УДК 52(15).003

INTEROPERABILITY ISSUES OF EARTH OBSERVATION GRID SYSTEMS

A. Shelestov, S. Skakun, M. Korbakov

Space Research Institute NASU-NSAU,
03680, Kyiv, Prospekt Glushkova 40.
Tel. (+380-44)526 2553, fax (+380-44)526 4124,inform@ikd.kiev.ua

Interoperability Issues of Earth Observation Grid Systems. A. Shelestov, S. Skakun, M. Korbakov. In this paper we review issues of Earth observation Grid systems integration. We describe different approaches for the solution of problems of certificate trust, data transfer, geospatial data access, task management, etc. As an example, we describe InterGrid system for environmental and natural disaster monitoring that integrates several regional and national Grid systems.

Интеграция Grid-систем исследования Земли. Шелестов А.Ю., Скакун С.В., Корбаков М.Б. В работе рассматриваются вопросы интеграции Grid-систем исследования Земли. Приведены возможные подходы к решению задач, возникающих при интеграции: обмен сертификатами, обмен данными, доступ к геопространственным данным, выполнение заданий и т.д. В качестве примера использования описанных решений описана InterGrid-система, которая объединяет несколько региональных и национальных Grid-систем и направлена на решения задач экологического мониторинга и мониторинга природных чрезвычайных ситуаций.

1. Earth observation Grid-based systems

At present, Grid technologies [1, 2] are widely applied in different domains, in particular EO domain. EU-funded European DataGrid Project (EDG) was one of the first Grid-enabled projects allowing European Space Agency (ESA) to gain firsthand experience in the use of emerging Grid technologies. To test the capabilities of the system, it was decided to use data from ERS-2's GOME instrument, consisting of global atmospheric-ozone measurements collected over several years of the mission [3]. This instrument generates over 400 terabytes of data products per year that have to be catalogued, archived and processed. EDG project evolved into EGEE and EGEE2 projects that developed own middleware – gLite – and integrate more than 90 institutions in over 30 countries.

Based on the gained experience European Space Agency (ESA) and European Space Research Institute (ESRIN) are developing Grid Processing on Demand (G-POD) for Earth Observation Applications [4]. Grid is considered as a comfortable "open platform" for handling computing resources, data, tools, etc., and not limited to only high performing computing. Online access to different data is enabled within this project, in particular to data provided by various instruments on Envisat satellite [5], SEVIRI instrument onboard MSG (Meteosat Second Generation) satellite [6], ozone profiles derived from GOME instrument, etc. One of the most important applications is the analysis long-term data. For example, the analysis of 8 years of GOME on-board temperatures (overall 525 Gb of data) took less than 2 days on 40 computer elements of ESRIN "Grid-on-demand" structure (overall 38460 files were processed). Grid Web Portal [4] provides access to the "Grid-on-demand" resources enabling: personal certification, time/space selection of data directly from the ESA catalogue, data transfer, job selection, launching and live status, data visualization, etc. At present "Grid-on-demand" infrastructure consists of more than 150 working nodes with ability to store and handle of about 70 Gb of data. As middleware Globus Toolkit 2.4 and LCG/EGEE components are used.

Spatial Information Grid (SIG), a research project supported by 863 projects of China government, is a series of special grid researches in the filed of Earth Observation. SIG has been designed to be the testbed of grid middleware research and grid-enable spatial information services and applications. There are 12 data centers have been involved SIG. The Web Portal has been developed in order to provide access to SIG resources [7]. This portal enables geo-data discover and processing, work monitoring, and grid resources (all service/job/node etc.) management.

DEGREE (Dissemination and Exploitation of GRids in Earth science) project [8] is initiated within EGEE/EGEE-II. A major challenge for DEGREE is to build a bridge linking the Earth Science and GRID communities throughout Europe, and focusing in particular on the EGEE-II Project. Grid is considered to be the appropriate platform for integration of heterogeneous data resources, processing tools, models, algorithms, etc. Moreover, Grid provides appropriate infrastructure enabling international cooperation within GMES and GEOSS. The following problems are within the scope of DEGREE: earthquake analysis, floods modeling and forecasting, influence of climate changes on agriculture

Japan Aerospace eXploration Agency (JAXA) and KEIO University started establishing "Digital Asia" system aimed at semi-real time data processing and analyzing. They use GRID environment to accumulate knowledge and know-how to process remote sensing data. The Digital Asia project is the part of bigger Sentinel Asia project that is targeting on building natural disasters monitoring system [9].

National Aeronautics and Space Administration (NASA) have created Information Power Grid (IPG) targeting an operational Grid environment incorporating major computing and data resources at multiple NASA sites in order to provide an infrastructure capable of routinely addressing larger scale, more diverse, and more transient problems than is

© A. Shelestov, S. Skakun, M. Korbakov, 2008

possible today. Nowadays IPG have approximately 600 CPU nodes of Computing resources and 30-100 Terabytes of archival information/data storage resources.

CEOS Wide Area Grid (WAG) project is initiated by CEOS Working Group on Information Systems and Services (WGISS), and aims at providing horizontal infrastructure enabling efficient integration of resources of different space agencies. WAG testbed infrastructure is currently under development within ESA Cat-1 project "Wide Area Grid Testbed for Flood Monitoring Using Spaceborne SAR and Optical Data" (no. 4181) [10].

Within WAG project Space Research Institute NASU-NSAU have developed testbed that integrates resources of Ukrainian Grid segment (Ukrainian Academician Grid) with resources of international organisations (ESA, RSGS-CAS).

Nowadays Earth Observation (EO) data play a major role in solving problems in different domains. Satellite observations enable acquisition of data for large and hard-to-reach territories, can provide continuous measurements and human-independent information, etc. EO domain, in turn, is characterized by large volumes of data that should be processed, catalogued, and archived. For example, GOME instrument onboard Envisat satellite generates nearly 400 Tb data per year [3]. EUMETCast system that is part of global GEONETCast system [11] of GEOSS enables acquisition of more than 50 Tb of unprocessed information per year. Moreover, the processing of satellite data is carried out not by the single application with monolithic code, but by distributed applications. This process can be viewed as a complex workflow that is composed of many tasks: geometric and radiometric calibration, filtration, reprojection, composites construction, classification, products development, post-processing, visualization, etc. [12]. For example, calibration and mosaic composition of 80 images generated by ASAR instrument onboard Envisat satellite takes 3 days on 10 workstations of Earth Science GRID on Demand that is being developed in ESA and ESRIN.

2. Tendencies of globalization: GEOSS, GMES, INSPIRE

Nowadays, there is a trend for globalization of monitoring systems with purpose of solving more complex problems and reducing collaboration expenses. EO data are naturally distributed over many organizations involved in data receiving and processing. This leads to the need of integration of existing systems for solution of complex problems.

The development of *GEOSS (Global Earth Observation System of Systems)* [13] is coordinated by Group on Earth Observations (GEO) [14] that was launched in response to calls for action by the 2002 World Summit on Sustainable Development and the G8 (Group of Eight) leading industrialized countries. GEO is a voluntary partnership of governments and international organizations that provides a framework within which these partners can develop new projects and coordinate their strategies and investments. It is recognised that GEOSS work with and build upon existing national, regional, and international systems to provide comprehensive, coordinated Earth observations from thousands of instruments worldwide, transforming the data collected into vital information for society.

GEOSS is based on the use of open standards for geospatial data, in particular OGC. *OGC (Open Geospatial Consortium)* is non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services. WCS (Web Coverage Service) is one of the OGC standards (among with Web Mapping Service and Web Feature Service) governing network access to geospatial data. WCS describes interface to allow access to geospatial "coverage" that represent values or properties of geographic locations. Current efforts in the development of this standard are turned upon extending it to allow access to multidimensional geospatial data.

GMES (Global Monitoring for Environment and Security) [15] is a European initiative for the implementation of information services dealing with environment and security. GMES is based on observation data received from EO satellites and ground based information. These data are coordinated, analysed and prepared for end-users. GMES provides the following services that can be grouped in three major categories:

— Mapping, including topography or road maps but also land-use and harvest, forestry monitoring, mineral and water resources that do contribute to short and long-term management of territories and natural resources. This service generally requires exhaustive coverage of the Earth surface, archiving and periodic updating of data.

— Support for emergency management in case of natural hazards and particularly civil protection institutions responsible for the security of people and property. This service concentrates on the provision of the latest possible data before intervening.

— Forecasting is applied for marine zones, air quality or crop yields. This service systematically provides data on extended areas permitting the prediction of short, medium or long-term events, including their modelling and evolution.

INSPIRE (Infrastructure for Spatial Information in Europe) initiative intends to trigger the creation of a European spatial information infrastructure that delivers to the users integrated spatial information services [16]. These services should allow the users to identify and access spatial or geographical information from a wide range of sources, from the local level to the global level, in an inter-operable way for a variety of uses. The target users of INSPIRE include policy-makers, planners and managers at European, national and local level and the citizens and their organisations. Possible services are the visualisation of information layers, overlay of information from different sources, spatial and temporal analysis, etc.

3. Integration of Grid systems

Modern tendencies of globalization and development of “system of systems” GEOSS lead to the need of integration of heterogeneous Grid systems.

Interoperability of Grid systems supposes running applications on distributed computational resources provided by different domains (Fig. 1). Since many of the existing Earth observation system rely on Grid technologies appropriate approaches and technologies should be evaluated and developed to enable Grid system integration (so called InterGrid).

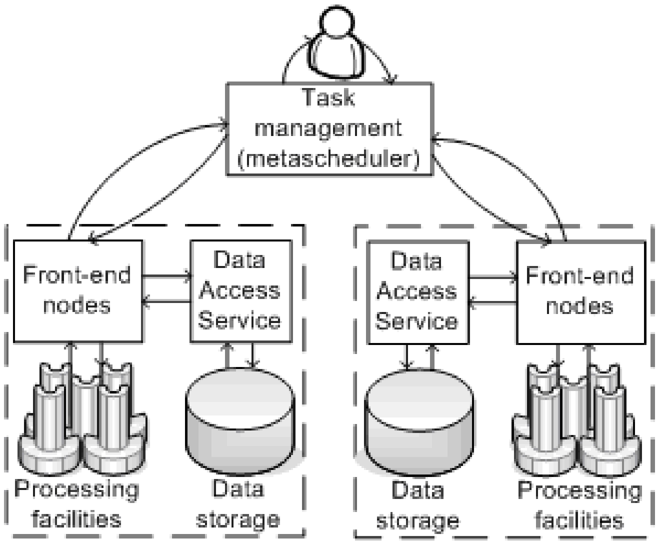

Fig 1. Task management level

4. Possible solutions for Grid system interoperability

In this section we present main issues and possible solutions for Grid-system integration. Main prerequisite of such kind of integration is certificates trust. It could be done, for example, through EGEE infrastructure that nowadays brings together the resources of more than 70 countries. Another problems concerned with different Grid systems integration are as follows:

— enabling data transfers and high-level access to geospatial data;
— development of common catalogues;
— enabling jobs submission and monitoring;
— enabling information exchange.

Data transfer. GridFTP is an appropriate and reliable solution for data transfer. The only limitation is the requirement of transparent LAN (local area network) infrastructure.

Access to geospatial data. High-level access to geospatial data can be organised in two possible ways: using pure WSRF services or using OGSA-DAI container [17]. Each of this approach has its own advantages and weaknesses. Basic functionality for WSRF-based services can be easily implemented (with proper tools), packed and deployed. But advanced functionality such as security delegation, third-party transfers, indexing should be implemented by hands. WSRF-based services can also pose some difficulties if we need to integrate them with other data-oriented software.

OGSA-DAI framework provides uniform interfaces to heterogeneous data. This framework makes possible to create high-level interfaces to data abstracting hiding details of data formats and representation schemas. Most of problems in OGSA-DAI are handled automatically, e.g. delegation, reliable transfer, data flow between different sources and sinks. OGSA-DAI containers are easily extendable and embeddable. But comparing to WSRF basic functionality implementation of OGSA-DAI extensions is more difficult. Moreover, OGSA-DAI requires preliminary deployment of additional software components.

Task management. There are two possible approaches for task management. One of them is to use Grid portal (Fig. 2) supporting different middleware platforms, such as GT4, gLite, etc. Grid portal is an integrated platform to end-users that enables access to Grid services and resources via standard Web browser. Grid portal solution is easy to deploy and maintain, but it doesn’t provide application interface and scheduling capabilities.

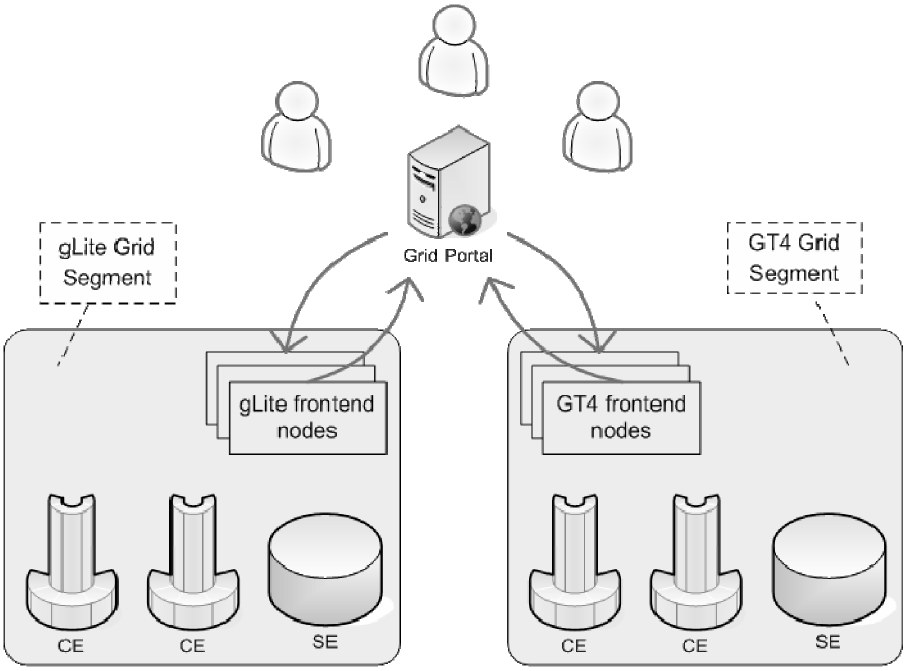

Fig. 2. Portal approach to Grid system integration

Another approach is to develop high-level Grid scheduler (Fig. 3) that will support different middleware by providing some standard interfaces. Such metascheduler interacts with low-level schedulers (used in different Grid systems) enabling in such way system interoperability. Metascheduler approach is much more difficult to maintain comparing to portals; however, it provides API with advanced scheduling and load-balancing capabilities. At present, the most comprehensive implementation for the metascheduler is a GridWay system. The GridWay metascheduler is compatibility with both Globus and gLite middlewares. Starting from Globus Toolkit v4.0.5 GridWay become standard part of its distribution. GridWay system provides comprehensive documentation for both users and developers that is a important point for implementing new features.

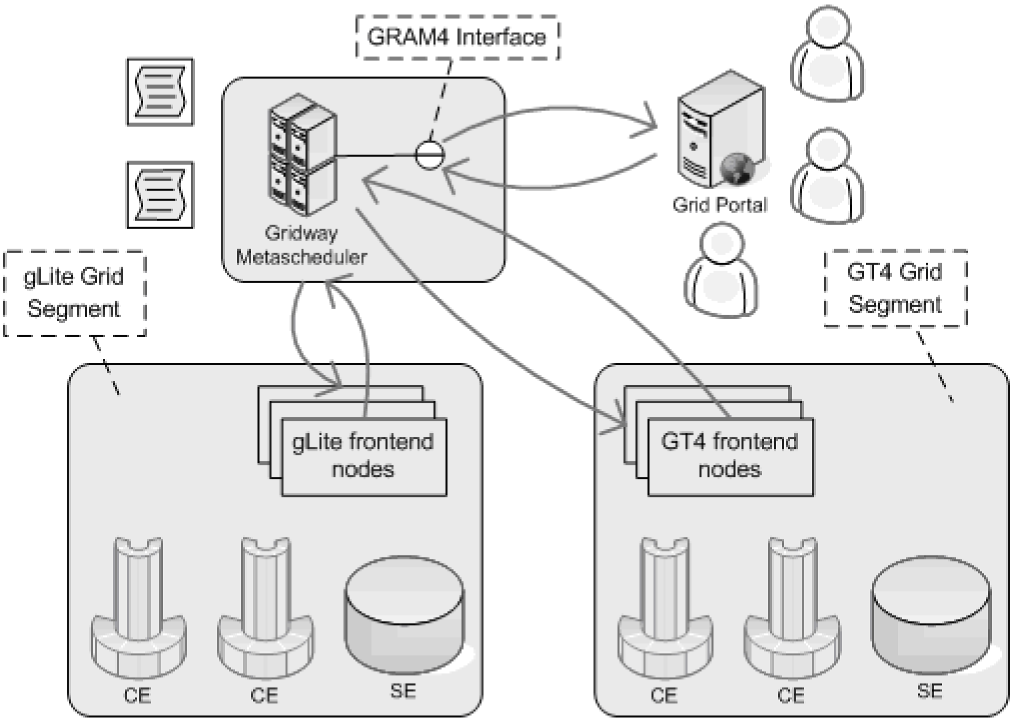

Fig. 3. Metascheduler approach

In the next section we show the example of application of described approaches to integration the development of InterGrid environment.

5. Implementation: lessons learned

Technologies described in the pervious sections were used for the development of InterGrid for environmental and natural disaster monitoring. InterGrid integrates Ukrainian Academician Grid (with Satellite data processing Grid segment) and RSGS Grid (Chinese Academy of Sciences) and is considered as a testbed for Wide Area Grid (WAG) implementation—a project initiated within CEOS Working Group on Information Systems and Services (WGISS).

The important application that is being solved within InterGrid environment is flood monitoring and prediction [18, 19]. This task requires adaptation and tuning of existing hydrological and hydraulic models for corresponding territories and the use of heterogeneous data stored on multiple sites. Flood monitoring and prediction requires the use of the following data sets: NWP modelling data (provided by Satellite data processing Grid segment), SAR imagery from Envisat/ASAR and ERS-2/SAR satellites (provided by ESA), products derived from optical and microwave satellite data such as soil moisture, precipitation, flood extent etc., in-situ observations from meteorological ground stations and digital elevation model (DEM). The process of model adaptation can be viewed as a complex workflow and requires the solution of optimization problems (so called parametric study). Satellite data processing and products generation tasks also represent complex workflow and require intensive computations. All these factors lead to the need of using computational and informational resources of different organizations and their resources into joint InterGrid infrastructure. The architecture of proposed InterGrid is depicted in Fig. 4.

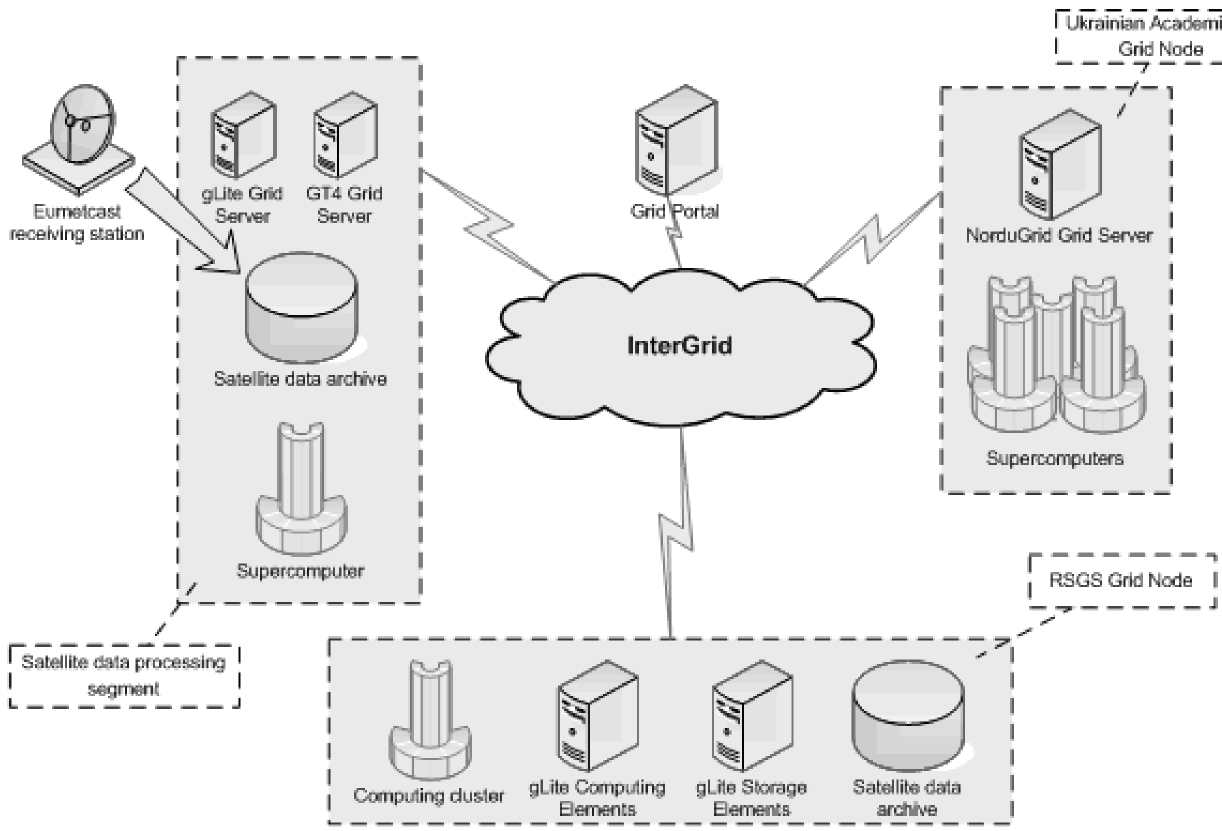

Fig. 4. InterGrid architecture

GridFTP was chosen to provide data transfer between Grid systems. In order to enable interoperability between different middleware (for example, Satellite data processing Grid segment is using GT4; RSGS Grid is using gLite 3.x; Ukrainian Academician Grid is based on NorduGrid) we developed Grid portal that is based on GridSphere portal framework [20]. The developed Grid portal allows users to transfer data between different nodes and submit jobs on computational resources of the InterGrid environment. The portal also provides facilities to monitor statistics of the resources such as CPU load, memory usage, etc. The further works on providing interoperability between different middlewares are directed to the development of metascheduler using GridWay system [21]. In the nearest future we are intended to provide integration with ESA's EO Grid-on-Demand infrastructure.

In order to provide visualisation of data and derived products for flood monitoring we develop user interface that is based on OpenLayers [22]. The example of OpenLayers visualization for flood application is depicted in Fig. 5.

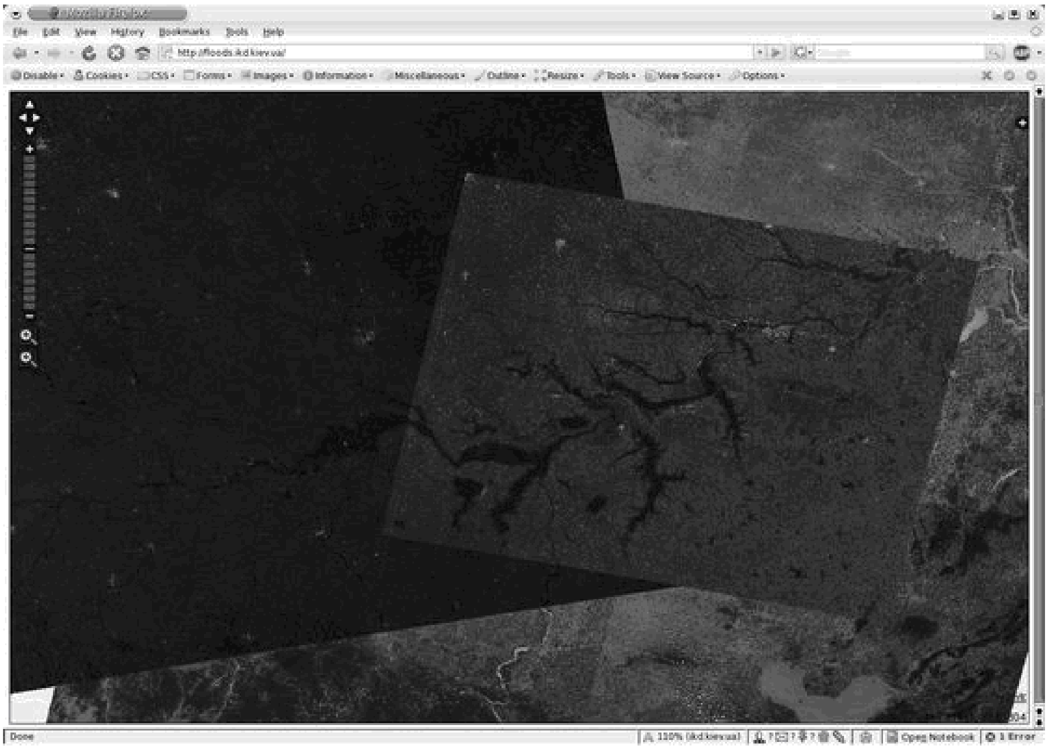

Fig. 5. Flood application within InterGrid testbed. Flood event: Huaihe River, China, July, 2007. Data sources: Envisat/ASAR (© ESA, 2007) and RADARSAT-1 (© CSA, 2007)

6. Conclusions

This paper focuses on the problems of integration of Earth observation systems, in particular those using Grid platform. We reviewed two solutions for providing interoperability between Grid systems managed by different domains, namely portal-based and metascheduling approach. We implemented portal solution based on GridSphere framework for the InterGrid environment that integrates several regional and national Grid systems. In order to provide advanced scheduling and load-balancing capabilities the further works will be directed to the implementation of metascheduler based on GridWay system. Also we are intended to provide integration with ESA's G-POD. Further investigations will be also directed to the integration of distributed monitoring systems with SensorWeb networks in order to provide automatic delivery of data from heterogeneous sources.

Acknowledgment. This work is supported by ESA CAT-1 project "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (#4181) and by INTAS-CNES-NSAU project "Data Fusion Grid Infrastructure" (Ref. Nr 06-1000024-9154).

1. *Foster I., Kesselman C.* "The Grid: Blueprint for a New Computing Infrastructure", 2nd Edition, Morgan Kaufmann, 2004.
2. *Shelestov Andrey Yu., Kussul Nataliya N., Skakun Sergey V.* "Grid Technologies in Monitoring Systems Based on Satellite Data" J. of Automation and Information Science, 2006. – vol. 38. – issue 3. – P. 69-80.
3. *Fusco L., Goncalves P., Linford J., Fulcoli M., Terracina A., D'Acunzo G.* "Putting Earth-Observation on the Grid", ESA Bulletin, 2003. – 114. – P. 86-91.
4. *ESA* Grid Processing on Demand, http://gpod.eo.esa.int.
5. *ESA* Envisat, http://envisat.esa.int.
6. *EUMETSAT*, http://www.eumetsat.int.
7. *SIG* portal, http://159.226.224.52:6140/Grid/application/index_en.jsp.
8. *Dissemination* and Exploitation of GRids in Earth sciencE. — http://www.eu-degree.eu.
9. *Fukui H.* "Sentinel Asia/Digital Asia: Building Information Sharing Platform by Geo web services and contributing to Disaster Management Support in the Asia-Pacific Region".
10. *Kopp P., Petiteville I., Shelestov A.,* Li G. "Wide Area Grid (WAG)", Proc. The 7th Ukrainian Conference on Space Research, National Flight and Control Center, Evpatoria, Ukraine, 2007. – 209 p.
11. *GEONETCast*, http://www.earthobservations.org/progress/GEONETCast.html.
12. *Rees* W.G. "Physical Principles of Remote Sensing", Cambridge University Press, 2001.
13. *GEOSS* Portal, http://www.geossportal.com.
14. *Group* on Earth Observations (GEO), http://www.earthobservations.org.
15. *GMES*, http://www.gmes.info.
16. *INSPIRE*, http://inspire.jrc.it.
17. *OGSA*-DAI Project, www.ogsadai.org.uk
18. *Kussul N., Shelestov A., Skakun S., Kravchenko O.* "Data Assimilation Technique For Flood Monitoring and Prediction" International Journal on Information Theory and Applications, 2008, volume 15. (in print).
19. *Kussul N., Lupian E., Shelestov A., Skakun S., Tischenko Yu., Hluchy L.* "Flood extent extraction using data from different sources" // Journal of Automation and Information Sciences, 2007, Issue 6, P. 117-126.
20. *GridSphere*, http://www.gridsphere.org.
21. *GridWay* Metascheduler, http://www.gridway.org.
22. *OpenLayers*, http://www.openlayers.org.