SIB-ESS-C – a Siberian Earth System Science Cluster to process satellite data and provide access to derived products in the web

Jonas Eberle, Roman Gerlach, Prof. Dr. Christiane Schmullius

Friedrich-Schiller-University, Department for Earth Observation

Loebdergraben 32, 07743 Jena, Germany

EMail: jonas.eberle@uni-jena.de

ABSTRACT

The Siberian Earth System Science Cluster (SIB-ESS-C) established at the University of Jena (Germany) is a spatial data infrastructure implementing standards published by the Open Geospatial Consortium (OGC) and the International Organization for Standardization (ISO) aimed at providing researchers with focus on Siberia with the technical means for data discovery, data access and data analysis. At the current development stage the SIB-ESS-C system comprises a federated metadata catalogue accessible through the SIB-ESS-C Web Portal or from any OGC-CSW compliant client. The Web Portal also contains a simple map-like visualization component which is currently being extended to a comprehensive visualization and analysis tool. The visualization component enables users to overlay different dataset found during a catalogue search. All data products are accessible as Web Mapping, Web Feature or Web Coverage Services allowing users to directly incorporate the data into their application. This paper describes previous developments and future ideas to build up a usefull and userfriendly access to satellite data and derived products for Siberia with state of the art web technologies and standards of the OGC.

Keywords: Earth System Science, Siberia, satellite products, gmes, geoss, ogc, iso, processing

INTRODUCTION

The Siberian Earth System Science Cluster (SIB-ESS-C) established at the University of Jena (Germany) is a spatial data infrastructure to provide earth observation products in the area of Siberia. The spatial data infrastructure implements standards published by the Open Geospatial Consortium (OGC) and the Internationale Organizsation for Standardization (ISO) for data discovery, data access and data analysis. The objective of the Siberian Earth System Science Cluster is to faciliate environmental research and Earth system science in Siberia. Working with international standards the system can be also published for other regions. The region for this project covers the entire Asian part of the Russian Federation approximately between 58°E - 170°W and 48°N - 80°N. Products from several earth observation projects (e.g. SIBERIA-I & II, SibFORD) were described with metadata, integrated und published with OGC-compliant services to provide possibilities to exchange the data with other users and software products.

To provide discovery, access and analysis services a webportal was published for searching and visualisation the available data. This webportal is based on current web technologies like AJAX and a user-friendly surface with Drag-n-Drop and further mouse events. Over the webportal the user can search for the published products showing results in a grouped table or a timeline and add these products to the map interface. As further development an analysis service for MODIS timeseries data is currently integrated based on the catalogue search results.

New developments in the next years are the integration of satellite data with an integrated product processing system as well as a monitoring system of the processed products. First of all satellite data and algorithms as OGC-compliant Web Processing Service (WPS) can be registered and described with metadata. In the second step the system should detect which products it can process when new data is available. At the end this system should provide permanent derived products from satellite data as well as change and trend-analysis to monitor the earth.

This paper gives an introduction to the architecture and the webportal of the Siberian Earth System Science to provide Earth Observation products and their metadata. New developments are components of a system to create and monitor permanent earth observation products.

SIBERIAN EARTH SYSTEM SCIENCE CLUSTER

At the current development stage the SIB-ESS-C system comprises a federated metadata catalogue accessible through the SIB-ESS-C Web Portal or from any OGC-CSW compliant client. The Web Portal also contains a simple map-like visualization component, which is currently being extended to a comprehensive visualization and analysis tool. The visualization component enables users to overlay different dataset found during a catalogue search. All data products are accessible as Web Mapping, Web Feature or Web Coverage Services allowing users to directly incorporate the data into their application.

Data and Products

The SIB-ESS-C-infrastructure provides access to data products created during a number of research projects over the last decade (e.g. SIBERIA-I & II, SibFORD). Most of these data products were derived from multiple Earth Observation sensors (e.g. land cover, biomass, forest disturbances, water bodies, phenology) or resulted from model experiments (e.g. Carbon emissions from forest fires). Additional data has been integrated from external collaborators such as the “Land Resources of Russia” collection contributed by the International Institute for Applied System Analysis [1].

Table 1 lists the integrated earth observation data products with specification of source, temporal coverage and spatial resolution. Several data sets are available for multiple consecutive time steps covering central Siberia and beyond. A detailed description of each dataset is available at the SIB-ESS-C Web page (http://www.sibessc.uni-jena.de/).

EO Product	Source	Temporal coverage	Spatial resolution	Partner responsible
Phenology	SPOT-VGT AVHRR	2000-2003 annual	1km & 10km	Center for the Study of the Biosphere from Space (CESBIO), France
Disturbances	MODIS AVHRR ATSR-2	1992-2003 on yearly basis	1 km	Centre for Ecology and Hydrology Monks Wood, UK
Freeze/ Thaw	QuikSCAT	2000-2003	10km	TU Wien, Institute of Photogrammetry and Remote Sensing (IPF), Austria
Water bodies	ASAR WS	2003/2004	150m	TU Wien, Institute of Photogrammetry and Remote Sensing (IPF), Austria
Snow Depth	SSM/I	2000-2003	25km	Center for the Study of the Biosphere from Space (CESBIO), France
Snow Melt	SSM/I	2000-2003	25km	Center for the Study of the Biosphere from Space (CESBIO), France

Land cover	MODIS	2001-2004 annual	500m	University of Wales Swansea, UK
Continuous Field Land Cover	MODIS VCF MODIS LC	2003	500m	FSU Jena, Institut for Geography, Germany
Topography	SRTM / GTOPO	2000	3arcsec<60° N I km > 60° N	Gamma Remote Sensing AG, Switzerland
Burnt severity	MODIS	2003-2008 annual	500m	Space Research Institute of the Russian Academy of Sciences
Clearcuts	MODIS	2005-2008 annual	250m	Space Research Institute of the Russian Academy of Sciences

Table 1 – SIB-ESS-C data products

Architecture

The architecture of SIB-ESS-C is mainly based on the usage of standards to ensure interoperability and the implementation of free and open source software components whenever possible. Technically SIB-ESS-C follows a service-oriented architecture (SOA) approach to be interoperable and adaptive when new developments and requirements arise. The components for data discovery, access and analysis are impleneted as Web services. Theses services are based on international standards and specification published by the Open Geospatial Consortium (OGC), the International Organisation on Standardization (ISO) and the World Wide Web Consortium (W3C). This approach allows users to access the services through a common Web browser or with any other GIS Desktop software implementing these international OGC standards.

In its current development stage the SIB-ESS-C architecture comprises components for data discovery, data access, data analysis, as well as a Web portal for direct user interaction. This interaction is based on OGC-compliant services, which are the connection between the Web portal and the provided information (Fig. 1). Further details on the individual components are given in the sections below.

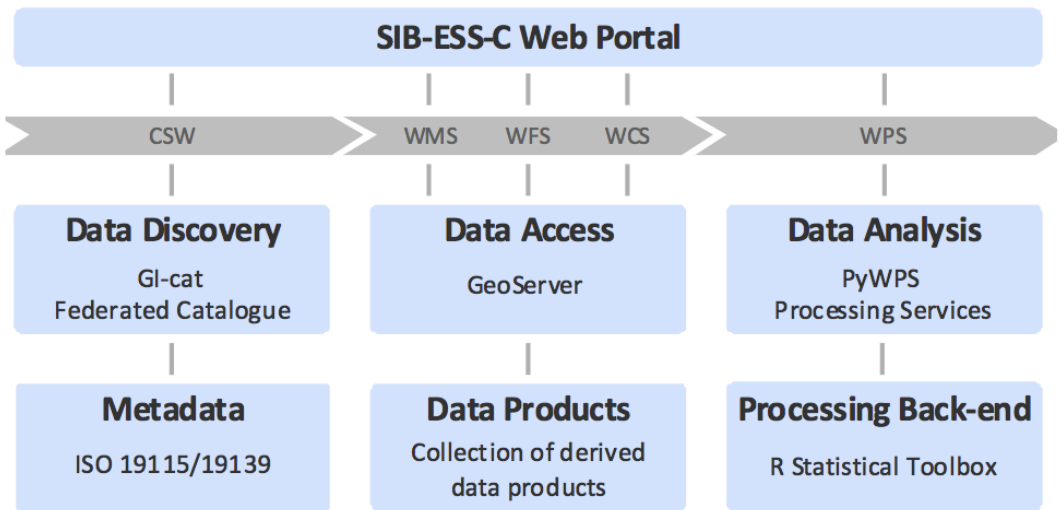

Fig. 1: SIB-ESS-C architecture, its components and interfaces

The developed systems and the architecture are in-line with the Global Earth Observation System of Systems (GEOSS) approach to establish a "system of systems" consisting of distributed but interconnected Earth observation systems [2].

SERVICES

The key components of data discovery and data access are the services published at the service tier. Based on files or databases the service tier providing possibilities to make use of this data. The service tier stands also for interoperability and flexibility in heterogeneous environments customary to interdisciplinary research. Based on standards and specification published by the Open Geospatial Consortium the implementation of several services is necessary for data access and data discovery.

In the following sections the implemtentation of the integrated services is described. For discovering data the Catalogue Service for Web (CSW) [6] is used for providing and querying metadata. This metadata can be described with different standards, e.g. the ISO 19115 standard [7] for describing spatial data and can also be linked to their datasets. For data visualisation the OGC Web Map Service (WMS) [3] is used for displaying raster and vector data as normal images. For real data access (data download) Web Feature Service (WFS) [4] for vector data and Web Coverage Service (WCS) [5] for raster data is requiered.

Data Discovery Service

The SIB-ESS-C catalogue implementation builds on GI-cat, an active open source software project developing a federated catalogue and broker for distributed resource discovery and access [8,9]. With its modular design and pluggable accessors GI-cat supports a number of community standards and specifications, e.g. OGC's CSW/ISO and ebRIM profiles, GBIF, THREDDS/OPenDAP [9]. Thus GI-cat enables users to perform queries on external catalogues and in turn allows other registries to harvest information of all services federated into GI-cat. Federated resources are internally mapped to the ISO 19115 data model [9]. This service is capable of publishing metadata, but also functioning as a mediator between different metadata models on both, the server and the client side. These datasets are described with metadata including a link to OGC-compliant visualization and downloading service.

Data Access Service

Accessing data follows a Service Oriented Architecture (SOA) approach and introduces a service layer between the client and the data level providing standardized interfaces to a number of data types and formats underneath (Fig 2). Such Web Services enable clients to retrieve portions of a server's data holdings based on spatial constraints and other criteria. Thus users retrieve only data they really need and not files of arbitrary size defined by a provider. The implemented services offers three different access possbilities conforming to OGC standards. All gridded data is distributed through Web Coverage Services whereas vector data is published through Web Feature Services. Additionally, both data types are available as Web Map Services delivering a portrayed representation of the actual data values and features.

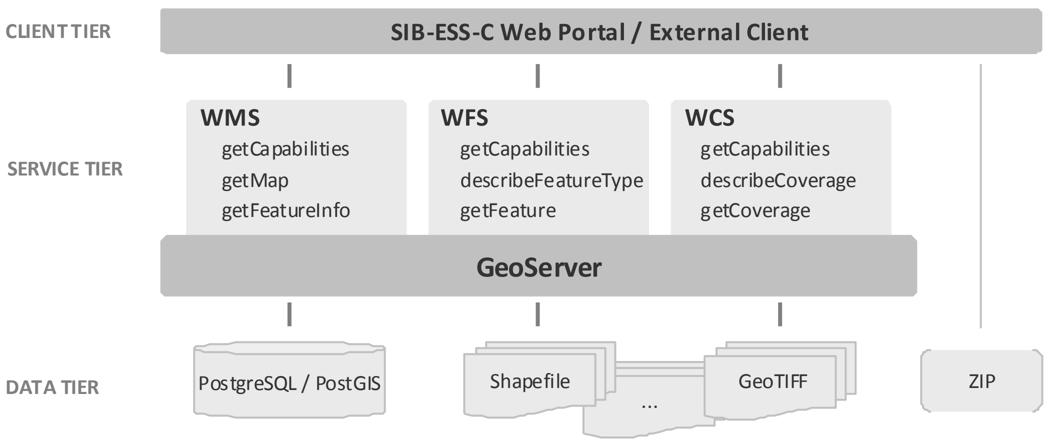

Fig. 2: Data access services of SIB-ESS-C showing the 3 tier architecture and OGC web services including their operations available from GeoServer.

These services are provided from the open-source project GeoServer (http://www.geoserver.org). Each of the services has different methods, which the user can call like getCapabilities to show the available datasets and other important information on the whole service. The spatial data can be stored as files (e.g. formats like ESRI Shapefile or GeoTIFF) or in a spatial database like PostgreSQL with PostGIS extension.

The datasets provided by these services are linked in metadata documents integrated in the discovery service. A user can search within the discovery service and gets the available service addresses from GeoServer so that he can visualize or download the searched data, either in the Web portal or in his own (online or offline) GIS software.

Data Analysis Service

The implementation of the analysis component adopts the overall design strategy of the SIB-ESS-C infrastructure by utilizing standards, open source technology, and well established tools and methods. The creation of graphs is based on a client-server architecture using standard compliant interfaces for interaction. The central element in this architecture is the PyWPS server (http://pywps.wald.intevation.org/). With the aid of this open source package, written in Python, processing functionality is published via an OGC Web Processing Service (WPS) [10]. The WPS functions as a wrapper providing a standard interface to a processing engine underneath. The back-end integrated here is the statistical package "R" (http://www.r-project.org/) with its powerful analysis and visualization capabilities. Alternatively, other existing packages could be used; for example GRASS GIS for which PyWPS provides native support already. The generic framework developed with PyWPS and R is easily extendible with additional functionality (e.g. plot types).

From a user perspective the time series analysis functions are accessible from any client capable of specifying and transmitting an execution request to the server. One such client is the SIB-ESS-C Web Portal assisting users in formulating the processing request. The request is sent using HTTP GET or POST protocol and contains information like plot type, input data, and additional parameters specific to each service. To ensure interoperability in terms of data access, the input data is linked in via Web Coverage or Web Feature Services allowing also disparate WCS and WFS servers to be incorporated. The users selection of the input data sources is supported by the SIB-ESS-C Web Portal and the catalogue search functionality implemented therein. After successful execution the service returns a graphics file to the Web portal. For time-series plots also an XML file with actual data values can be retrieved allowing rich clients to perform customized rendering. The development and testing of the analysis service and its linkages to other services for data discovery and access are still under way and have not been released yet on the SIB-ESS-C Web Portal.

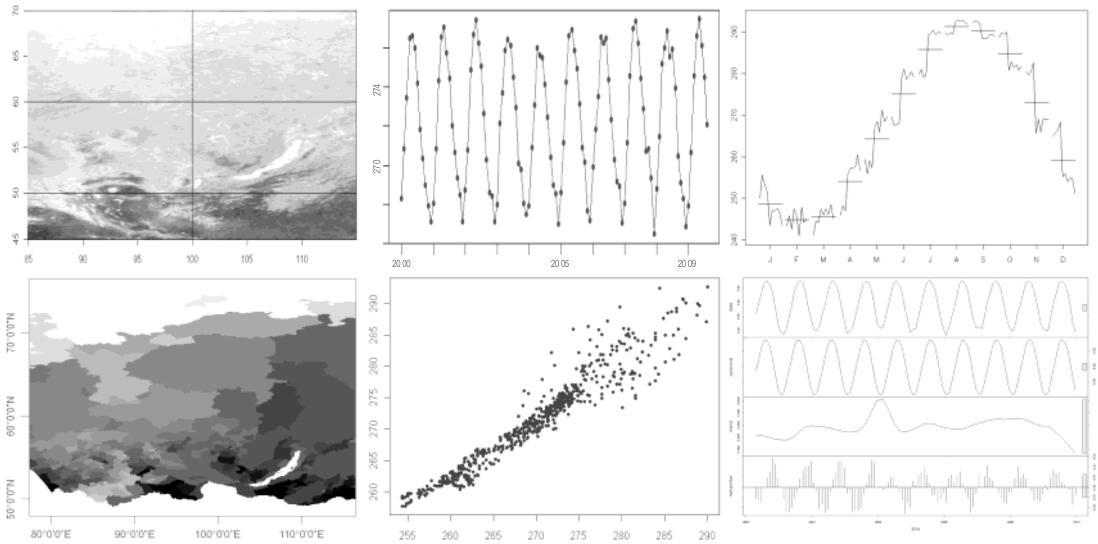

Fig. 3: Samples of SIB-ESS-C data analysis plots

The primary goal of this analysis service is the spatio-temporal exploration of distributed Earth observation time series data. Time series data from remote sensing represents a continuous measurement in space and thus differs considerably from station data (e.g. runoff measurements) collected at a single location. Visualizing and exploring the spatial and temporal characteristics of these remote sensing time series datasets typically requires aggregation in time, space, or both. The analysis tools available from SIB-ESS-C contain both ways of aggregation depending on the visualization type selected. There are two-dimensional (2D), map-like visualizations of a user specified area of interest that are averaged across the temporal range selected, delivering, for example, an annual mean value for each pixel. In contrast, SIB-ESS-C offers visualization along a timeline that is averaged in space according to the user defined area of interest. A few sample plots can be seen in figure 3.

SIB-ESS-C WEBPORTAL

The Siberian Earth System Science Cluster contains a Web-based client for human interaction with the Web-services described earlier in this paper. Web-services are typically targeted at machine-to-machine interaction, hence an additional user interface designed for human interaction is required for such systems. The SIB-ESS-C Web Portal is the central access point to all services provided by SIB-ESS-C servers, but can also integrating federated services from external resource providers, given that those services comply with OGC standards (e.g. CSW, WFS, WCS, and WMS). Some screenshots of this Web Portal can be found in following figure 4. The technical implementation is described below.

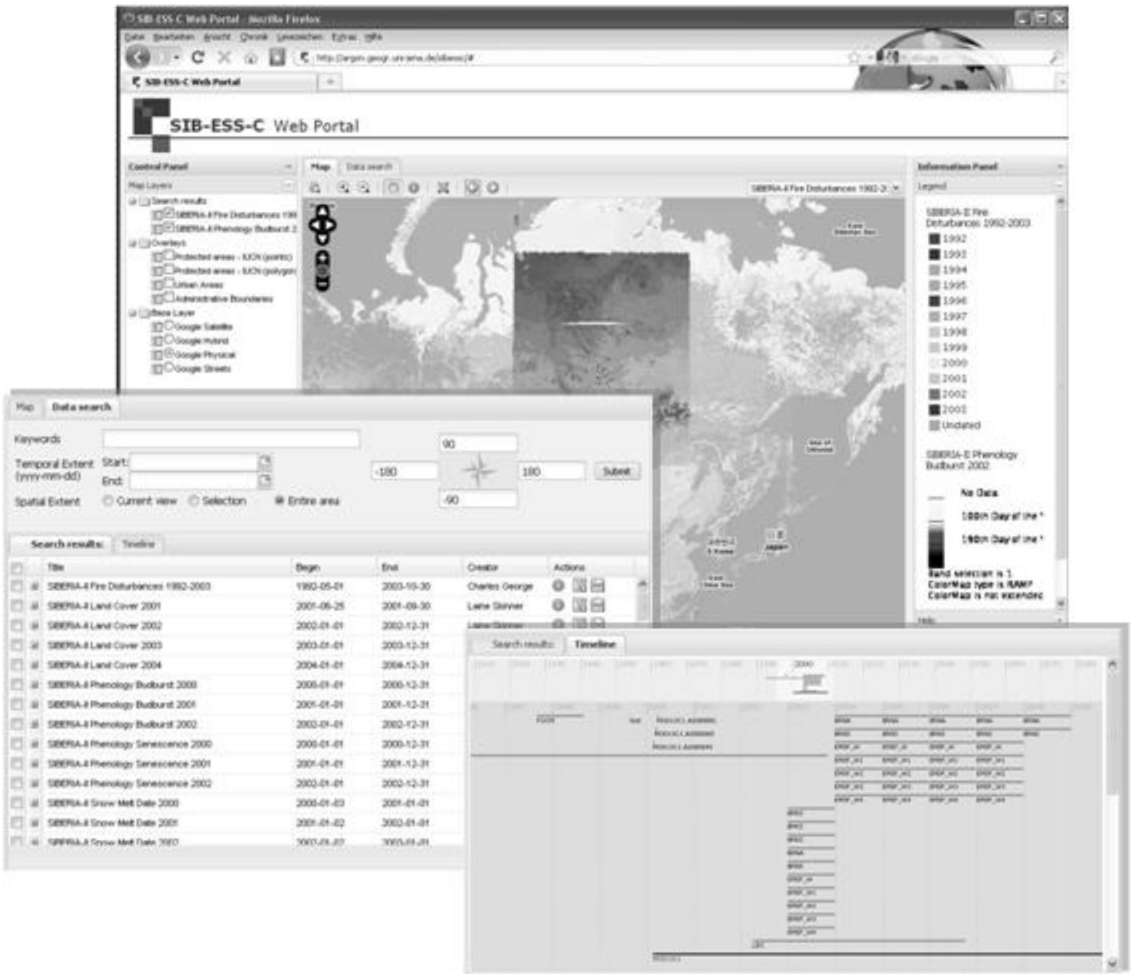

Fig. 4: View panels of the SIB-ESS-C Portal: map view (top), search form and results (left), timeline view (right)

The Web portal offers functionalities for data search, data access, data visualisation and analysis. The technical implementation of the Web Portal is fully in line with the overall design strategy of the SIB-ESS-C infrastructure outlined earlier. The portal integrates standard compliant Web services (e.g. OGC CSW, WMS, WCS), has a modular, extendible architecture, and builds on existing free and open source toolkits. The dynamic and interactive user interface has been achieved by utilizing the MapFish framework (incl. Openlayers, GeoExt, ExtJS), the SIMILE Timeline Widgets, and additional

JavaScript/AJAX code developed in-house. These state-of-the-art Web technologies contribute to a rich, user-friendly Web application with desktop-like features, such as, drag & drop between search results and the map's layer tree, grouping and sorting of columns in the search result table, interactively adding and removing layers in the map, context menus upon right click with the mouse, or adjustable map layer opacity for visual comparison of overlaying datasets. All these tools and features of the Web portal have been implemented with the objective of optimizing the user interaction with SIB-ESS-C Web Services.

FURTHER AND NEW DEVELOPMENTS

The current infrastructure provides manually created products, which were manually described by metadata. Products that can be created permanently (e.g. monthly or yearly) have to be updated by hand at the moment, the product-creation as well as the metadata-creation. The workflow for every single timestep of the product-creation is the same, therefore this system should be extended with an automatically running workflow for product-creation. For this several subsystems are planned (figure 5).

The first subsystem is the "Data Registry System" where input data, algorithms and derived products have to be registered and added with additional metadata. The second is the processing system, in which the algorithm has to be integrated to create the desired product. When products were created permanently a third subsystem can analyse these available products, for example for trend analyses or change detection algorithms. In the following sections these three subsystems are presented.

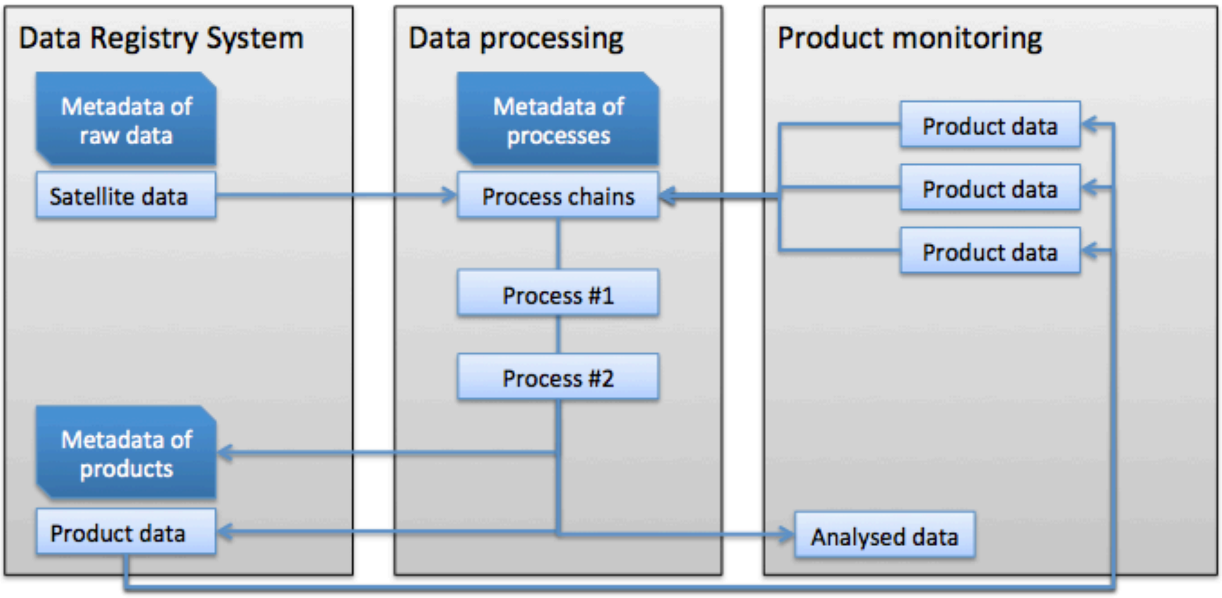

Fig. 5: Interaction of system to derive and monitor products from processed input data

Data Registry System

For creating products data is needed as input data. For the derivation of earth observation products mainly satellite data as well as maybe additional auxiliary data is necessary. To know which data is available for processing this input data must be described with additional information like the available bands with which kind of spectral information and the resolution of the data. This information can for example be described standard-compliant with part two of ISO 19115 titled "Extensions for Imagery and Gridded Data" [11].

On the other side the algorithm integerated in the Data Processing System must provide conditions for the suitable input data, which can be found in the metadata of the registered input data. If both are described detailed enough the system know which algorithms can be executed at new incoming data. Furthermore the system knows if data is available for any registered algorithm, for example if a user requests a specific product for a desired region. After product creation the product will also be registered within this system and additional metadata is added. For product metadata the lineage information is then very important to pursue the process of creation.

Data and algorithms will be connected with information provided as metadata. This metadata system will than be mapped to metadata standards and integrated in the Data Discovery Service for querying purposes. In this interactive system the user should query this discovery service to get information about the available input data in an area of interest as well as the available algorithms the user can execute himself on the Web portal.

Data Processing System

When data and suitable algorithms are registered, the system can executed the algorithm integrated in the "Data Processing System". Thus the whole system is based on standards from the OGC also the processing system deals with a specification from the OGC – the Web Processing Service (WPS) [10]. Within a WPS the algorithm can be implemented in the programming language of the software implementation or, alternatively, as executable binary which is than ran like any other command line execution. For analysing satellite data the integration of (partly commercial) software like IDL, Ecognition, GRASS GIS, GAMMA and ArcGIS Server is planned to test.

Working with complex processings the algorithm can be split up in several processes. To connect these individual process to one complex process orchestration engines are necessary to provide this processing as workflow. A few software components are available to orchestrate OGC-compliant Web Processing Services like the 52 North Orchestration Engine and Apache ODE. The second component follows an IT-standard named "Business Process Execution Language" which is a complex orchestration engine implementing logical elements like if-conditions and loops.

Dealing with automatically processing the algorithms have to be robust and applicably on more than only one dataset. For the processing it is necessary to have a structured data management to handle the exchange of large data to processes within a workflow.

Data Monitoring System

Earth observation products can be created every time the satellite delivers new data. If a product is available on several times change analyses can be done on this product. Theses change analyses are for example the calcuation of trends or detection of changes within the products and can be integrated as processes in the Data Processing System. Every time the system creates a new product, registered analyses will also be executed. After processing the changes can be visualized in the Web portal or any other GIS destop software or downloaded to the users desktop. A further step is also the integration of threshold values within the changes to create automatically an alert message (e.g. via email).

This system shall go in-line with the Global Monitoring of Environment and Security (GMES), a European programme for the establishment of a European capacity for Earth Observation [12]. Goal of the "Data Monitoring System" is the permanent observation of implemented products to monitor specific regions. Therefore satellite data to produce these products have to be constantly available.

SUMMARY & CONCLUSION

In this paper we provided the conceptual overview of the Siberian Earth System Science Cluster and described the capabilities and services available. SIB-ESS-C has been implemented as an interoperable spatial data infrastructure to facilitate distributed multidisciplinary research in Siberia. With its standard OGC Web services for data discovery (CSW), data access (WMS, WCS, WFS) and data analysis (WPS) the system can be integrated into a broader context, such as GEOSS.

Future developments of SIB-ESS-C will focus on improving performance and usability, extending data holdings by in-house processing of relevant Earth observation products, extending the analysis and visualization capabilities, Web service orchestration, and interoperability experiments with similar systems and services.

A main new part will be the automatical derivation of earth observation products and the monitoring services at the same time. Therefore robust algorithms for the desired products have to be found and implemented as Web services. Another important factor is the description of input data and algorithms to provide a direct and automatical link between themselfes.

REFERENCES

[1] – V. Stolbovoi, I. McCallum:. CD-ROM Land Resources of Russia. http://www.iiasa.ac.at/Research/FOR/russia_cd/guide.htm (2002)

[2] – GEO: Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan. ESA Publications Division, the Netherlands. (2005)

[3] – OGC: Web Map Service Implementation Specification, Version 1.1.1. OGC 01-068r3. (2002)

[4] – OGC: Web Feature Service Implementation Specification, Version 1.1.0. OGC 04-094. (2005)

[5] – OGC: Web Coverage Service (WCS) Implementation Specification, Version 1.1.1c1. OGC 07-067r2. (2007)

[6] – OGC: OpenGIS® Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile. OGC 07-045. (2007)

[7] – ISO: Geographic information - Metadata (ISO 19115:2003). DIN CEN. (2005)

[8] – L. Bigagli, S. Nativi, P. Mazzetti, G. Villoresi: GI-Cat: a Web service for dataset cataloguing based on ISO 19115. In Proceedings of the 15th International Workshop on Database and Expert Systems Applications (DEXA'04), Saragozza, Spain, 30 August – 3 September 2004, pages 846 – 850. IEEE Computer Society Press. (2004)

[9] – S. Nativi, L. Bigagli, P. Mazzetti, E. Boldrini, F. Papeschi: GI-Cat: A Mediation Solution for Building a Clearinghouse Catalog Service. pages 68 –74. (2009)

[10] – OGC: Web Processing Service (Standard), Version 1.0.0. OGC 05-007r7. (2007)

[11] – ISO: Geographic Information - Metadata - Part 2: Extensions for Imagery and Gridded Data (ISO 19115-2:2008). (2008)

[12] – GMES: Global Monitoring of Environment and Security – Home. http://www.gmes.info (2011)

Acknowledgement: The development of Siberian Earth System Science Cluster has been funded by the Friedrich-Schiller-University. Additional support was provided by the Italian National Research Council.

Author's biography: Jonas Eberle, Master of Science in Geoinformatics with focus of remote sensing and spatial data infrastructures and Bachelor in Applied Computer Sciences with focus of environment informatics. He did his Master thesis at the German Aerospace Center (DLR) within the Remote Sensing Data Center concerning the creation of process chains to derive products from MODIS satellite data. He is working since several years as web developer with focus on Content Management Systems and user-friendly frontends.