Grid Technologies for GEOSS

The User and the GEOSS Architecture XXV

Valencia, Spain, December 3, 2008

Craig Lee, President OGF
lee@aero.org

© 2006 OpenGridForum
Geospatial Data: Immense Applicability

Estimates vary, but ~80-90% of all data collected or produced by the human race is geospatially referenced.

- Natural exploration, e.g., oil & gas
- Public administration
- Civic planning and engineering
- Weather and aviation
- Environmental planning
- Disaster management
- Earth Observation systems
An Example of a Global Grid: EGEE

Bob Jones, EGEE Project Director

How to Integrate Grid and Geospatial Technologies?
The OGC-OGF Collaboration

- Promote best practices and international standardization for distributed geospatial data processing capabilities that is:
  - *Transparent* -- users is not aware of the infrastructure
  - *Interoperable* -- the resources work together
  - *Scalable* -- small local, to massive distributed platforms
Researchers face increasingly large repositories of geospatial data stored in different locations and in various formats. To address this problem, the Open Geospatial Consortium and the Open Grid Forum are collaborating to develop standards for distributed geospatial computing.

**NEED FOR STANDARDS**

The ability to access, integrate, analyze, and present geospatial data across a distributed computing environment using common tools has tremendous value. Indeed, with the growing connectivity of our world through data-collecting instruments, data centers, supercomputers, departments, machines, and personal devices such as cell phones, PDAs, and smart phones—as a society we expect a wide range of information to be instantly accessible from anywhere. This expectation is motivated not only by personal convenience; there are also solid business and policy reasons for enabling such distributed geospatial applications.

Achieving such ease of access and interoperation will require best practices that are codified into widely adopted standards. In the context of e-Science, the National Science Foundation’s Cyberinfrastructure Council argues that “the use of standards creates economies of scale and scope for developing and deploying common resources, tools, software, and services that enhance the use of cyberinfrastructure in multiple science and engineering communities. This approach allows maximum interoperability and sharing of best practices. A standards-based approach will ensure that access to cyberinfrastructure will be inkl
Two Issues

• What geospatial and computing technologies must be integrated?

• How do we manage progress to accomplish this?
Some Key OGC Standards

- **Web Map Service (WMS)**
  - Display of registered and superimposed map-like views of information that come simultaneously from multiple remote and heterogeneous sources.

- **Web Feature Service (WFS)**
  - Retrieval and update of digital representations of real-world objects referenced to the Earth’s surface.

- **Web Coverage Service (WCS)**
  - Access to a grid coverage, usually encoded in a binary format, and offered by a server.

- **Catalogue Service for the Web (CSW)**
  - Common interfaces to publish, discover, browse, and query metadata about data, services, and other resources.
Key OGC Standards

- **Sensor Observation Service (SOS)**
  - Access observations for a sensor or sensor constellation, whether in-situ or dynamic (e.g., water monitoring or satellite imaging). Optionally access associated sensor and platform data.

- **Sensor Planning Service (SPS)**
  - An interface to task sensors or models. Using SPS, sensors can be reprogrammed or reconfigured, sensor missions can be started or changed, simulation models executed and controlled.

- **Sensor Alert Service (SAS)**
  - Defines an interface for publishing and subscribing to alerts from sensors. If an event occurs the SAS will notify all clients subscribed to this event type.

- **Observations and Measurements Encoding Standard (O&M)**
  - Defines an abstract model and an XML schema encoding for observations and it provides support for common sampling strategies. It also provides a general framework for systems that deal in technical measurements in science and engineering.
Key OGC Standards

• **Transducer Markup Language (TML)**
  - TML defines the conceptual model and XML Schema for describing transducers and supporting real-time streaming of data to and from sensor systems

• **Web Processing Service (WPS)**
  - WPS defines a standardized interface to facilitate the publishing of geospatial processes, and the discovery of an binding to those processes by clients
Key OGF Standards

- **HPC-Basic Profile**
  - Cluster scheduling interoperability layer supported by Microsoft, Platform Computing, Altair Computing, and others

- **Simple API for Grid Applications (SAGA)**
  - Complete distributed programming environment with C++ and Java bindings

- **GridRPC**
  - Grid-enabled Remote Procedure Call

- **Data Access and Integration specs (WS-DAI-*)**
  - Set of specs for accessing remote files, databases, XML documents

- **Web Services Resource Framework (WSRF)**
  - Allow remote data and services to be independently managed
OGF Data-Related Standards

- **Data Format Description Language**
- **Storage Network Community**
- **Information Dissemination InfoD (GFD 110)**

- **Movement**
  - DMI (GFD 134)

- **Transport**
  - GridFTP (GFD 20)

- **Access**
  - Byte-I/O (GFD 87/88)

- **Storage Management**
  - SRM (GFD 129)

- **Naming**
  - RNS (GFD 101)

- **Access & Management**
  - DAIS (GFD 74, 75, 76)

- **Unstructured Data** (Disk, Tape)
- **Structured Data** (Database)

- **Existing**
- **Possible**
HPC Basic Profile

- **HPC Specific Job Management**
  - HPC Basic Profile (GFD 114)

- **Security**
  - WS-Security (OASIS)
  - SSL/TLS (IETF)

- **Job Management**
  - OGSA-BES (GFD 108)

- **References**
  - Education ISV Primer (GFD 141)

- **Job Definition**
  - Application Description
    - HPC Application (GFD 111)
  - Job Description
    - JSDL (GFD 56/136)
  - File Transfer
    - HPC File Staging (GFD 135)

- **Combines**
  - Compute Resources

- **Uses**
  - Provides Access To

- **Extend**
Running Applications Remotely

- ISV Remote Computing Usage Primer, GFD-I.141
- How to use sets of OGF standards for different scenarios
- www.ogf.org/RemoteExecution
How Do We Manage Progress for Distributed Geospatial Systems?

- **Targeted Projects on Decisive Issues**
- **Build Critical Mass of Key Stakeholders**
  - Continual polling and coordination across the community
- **They must agree on:**
  - Clear Goals
  - Clear Schedule ("time-box" the process)
  - Clear Responsibilities
  - Properly Provisioning the Effort
- **Return on Investment**
  - Time, Money & People $\Rightarrow$ Concrete Results
  - Early influence, skills building, visibility, early deployment in the marketplace
  - *Get more then they put in*
OGC Web Service Testbed

- OWS is an annual process where sponsors identify specific demonstration targets
  - Larger number of participants supply in-kind resources
- OWS-5 (finished March 2008)
  - Sponsors
    - BAE Systems - National Security Solutions
    - Federal Geographic Data Committee (FGDC/USGS)
    - Google
    - Lockheed Martin Integrated Systems & Solutions (Lead Org.)
    - Northrop Grumman
    - US National Aeronautic and Space Administration (NASA)
    - US National Geospatial-Intelligence Agency (NGA)
  - ~35 Participating Organizations
- 3x-4x Return on Investment
  - Strength of Collaboration
Satellite EO-1 tasked to collect imagery on Northern San Diego County Wildfires that was integrated with UAV track data (red lines) in Google Earth. Demo drove issues around sensor networks, data interoperability, and command & control.

OWS-6 Topics

- **Sensor Web Enablement (SWE)**
  - Georeferenced sensors, event notification, security

- **Aviation Information**
  - Aeronautical Information Exchange Model (AIXM) for next generation flight control system (FAA and EuroControl)

- **GeoProcessing Workflows (GPW)**
  - Web Processing Service (WPS), Workflow and Grids (“WPS to Grid”)
  - Distributed resource management, e.g., data, services, hosts, networks, for geospatial applications

- **Decision Support Services**
  - 3D Visualization with Fly Through
  - Open Location Services
  - Integrated Client for multiple OWS services

- **Compliance, Interoperability & Test Evaluation (CITE)**
  - Suite of tests and reference implementations for all approved standards
Many Possible Geospatial/Grid Topics

- Integration, interoperability of geospatial and computing resource
  - Information models (CIM/GLUE)
  - Distributed and federated catalogs (ebRIM, MDS, etc.)
  - Discovery services
- Implementation of WPS on various grid tools
  - SAGA, GridRPC, HPC-Basic Profile
- Integration of WS-Eventing, WS-Notification, INFOD
  - Sensor Web Enablement
- Integration of configuration and lifecycle management tools
  - Configuration Description, Deployment and Lifecycle Management (CDDLM)
- Integration of WPS with workflow management tools
  - BPEL, Kepler, Taverna, Triana, Pegasus, etc.
  - Workflow design tools, execution engines, planning (data virtualization), data provenance
- Integration of WPS with grid security models
  - SAML, XACML, VOMS, GSI
  - Support for virtual organizations
- Integration of geospatial data with rule-based, data management policy engines
  - iRODS
Emerging Technologies!

- Green IT and Green Grids
  - Energy costs are becoming defining issue
    - Beyond costs -- *location feasibility*
  - Green Grids have control structure actually quite relevant to EO and environmental mgmt systems

- Virtualization and Clouds
  - Virtualization of resources at any level
    - Infrastructure, platform, services

**OGF is actively engaging these areas**
- Avenue for GEOSS to stay current
OGF Green VM project

- Effort to coordinate existing projects on virtualization, clouds and green IT under the OGF umbrella
- Organizations currently involved:
  - Atos Origin, IBM Haifa, UCM, ENS Lyon, Intellect UK
- Initial Goals
  - Development of a "reference model" for energy monitoring and policy enforcement engine
  - Enable a "common framework" in which to coordinate projects, e.g., Reservoir, OpenNebula, Green-Net
  - Use VM initial placement and migration, and other mechanisms, to enforce energy policy
  - Coordinate with EU Data Centre Code of Conduct
An Energy Management Arch.

Synergies

• Major Architectural Components
  • Environmental Monitoring
    • For data centers but applicable to other scenarios
  • Decision Making
    • Interpreting environmental data and events against policy
  • Enacting Responses
    • Controlling elements of the data center/grid/cloud

• Strong industry motivation will drive commonality
• Leverage for all manner of distributed control systems
  • e.g., Earth Observation systems
Establishing the Lineage from Grids to Clouds

- Clouds have tremendous industry interest
  - Outsourcing (virtualizing) at any system level
    - Infrastructure, platform, services
  - Primary business model is simple client-provider
  - Huge market potential
- *Multiple clouds* start to look very grid-like!
  - Inter-cloud discovery and interoperability
  - Security in virtual organizations
  - Data access and integration
  - Workflow management, ...
- Grids as Cloud Frameworks
  - Challenge: How to integrate/leverage cloud technology in general distributed systems?
A “Cloudy” Solution Space

OS Virtualization

Parallel Frameworks

Software as a Service

Amazon S3/EC2
RightScale, GigaSpaces, Elastra, 3Tera

Hadoop over EC2
Cohesive

Hadoop
Sun’s Caroline

MapReduce, GFS, BigTable
MS Astoria

Google AppEngine
Mesh

Vendors of high-level application configuration tools are potential “cloud framework providers”

A Host of Cloudy Issues

- Data access and interoperability
  - Will have to be approached at the application domain level, by the domain users involved

- Security
  - Can cloud providers provide sufficient security for sensitive applications?

- Reliability
  - Can cloud providers provide sufficient reliability for critical applications?

- Frameworks
  - How to manage sets of VMs and VOs?
  - These are essentially “cloud frameworks” and call for "cloud framework providers"

- Performance management
  - Clouds tend to abstract away location
  - Managing the compute-data locality (affinity) can be important

- Costing models
  - How to compare your own infrastructure costs with a cloud computer?
  - How to compare two clouds?
OGF Cloud Efforts

- Avoid reinventing multiple wheels …
- Leverage cloud benefits in general distributed systems
- Understand market direction and adoption paths
- Cloud Interoperability Forum “unconference”
  - Organized by Reuven Cohen & Shishir Garg
  - Sept. 24, Orange Labs, South San Francisco
  - What is the cloud core constituency ready for?
    - Informal meetings, organized forum, standards?
  - What are the…
    - Technical drivers? Market drivers? Social network drivers?
- O’Reilly book chapter from OGF
  - Reuven Cohen, enomaly.com, (ed.)
  - Co-authors: Lee, Gannon, Fox & Strong
  - Establishing a lineage from grids to clouds
- OGF-25 cloud workshops
Summary

• Significant work already underway for standard, distributed geospatial computing infrastructures
• Highly relevant to many application domains, including EO!
• Opportunity to leverage the work of key participants through collaboration
• Stay current with emerging technologies
Come Work with Us!

- Digital Repositories: Evolving a community
  - Workshop @ IDCC, Edinburgh, UK, December 1-2, 2008
- Cloudscape Workshop
  - Brussels, January 14-15, 2009
- Grid Interoperation Now workshops

- Grid-to-Clouds Workshop
- Global Lambda Integrated Facility Workshop
- Green IT in Grids Workshop
- Geospatial Workshop
- Workings groups, BoFs, …
Back-ups
Example: Grid in a Box at a Windmill

Green Power is the Future

- wind
- solar
- hydrogen

Sustainable and renewable uninterruptible power supply

- Extensive use of virtualization and fiber optics to put the computing where the power is
- Enable the workload to “follow the wind” or “follow the sun”

Source: www.Bastionhost.com

Bill St. Arnaud
CANARIE

Source: www.Bastionhost.com