CANADIAN FOREST GEOSPATIAL DATA INFRASTRUCTURE

David G. Goodenough[1, 2], Hao Chen[1], Liping Di[3], Andrew Dyk[1], Aimin Guan[1], Geordie Hobart[1], Wesley Gilbert[1] and Randall Sobie[4]
[1]Pacific Forestry Centre, Natural Resources Canada, Victoria, BC
[2]Department of Computing Science, University of Victoria, Victoria, BC
[3]George Mason University, Greenbelt, Maryland, USA
[4]Department of Physics, University of Victoria, Victoria, BC
506 West Burnside Road, Victoria, BC, Canada, V8Z 1M5
dgoodeno@nrcan.gc.ca

ABSTRACT

A Canadian Forest Geospatial Data Infrastructure has been developed for managing, archiving and distributing remote sensing data and derived information products. The System of Agents for Forest Observation Research with Advanced Hierarchies (SAFORAH) is a grid-enabled environment that makes optimum use of distributed data storage facilities and transparently presents these data sets across multiple organizations and government agencies as one single data archive to users. The SAFORAH implementation was built with an extensible data grid network based on Globus Toolkit 4, GMU's Grid-enabled OGC services, MDA's CUDOS metadata management software and customized software operating over a high bandwidth network. The Grid-enabled OGC services follow the Open Grid Service Architecture and Web Service Resources Framework. The OGC portals are fully compliant with Open Geospatial Consortium (OGC) Standards. Authorized users and geospatial information systems can access current and consistent Canadian forest information in the SAFORAH data grid by using the Grid-services or going through the OGC portals or CUDOS FGDC route. Currently, four Canadian Forest Service centres (Victoria British Columbia, Cornerbrook Newfoundland, Edmonton Alberta, and Québec City), Environment Canada's Canadian Wildlife Service in Ottawa, and the University of Victoria (UVic) are operationally connected to the SAFORAH data grid. The connection to a large petabyte storage facility under development at UVic was also established. With respect to earth observation data for collaborative research on forests, SAFORAH supports the management and distribution of data and derived products from many spaceborne and airborne optical and radar sensors.

INTRODUCTION

The Canadian Forest Service (CFS) of Natural Resources Canada (NRCan) has national and international obligations to report on the state of Canada's forests which comprise 10% of the world's forests. These reporting obligations are met through such programs as: the National Forest Inventory (NFI); national forest carbon monitoring, accounting and reporting; and Earth observation for sustainable development. A large amount of earth observation data (EO) data, collected by CFS, NRCan, is stored at various locations across the country. These data sets are vitally important to the study of Canada's forests, thus facilitating research in support of our national forest monitoring activities. The challenge of creating an infrastructure to effectively manage and seamlessly share this very large and distributed set of EO data presents a unique set of problems.

New efforts and programs leading to a greater pooling of EO resources are being developed to enhance Canadian Government's collaboration framework and enhance the government's return on investment of EO data through increased utilization due to ease of data access. In 2002, we developed an online data management system, the System of Agents for Forest Observation Research with Advanced Hierarchies (SAFORAH). The initial implementation of the system was to build a dynamic virtual organization (VO) in a grid-enabled networking environment with Globus Toolkit 2 and employed MDA's Catalogue and User Data Ordering System (CUDOS) (MDA, 2003) as a Web portal to allow users to access EO data collected by CFS. SAFORAH transparently presents the distributed data storage facilities across multiple organizations and government agencies as a single archive to users. Four Canadian Forestry Centres, CFS Victoria, CFS Cornerbrook, CFS Northern and CFS Laurentian, and the petabyte data storage at UVic were connected initially to the original SAFORAH data grid.

GeoConnections (http://www.geoconnections.org/) is an initiative led by the Canadian government to link location-based information using the power of the Internet to increase access to geospatial data for the benefit all

Canadians and private enterprise. GeoConnections uses international standards, such as the Open Geospatial Consortium (OGC, 2007), in the Canadian Geospatial Data Infrastructure (CGDI) for spatial data interoperability and web services. CGDI includes hundreds of location-based databases throughout the country, such as Atlas Canada, GeoGratis, GeoBase and the National Forest Information System (NFIS). To increase the exposure of comprehensive and integrated EO data in SAFORAH, it is essential for SAFORAH to take advantage of CGDI's capabilities by leveraging OGC technology, data and services. In this paper, we introduce the new grid-enabled OGC Web services that have been implemented in SAFORAH. With the SAFORAH OGC portals available, the system provides standard OGC Web service interfaces to CGDI users and other OGC enabled geospatial information systems. The upgraded SAFORAH system now provides grid-enabled OGC services for various EO data and information products through a Grid-enabled Web Map Service, Grid-enabled Web Coverage Service, and Grid-enabled Catalog Service for Web.

SAFORAH SYSTEM ARCHITECTURE

Grid computing is a technology aiming to provide seamless and scalable access to wide-area distributed computing resources with enhanced performance. The SAFORAH project utilized grid technology to build a data grid that enables the sharing, selection and aggregation of a wide variety of geographically distributed computational resources, such as supercomputers, compute clusters, storage systems and data sources, and present them as a single and unified resource to users. This distributed grid architecture assists users and researchers working on large-scale national forest applications and EO data intensive computing projects to store, catalogue and share their data.

The SAFORAH grid consists of grid services which comply with Open Grid Services Architecture (OGSA) and Web Services Resource Frame Work (WSRF). Any SAFORAH user, who is a member of a SAFORAH virtual organization (VO), can deploy and run these grid services to access EO data and related information products. A grid service in SAFORAH can also securely invoke or be invoked by a grid service in another VO if it is authorized.

A set of OGC Web portals and a FGDC portal through the MDA's CUDOS were developed and deployed to aid CGDI users, general public users and other geospatial information systems in accessing SAFORAH grid services and EO data. These users or systems can transparently use these grid services and access EO data through the portals without any knowledge of which VO is the home of the data repository.

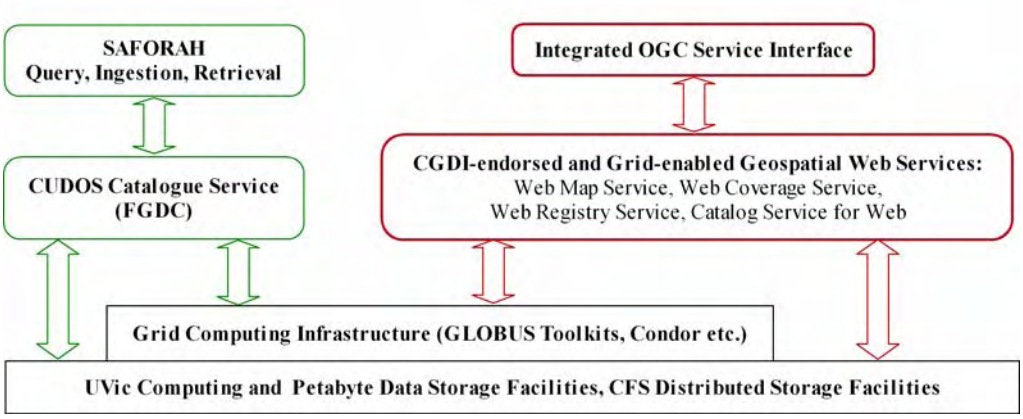

Figure 1. Architecture of the SAFORAH system with new OGC services.

Figure 1 is a high level representation of the newly implemented SAFORAH OGC services architecture. The black rectangles represent the Grid computing environment with distributed storage facilities and computing resources. The red rectangles are the grid-enabled Web services, including the integrated OGC services portals. The green rectangles are the metadata-catalog system, CUDOS, which is used as another Web portal to the SAFORAH data grid. The CUDOS metadata standard follows the Federal Geographic Data Committee (FGDC) Standards (FGDC, 2002).

SAFROAH DATA GRID IMPLEMENTATION

The Globus Toolkit (http://www.globus.org/), developed by the Globus Alliance, provides open source middleware, which was used to create the grid-based application environment. SAFORAH was originally built with Globus Toolkit 2 (GT-2), containing only the pre-Web-service components. The recent evolution of grid technologies to enhance interoperability through virtualization and service oriented architecture (SOA) to support the grid-service standards was the one of the driving forces for us to move from GT-2 to GT-4. Many grid services in the Globus Toolkit 4 (GT-4) meet most of the abstract requirements in OGSA and are implemented on top of WSRF, allowing dissimilar grid applications to communicate with each other. The migration from GT-2 to GT-4 allowed for the enhancement of the SAFORAH data grid to support newly developed grid-enabled OGC Web services.

Standard Geospatial Grid Services and Interfaces

Some of the initial geospatial grid service implementations were developed by the Centre for Spatial Information Science and Systems (CSISS) at George Mason University (GMU) (Chen, 2006). The services were designed to make remote sensing data available to GIS users by utilizing the OGC interoperability protocols. The following service components were developed and deployed in the SAFORAH GT-4 environment.

The Grid-enabled Catalogue Service for Web (GCSW) securely provides archiving, publishing, managing, and querying of geospatial data and services and facilitates transparent access to the replica data and related services under the grid environment. The information model of GCSW is based on OGC CSW (OGC, 2004 & 2005) and the standard service interface was based on the OGC interface standard: getRecord. Any geo-referenced EO data or information products can be registered with GCSW and served out by the Grid-enabled Web Coverage Service or Grid-enabled Web Map Service in the data grid.

The Grid-enabled Web Coverage Service (GWCS) provides access to all distributed EO data and information products registered with GCSW and managed by the SAFORAH data grid. The implementation of three standard service interfaces was based on the OGC WCS interface standards (OGC, 2003): getCapabilities, getCoverage and describeCoverage.

The Grid-enabled Web Map Service (GWMS) responds to users' rendering requests and dynamically produces static maps from distributed EO data, which are registered with GCSW and managed by the SAFORAH data grid. The implementation of the standard service interfaces was based on the OGC WMS interface standards (OGC, 2006): getCapabilities and getMap.

The above grid-enabled OGC compliant services can work both independently or collaboratively to allow authorized grid users to access and interact with the services directly. Other OGC clients or geospatial information systems may also get access to these grid-enabled services and EO data registered with GCSW through the SAFORAH OGC Web portals.

Integrated Grid Architecture with Grid-enabled OGC Services

Figure 2 below presents a simplified overview of SAFORAH's integrated grid architecture that supports the OGC services. It consists of the Control Grid Service (CGS), Replica Location Service (RLS), Replica and Optimization Service (ROS), Monitoring and Discovery Service (MDS), GCSW, and a series of GWCS and GWMS.

To retrieve data or invoke a service from the data grid, a user's request is sent to CGS. If the request is for real physical data, then the CGS directly forwards the request to the data serving services (GWCS/GWMS). However, when the query is dealing with a logical file name that has duplicated physical datasets located at different grid locations, CGS needs to query RLS and MOS. RLS is a distributed registry service that tells where and how many replicas of the datasets are available. RLS returns to CGS a list of the physical file names along with their host names and service IDs. Next, CGS uses ROS to determine which host node has the lowest workload and therefore, the fastest response time to serve the query request. ROS in turn uses MDS to provide CGS with the information about available grid nodes and their workload status to help CGS choose which node will respond the fastest. The physical filename from the selected node is then used to replace the initial logical filename in CGS. After that, CGS uses the host name and service ID, collected from RLS and ROS, to query GCSW for the grid service URL of the chosen host node. GCSW queries its database and defines the appropriate interface between the web-based client and the grid node for CGS. Finally, CGS packages all parameters and makes a grid service call to query the selected GWCS/GWMS host node. The resulting URL for accessing the required EO data will be returned to CGS, which then passes it back to the user (or displaying the link in the portal).

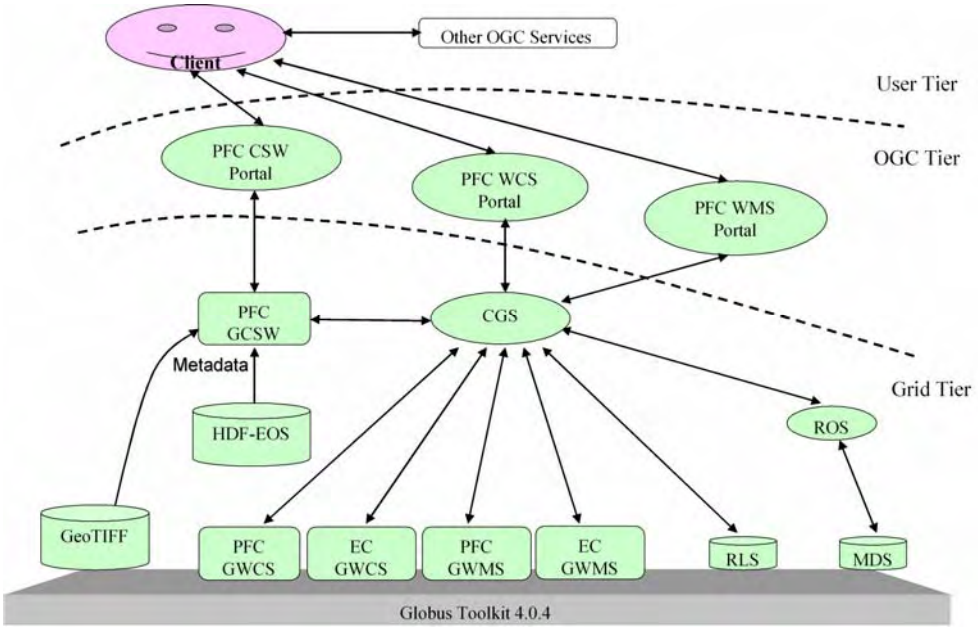

Figure 2. Integrating Grid technology with OGC Services.

OGC User Interfaces

SAFORAH makes use of the standards-based OGC technologies that presently comprise the Canadian Geospatial Data Infrastructure and facilitates EO data and information products sharing. Three OGC user interface portals were built to allow OGC users and OGC-enabled geospatial information systems to access EO data and services in the SAFORAH data grid. The portals first convert standard OGC requests to Grid-enabled service requests and then un-wrap Grid service responses back into standard OGC service responses. Thus, OGC users can seamlessly access all the distributed EO data in the SAFORAH grid without any knowledge of the underlying structure of the system.

The OGC interface portals were developed for registering, querying and serving EO data. The implementation was based on the OGC Web Data Service Specifications, including Web Coverage Service v1.0.0 (WCS), Web Map Service v1.3.0 (WMS), and Catalogue Service for Web v2.0 (CSW). The OGC Web services standards provide standard user interfaces in both Key-Value Pairs (KVP) and the XML document format. SAFORAH uses the standard Web-based SOAP protocols to send a GET or POST request and return the results for all the OGC services. The KVP standard format is for the GET method. The XML format is for POST and SOAP methods. All the service responses are returned in the XML format.

The SAFORAH OGC Web services can be accessed through the following links:

http://www4.saforah.org/WMSPortal/

http://www4.saforah.org/WCSPortal/

http://www4.saforah.org/GridCSFPortal/

Catalogue and User Data Ordering System (CUDOS)

In addition to the OGC user interfaces, MDA's CUDOS was modified and used as another Web portal to provide an alternative way for non-OGC users to access EO data in the SAFORAH data grid. CUDOS is an online data catalogue system initially developed by MacDonald, Dettwiler and Associates Ltd (MDA) for improving EO data operations at MDA's satellite ground stations. The main functions include the EO metadata catalogue, data archiving and ordering. In SAFORAH, CUDOS was modified to become a portal to the SAFORAH data grid. CUDOS provides metadata management, EO data dissemination and user authentication and has the ability to securely access distributed EO data in the SAFORAH data grid in a seamless and transparent fashion to users. Through the CUDOS portal, authorized users can easily query the EO metadata database, and ingest and download EO images to/from the SAFORAH data grid.

The EO metadata contents in CUDOS were designed to follow the FGDC standards. A Z39.50 server (http://www.loc.gov/z3950/) was implemented in CUDOS to allow other FGDC information systems to search and retrieve metadata information from SAFORAH.

The web address for the CUDOS portal is http://www.saforah.org.

SAFORAH Grid Credentials

The SAFORAH data grid uses digital certificates signed by Grid Canada, which is a partnership between CANARIE (http://www.canarie.ca/), the National Research Council (http://www.nrc-cnrc.gc.ca/), and C3.ca (http://c3.ca/). By using the Grid Canada credentials, the SAFORAH data grid is able to utilize the CANARIE network infrastructure (Ca*Net-4), to facilitate VO linkages between the Canadian government, universities and scientific institutions. Thus connecting the computing resources and data storage facilities of various grid VOs connected to CANARIE.

SAFORAH APPLICATIONS AND IMPACT

For the SAFORAH data grid, GT-4 was installed on two Sun Solaris machines at CFS's Pacific Forestry Centre, various Linux computer nodes at the other three CFS Forest Centres and an IBM P670 machine at UVic, which is connected to the UVic petabyte data storage facility. Figure 3 shows the implementation of the SAFROAH data grid. Currently, CFS Victoria, CFS Cornerbrook, CFS Northern, CFS Laurentian, National Wildlife Research Centre of Environment Canada, and UVic are operationally connected to the SAFORAH data grid. The grid connections to the Canadian Space Agency (CSA) in Saint-Hubert, Quebec, and the Research Branch of Agriculture and Agri-Food Canada (AAFC) in Ottawa are currently being implemented and will be operational in the near future.

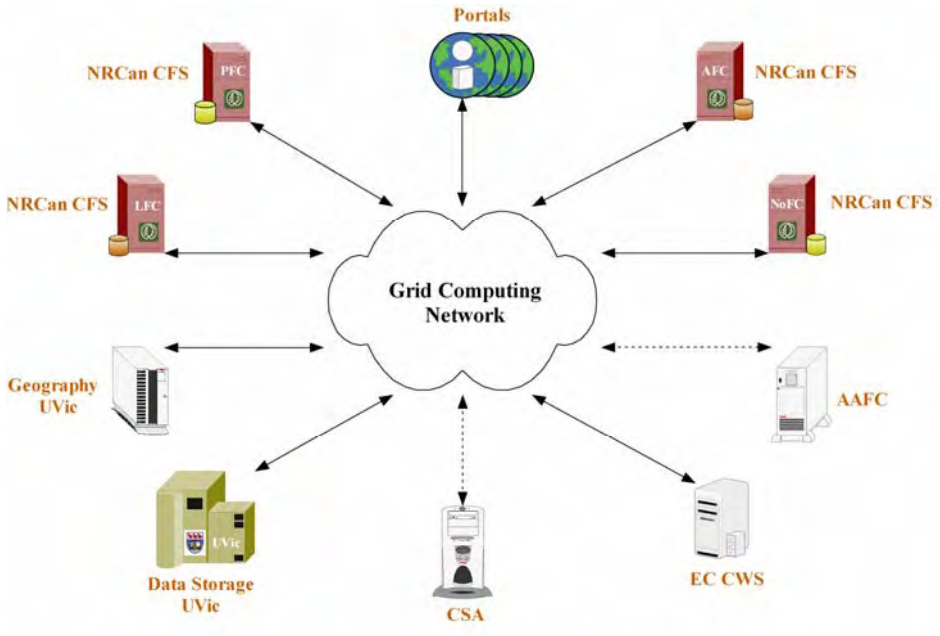

Figure 3. SAFROAH Data Grid.

The research and development of SAFORAH provides direct support to collaborative research of national programs by sharing EO data and information products within the Canadian government, universities and scientific institutions. One of the applications is the Earth Observation for Sustainable Development (EOSD) of Forests, a joint program between NRCan CFS and CSA for the development of a forest monitoring system for Canada (http://www.pfc.forestry.ca/eosd/index_e.html). The EOSD project has acquired 1700 Landsat TM/ETM+ images of Canada, which were used to create 630 EOSD landcover NTS map sheets. All these data and derived products are freely available to the public through the SAFORAH data grid via OGC services and CUDOS. Since many Canadian provincial and territorial mapping agencies recognized the importance of such mapping of Canada's forests, some have initiated mapping programs of their own by downloading and utilizing Landsat data and EOSD Landcover images through SAFORAH.

CONCLUSIONS

SAFORAH was built with GT-4 over CANARIE broadband networks to provide seamless, scalable and secure access to a wide variety of geographically distributed computation resources across Canada, such as data storage facilities, supercomputers, computer clusters and EO data sources. SAFORAH supports EO data collaboration for some Canadian national forest monitoring programs in a dynamic virtual-organization environment as part of Grid Canada. The recent implementation and deployment of GCSW, GWCS and GWMS in the grid-enabled environment is a first step. The next phase for SAFORAH will be the evolvution of real time EO data processing by using the supercomputer facilities and petabyte data storage utilities at UVic. SAFORAH provides the OGC standard portals, including WMS, WCS and CSW, to enhance information interchange with other geospatial information systems, including CGDI.

SAFORAH has enabled CFS researchers to effectively manage EO data and derived infromation products with a minimum of effort. It provides a practical and efficient tool for the SAFORAH partners and public to access EO data and forest information products. The expansion of SAFORAH to support OGC Web Services, significantly increased the accessibility and interoperability of remote sensing data collected by EOSD. The OGC services promote the use of the EOSD information products by various stakeholders and CGDI users. Through the SAFORAH implementation, CFS and its partners have made significant contributions to Canadian interests in forest applications and ensured availability of EO data and information products for the benefit of government organizations, scientific researchers, public users, and private enterprise.

ACKNOWLEDGMENTS

The participating partners in the SAFORAH project include CFS NRCan, CSA, GeoConnections, UVic, GMU, MDA, CITS, Environment Canada, AAFC and the Canada Foundation for Innovation. We thank CFS, CSA and GeoConnections for funding the development of SAFORAH. We thank the CSISS technical team and UVic grid team for their technical support to the SAFORAH project. We appreciate the support of the high-bandwidth connectivity from CITS. DGG acknowledges the support at UVic from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Chen A., L. Di, Y. Wei, Y. Bai, and Y. Liu (2006), Grid-enabled Standard-compliant Open Computing Environment for Earth Science Exploration and Applications, *Proc. IGARSS 2006*, pp. 237-240, Denver, Colorado, USA.

FGDC (2002), Geospatial metadata - Federal Geographic Data Committee Standards, *USA.gov*, *http://www.fgdc.gov/metadata*

MDA (2003), Catalogue and User Data Ordering System. *MacDonald, Dettwiler and Associates Ltd.*

OGC (2003), Web Coverage Service (WCS), Version 1.0.0, *Open Geospatial Consortium Inc.*: *http://www.opengeospatial.org/standards/wcs*

OGC (2004), OpenGIS Catalogue Service Specifications 2.0 – ISO19115/ISO19119 Application Profile for CSW2.0, *Open Geospatial Consortium Inc: http://www.opengeospatial.org/standards/cat*

OGC (2006), Web Map Service Interface, Version 1.3.0, *Open Geospatial Consortium Inc*., *http://www.opengeospatial.org/standards/wms*

OGC (2005), Catalogue Service, Version 2.0.0**,** *Open Geospatial Consortium Inc***.,** *http://www.opengeospatial.org/standards/cat*

OGC (2007), OpenGIS Specifications (Standards), *Open Geospatial Consortium* Inc., *http://www.opengeospatial.org/standards*