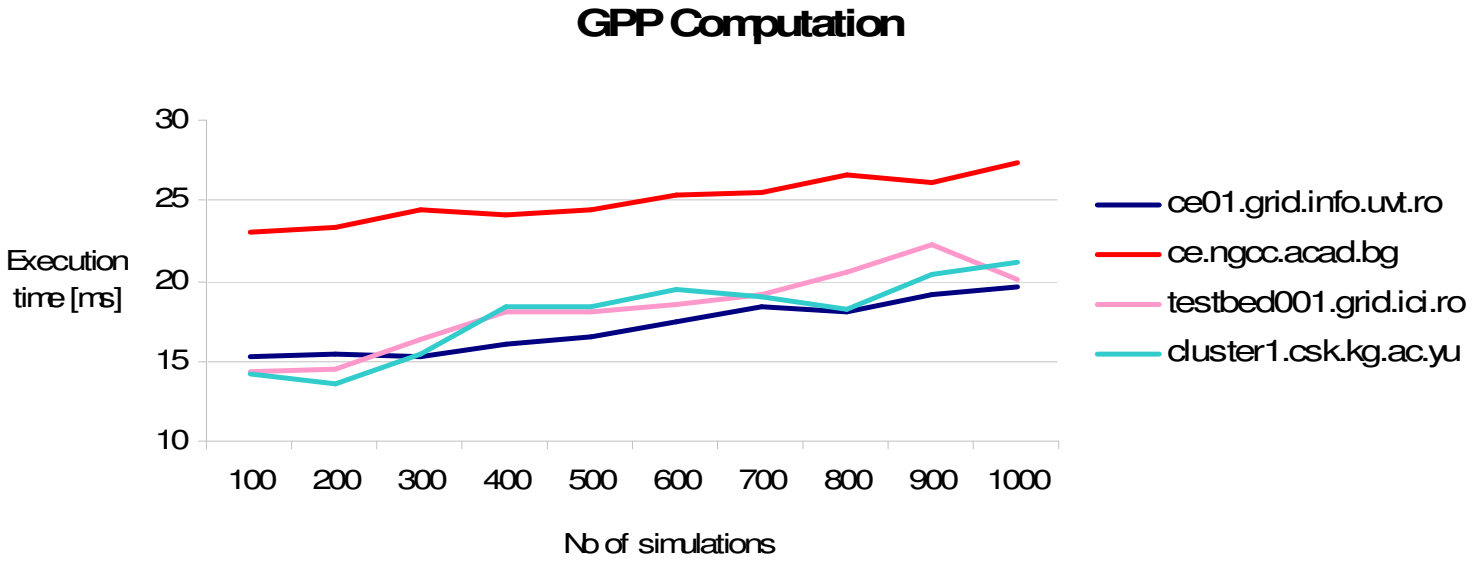
GreenView - GPP Computation

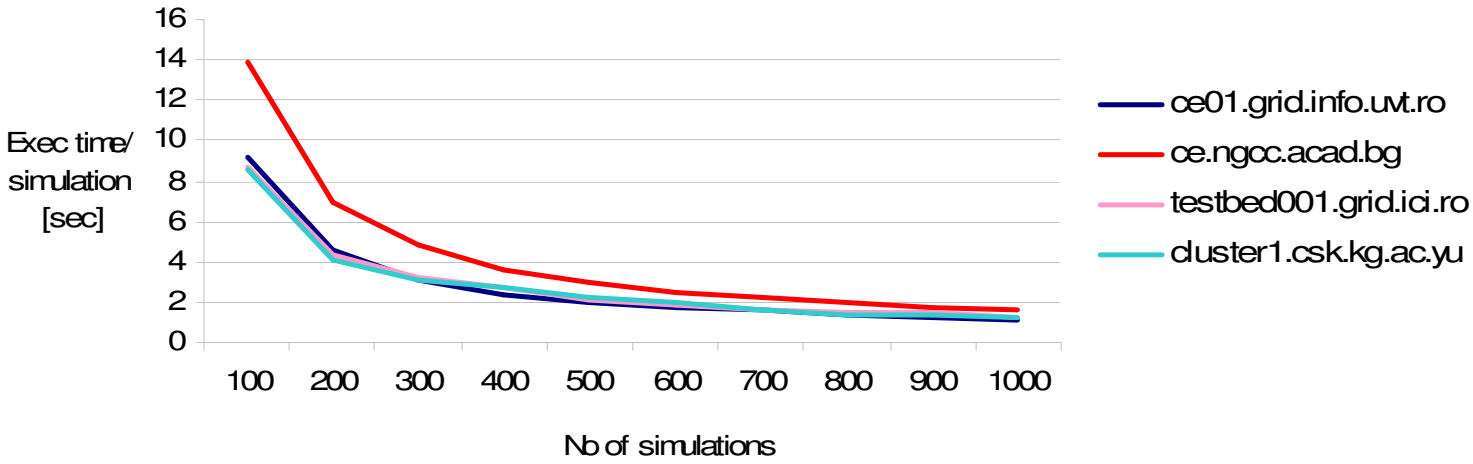


GreenView - Results



GreenView - GPP Computation





Research project

GiSHEO - On Demand Grid Services for Higher Education and Training in Earth Observation (http://gisheo.info.uvt.ro)

Supported by ESA PECS Program (http://pecs.esa.int/) for the period of 2008-2010

Research team (Romania):

1. West University of Timisoara (UVT)
2. Technical University of Cluj-Napoca (UTCN)
3. Romanian Space Agency (ROSA)
4. National Institute for Space Research (INCAS)

General Objectives

- Study the requirements and specifications for distance learning in Earth Observation training

  User requirements (i.e. professor, students), eLearning environment functional specifications, usability requirements, lesson structure, user interaction techniques.

- Tools development over GRID Technology to enable the creation and the execution of lessons for Earth Observation

  Create specialized tools that take advantage of GRID capabilities in a transparent manner for the user (student or teacher) in order to allow the processing of massive data.

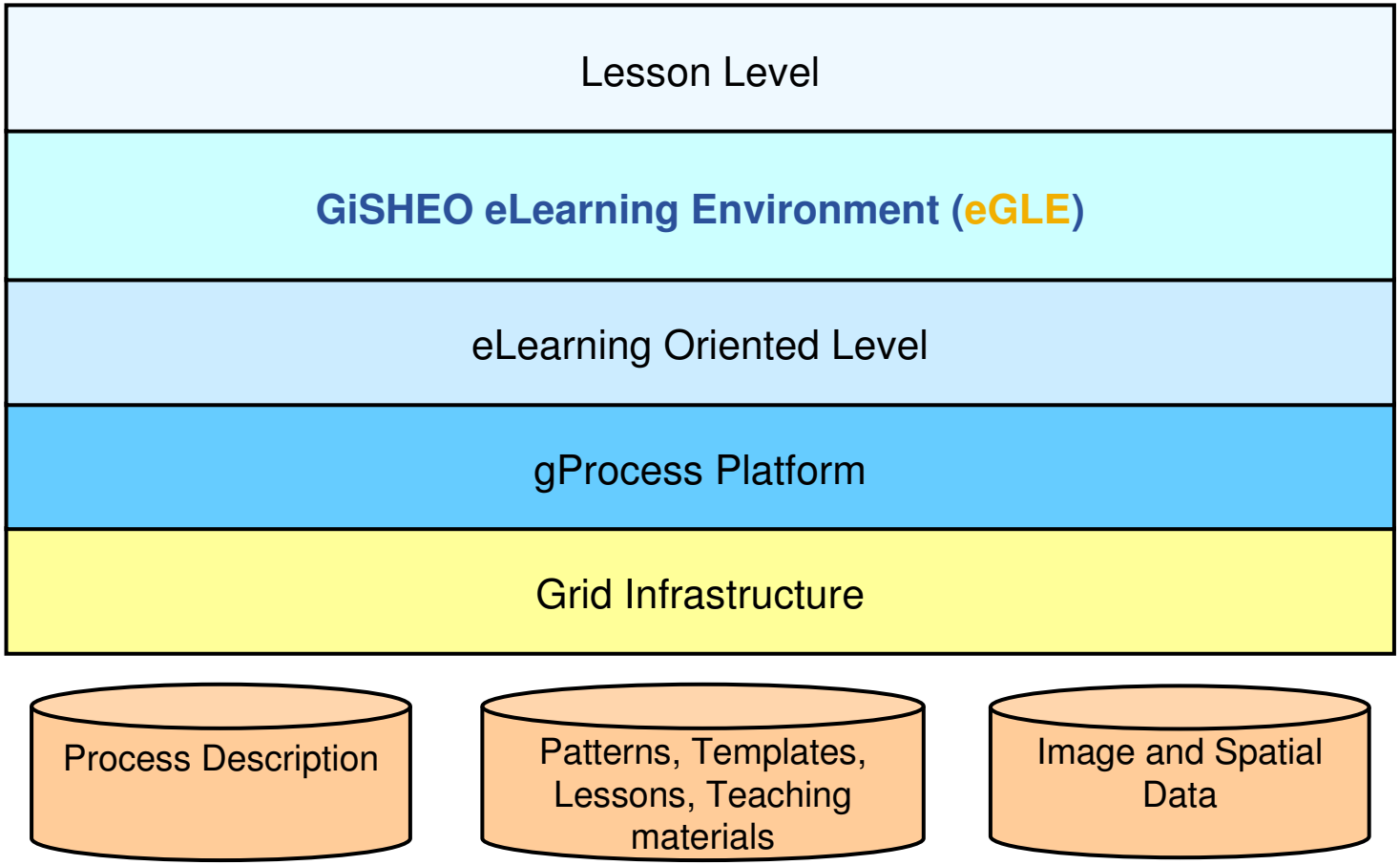
- Creation of lessons templates in order to allow non-technical teachers to create lessons for Earth Observation

  Create templates that allow teachers specialized in Earth Observation, but with non-technical studies, to develop new lessons by using GRID advantages and large amounts of data (satellite images).

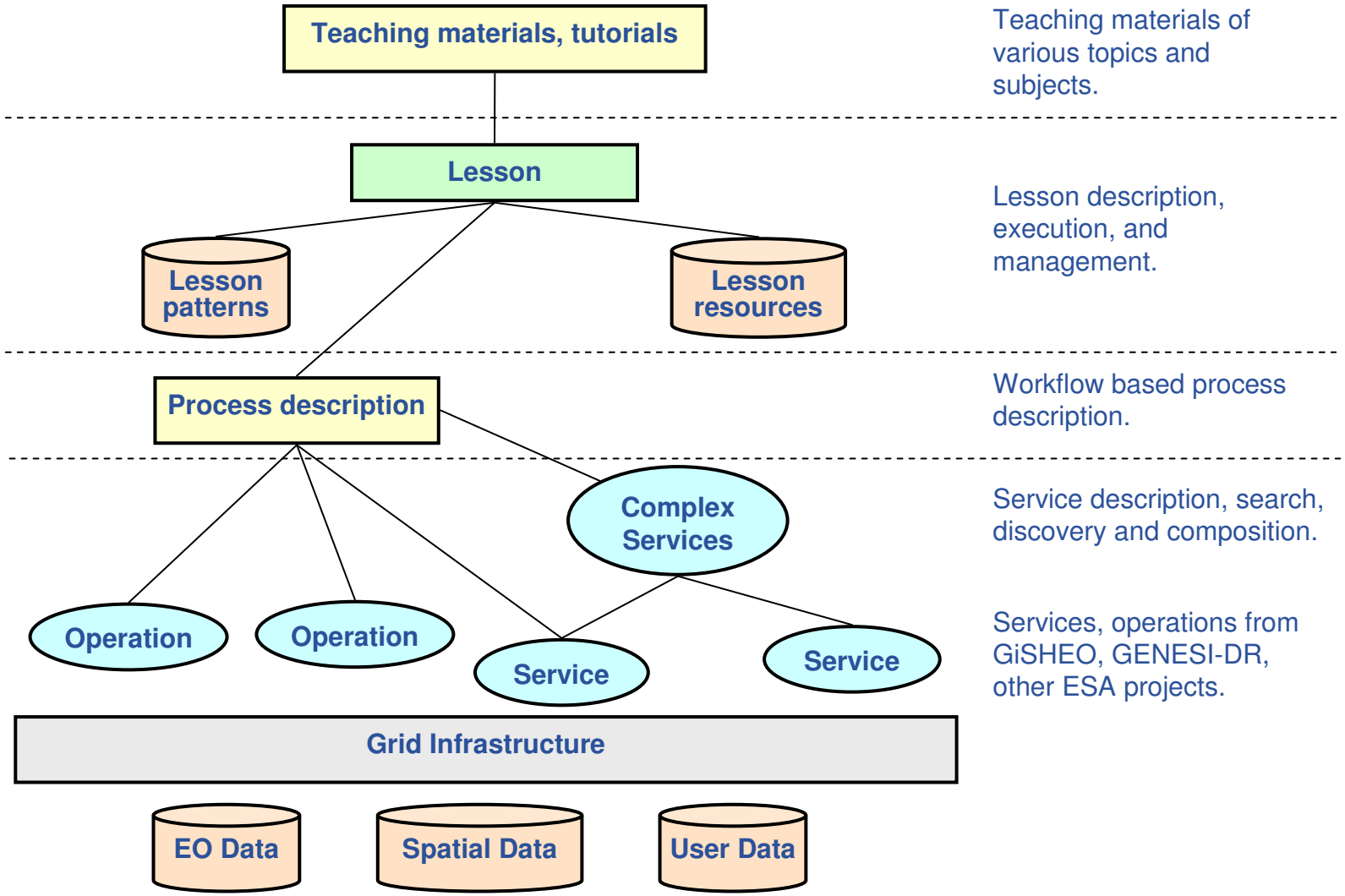
- Use Earth Observation data and services provided by GENESI-DR.

  Explore and experiment data availability and accessibility to GENESI-DR from remote Grid applications in terms of: accessing policies, types, formats, services, data replication, transfer performance, granularity, consistency, and efficiency

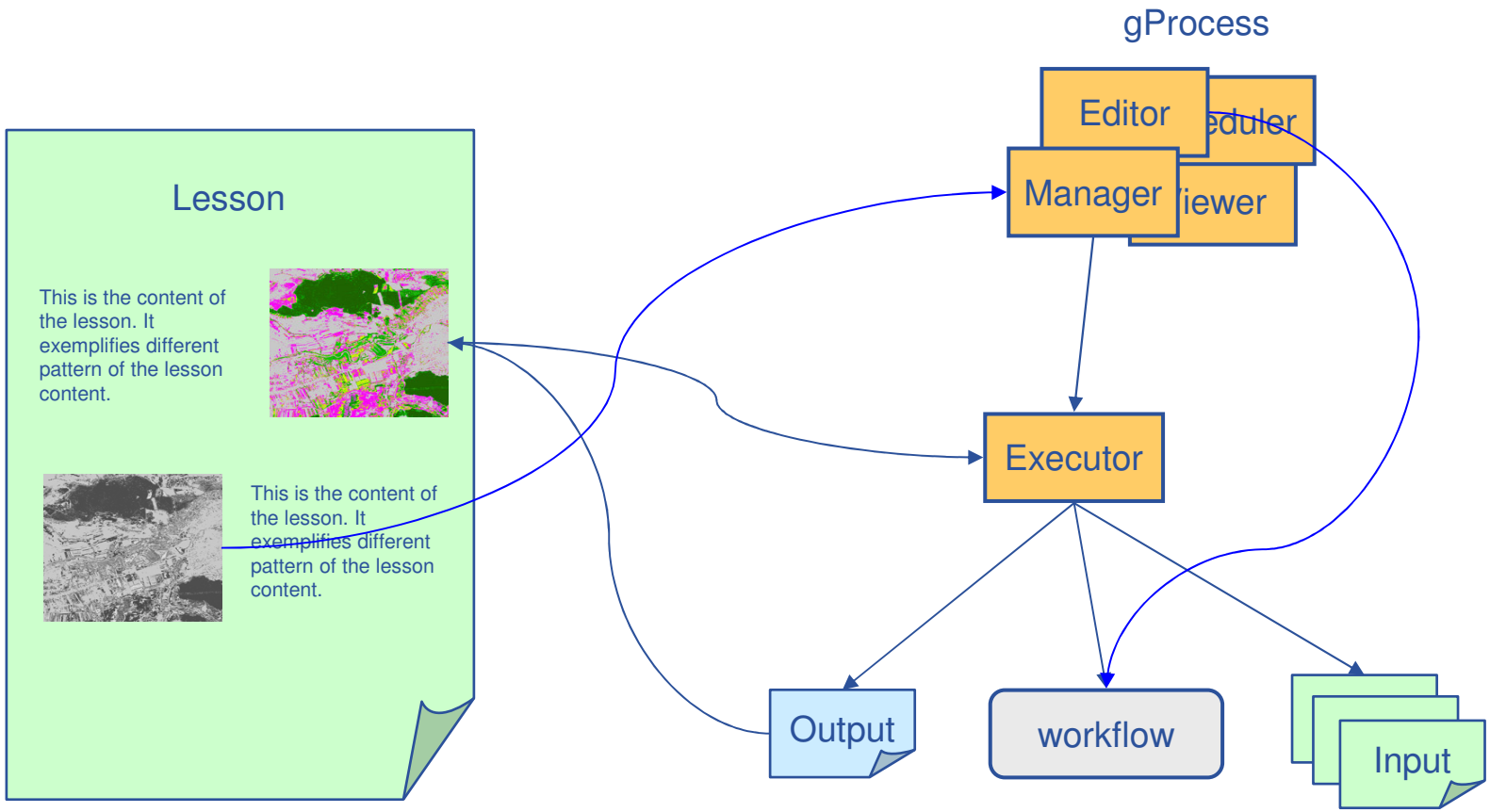
Functional Levels in eGLE Architecture



eGLE - Data and Process Flow



eGLE - Lesson Execution



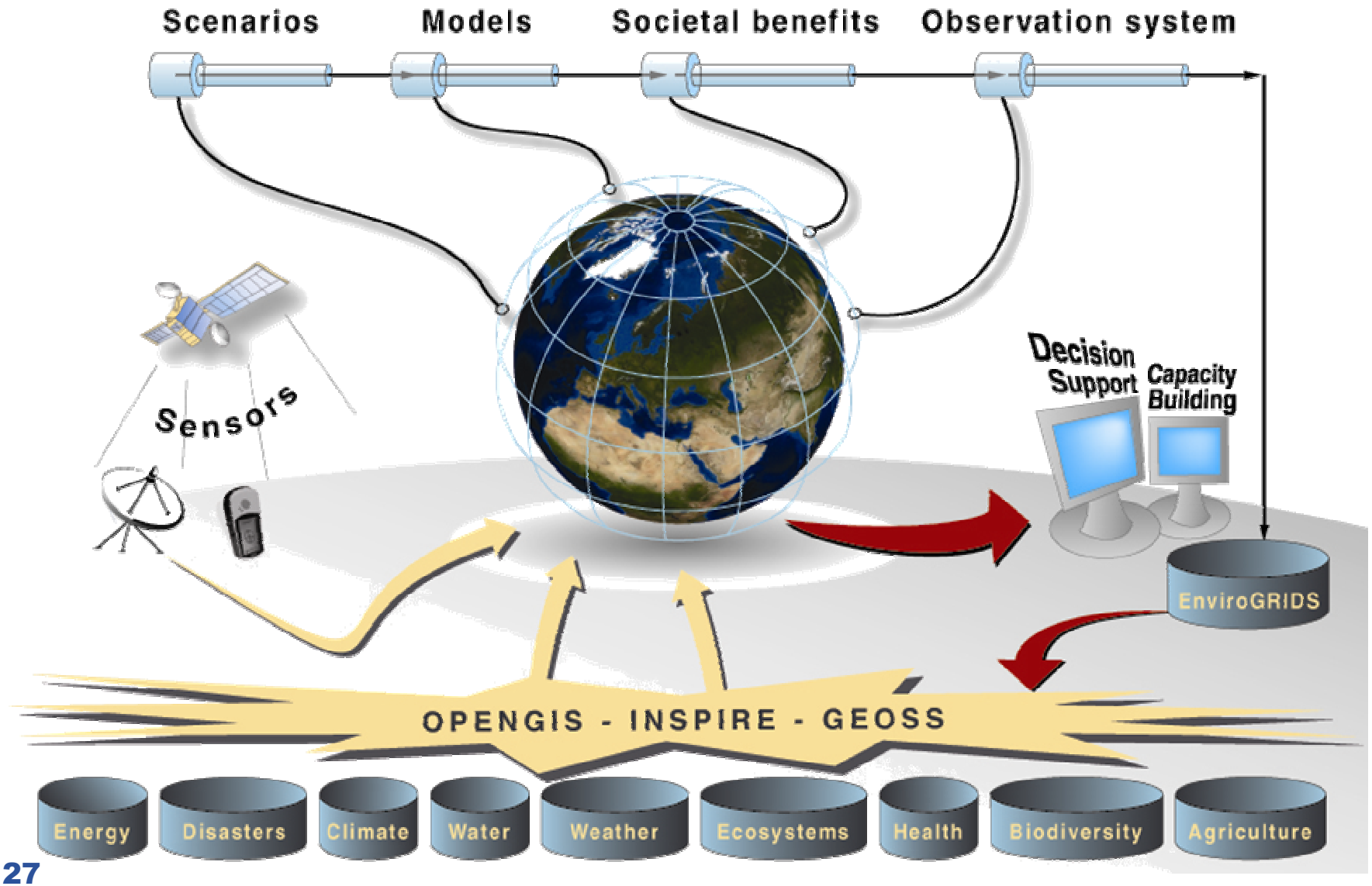
EnviroGRIDS Project

- **EnviroGRIDS - Gridifying the Black Sea catchment to support its sustainable development**

- Founded by the European Commission FP7 framework (Theme 6:environment), April 2009 – March 2013
- 27 partners

- **Coordinator**
  - University of Geneva, Switzerland

EnviroGRIDS Project



Building Capacity for a Black Sea Basin Observation and Assessment System supporting Sustainable Development

Map source: "Eutrophication in the Black Sea region; Impact assessment and causal chain analysis", Global International Waters Assessment (GIWA).

General Objectives

- Gap analysis
- Spatially explicit regional scenarios of development
- Modeling of large scale, high resolution distributed hydrologic processes
- Develop access to real time data from sensors and satellites
- Streamlining the production of indicators on sustainability and vulnerability of societal benefits
- Develop early warning and decision support tools at regional, national and local levels
- Build capacities in the implementation of many new standards and frameworks

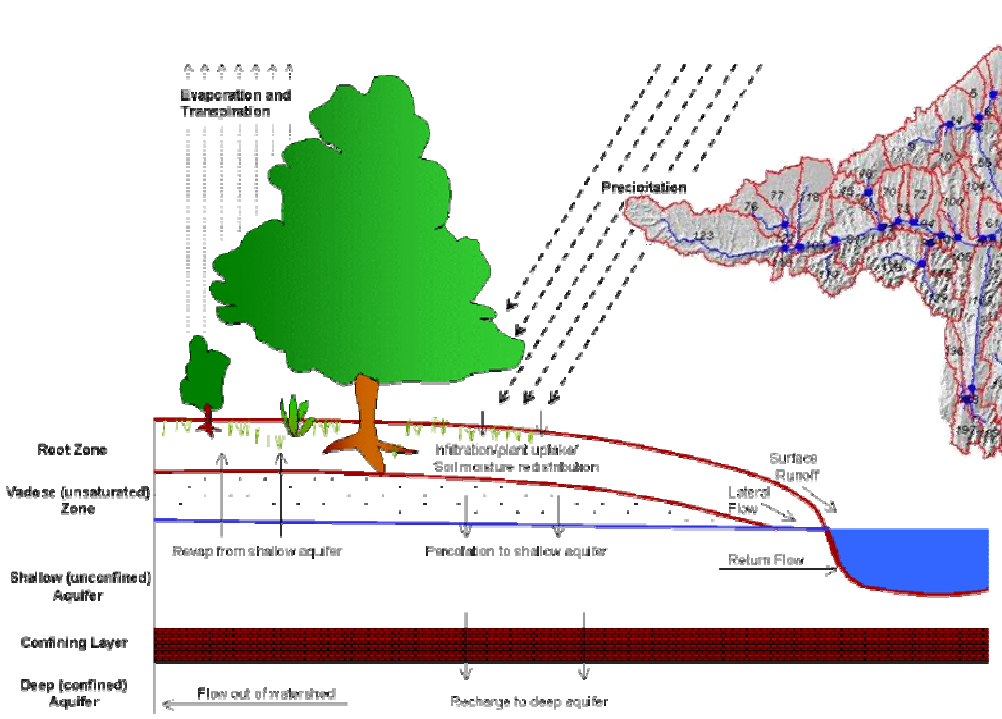
Grid Oriented Objectives and Tasks

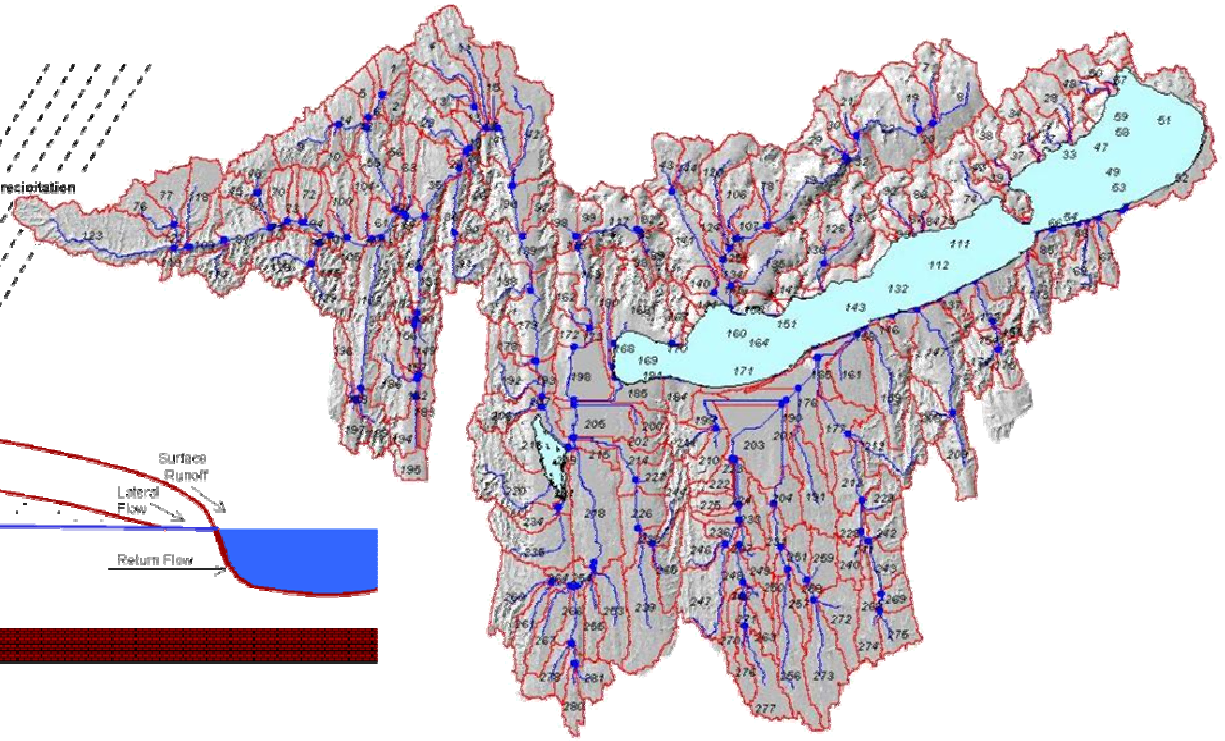
- Link, gather, store, manage and distribute key environmental data
- Gridification of applications
- Build capacities in the implementation of several new standards for sharing geospatial data.
- Main references are **GEOSS** (Open Geospatial Consortium), **OGS** (Open Geospatial Consortium), and **INSPIRE** Directive (Infrastructure for Spatial Information in the European Community)

Gridification of Applications

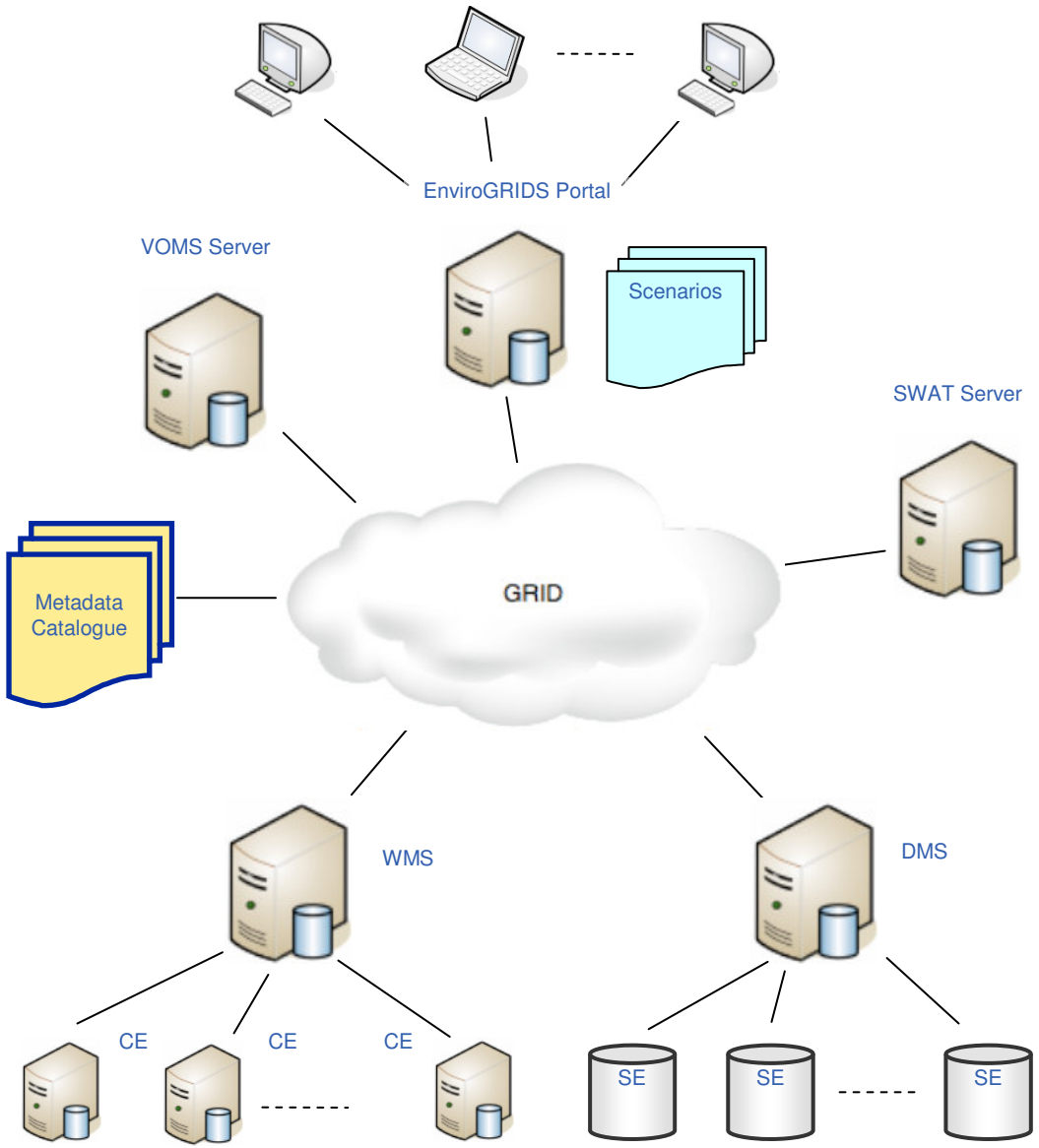
- Running hydrological model over the Grid

Soil & Water Assessment Tool | SWAT

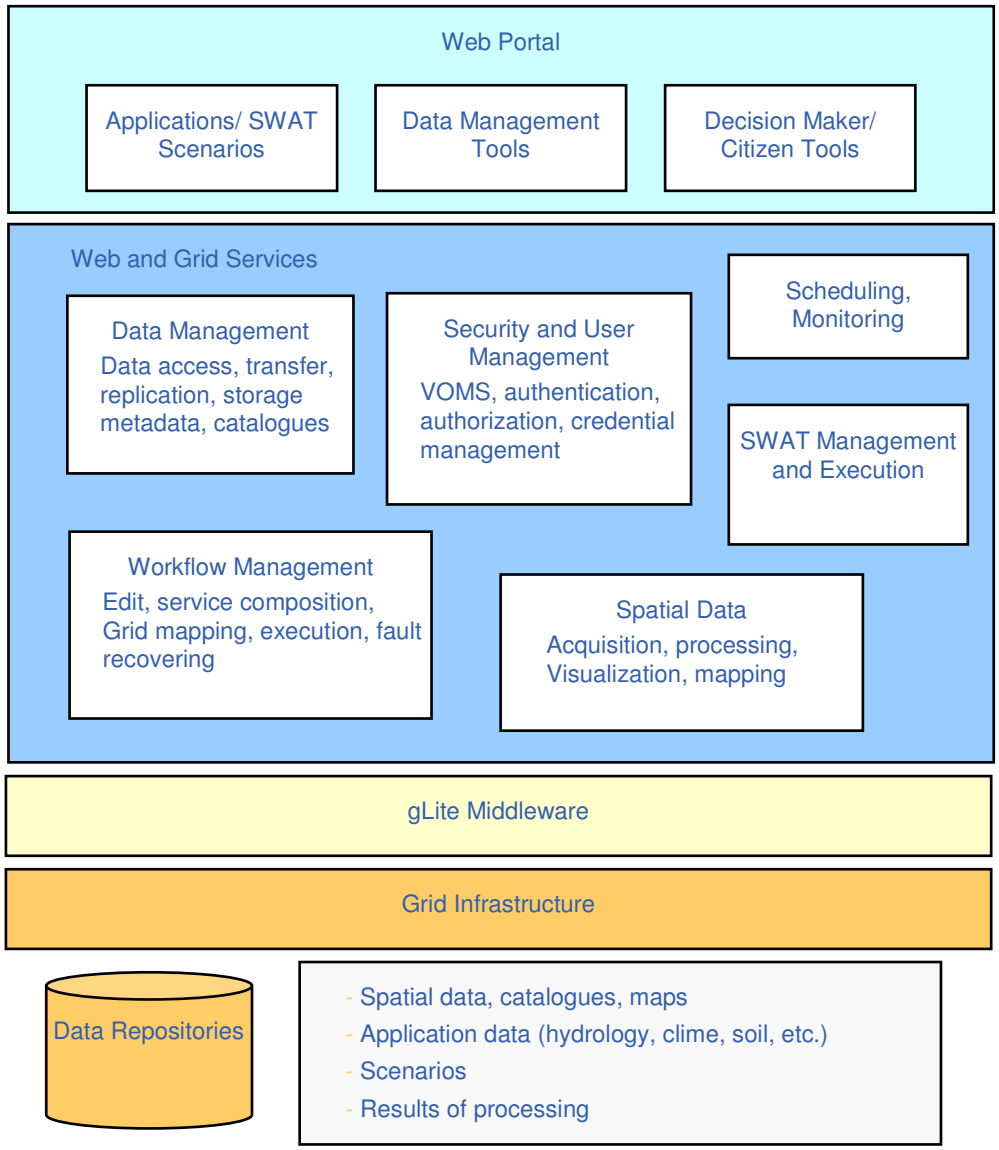




EnviroGRIDS System Architecture



EnviroGRIDS Functional Layers



Conclusions

- ESIP and gProcess platforms develop and experiment solutions to access and use efficiently GENESI-DR data and services
- gProcess based Earth Observation Application Development Methodology supports the using GENESI-DR data and services in Grid applications
- ESIP and gProcess will use spatial data and services provided by the GENESI-DR project, and will be basic platforms for the GISHEO (ESA-PECS) and EnviroGRIDS (FP7) projects
- GENESI-DR repositories contains huge quantities of data. More efficient and **user oriented approaches** for data searching and discovering have to be explored and developed
- The research projects and the developers need a GENESI-DR structure of **pilot data repositories** just for development and on-line tests. The pilot repositories must be resident on the same sites of GENESI-DR data providers. Data must be accessed through secured ways (i.e. certificates to access Portal and distributed repositories), and similar policies, as any other production data

Many thanks. Questions

SEE-GRID-SCI

SEE-GRID eInfrastructure for regional eScience

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

www.eu-egee.org

e-infrastructure