A COMPUTATIONAL MAPPING ENGINE PORTAL FOR ACCESSING GEOLIBRARIES

Charles O'Hara and Roger King
Mississippi State University
Mississippi State, MS 39762 USA
cgohara@erc.msstate.edu Fax: 662.325.7692

ABSTRACT-Computational portals can provide user customized front ends to very large geospatial image data archives. The computational portal differs from most portal technology in that it not only provides the user the interface for selecting an area and data type of interest, it can also permit the user the ability to manipulate or transform imagery data from a variety of sensors into a format most efficient for the desired end use. Portal technology allows tasks to be saved and configured as single process or as a sequence of chained processes to conduct advanced analysis on selected data for a defined area of interest. The approach offers advantages over conventional methods because it provides methods of generating and delivering data and information products the computational intricacies of which are totally transparent to the user. Mississippi State University has begun developing a computation portal for accessing a geospatial data library (geolibrary). The Computational Mapping Engine (CME) Portal at MSU began as a conventional interactive geospatial image archive and has successfully implemented computational portal technologies to address challenges in loading, storing, distributing, archiving, transforming, manipulating, and fusing remote sensing image data to facilitate the effective use of these data in research and in real-time decision making processes. This paper presents a general CME Portal architecture that can be used to develop and implement computational portals for accessing and processing remote sensing image data from local or distributed geolibraries.

I. INTRODUCTION

The demand for on-line and archived sources of geospatial image data is enormous and will increase. This demand has outpaced internet bandwidth increases and drives the need for user communities to be provided improved access to custom created data products computationally generated from stored image data.   A good example of the demand for geospatial image data is the TerraServer that was launched in 1998. The Distributed Active Archive Centers (DAACs) of NASA also demonstrate the demand for access to archived data providing users with the ability to query an archive database by data type, geography, and temporal range of interest and then reduce the selected set by reviewing qualitative and quantitative information about the data desired. The TerraServer and DAACs are outstanding examples of conventional image archives that provide access to metadata and the ability to request static data from on-line, near-line, or off-line sources. However, advances in remote sensing, computing, and networking technology have enabled the development of computational portals for accessing archives of geospatial data, requesting custom data products, and
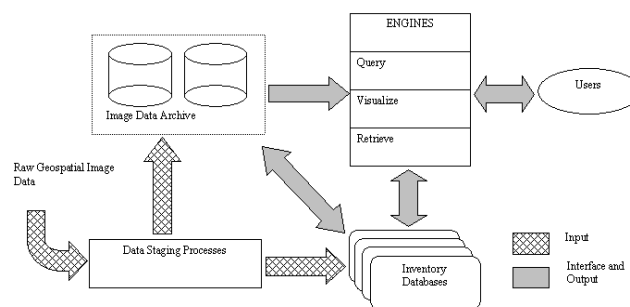

Figure 1. Simplified architecture of a conventional interactive geospatial image archive. (modified from [1])

acquiring the data product through computational processing and automated delivery mechanisms.

A. Conventional Geospatial Image Archive Architectures

Satellite and aerial image data users have conventionally requested data from archives through using interfaces that facilitate mining of archived data. Typical interfaces allow data to be mined through a process that involves selecting image sensor type, geographic areas of interest, temporal range of interest, and perusing reduced resolution "quick look" images that provide a general idea of the visual quality of the data accompanied by attribute information that provide additional qualitative and quantitative details about the data. The general architecture of a conventional geospatial image archive is illustrated in figure 1.

In conventional interactive geospatial image archives, remote sensing image data are staged to the archive, reduced resolution images and ancillary data are generated in the staging process, and the results are stored as images and attribute ancillary information in an inventory database [1]. User queries for data are performed against the data in the inventory database and browsing of image data is performed on the reduced resolution images to narrow the search down to a desired image that is requested. Delivery of requested data can be accomplished through on-line or off-line methods, but for most conventional geospatial image archives, image data for large image data is typically handled by media transfer using tape or more recently CD or DVD media.

B. New Technologies and User Communities

New remote sensing and geospatial data types and the vast technology strides that have been made in computer networks

and computational technologies are enabling the development of web-based capabilities to access staged image data, request custom data products from raw data, and get rapid data or information product delivery. Computational solutions are needed to generate custom data products for timely or near real-time input to spatial decision support systems used by a variety of user communities. Users and user communities for a wide variety of purposes, analyses, and decision-making processes require Geospatial data and products. Data are needed by different users and communities of users in different formats, as different transformation products, and/or fused with other data types to form complex data products.

C. Portal Technologies

In general, web portals provide users with methods to customize their interface to collections of content and services that have been organized into various categories and made available to user communities through links on web pages. Commercial web portals typically offer a broad array of resources and services, such as e-mail, forums, search engines, and on-line shopping malls. Users are able to create a personalized style to their interface that is ideal for them and provides them with desired content.

Portal technologies can provide the ability to design and deliver user interfaces to remote sensing image data archives that are ideally suited to specific user communities, but highly customizable by the individual user. The type of user community would generally dictate the focus, features, functionality, and complexity of the generally styled interface as well as the computational resources available to the user community for creating custom products [2].

Scientific web portals provide user communities with the ability to access (and in some cases stage) scientific data within their domain, to access advanced services, and in some cases to use computational resources for problem solving and analysis. Scientific portal development efforts seek to build improved web-based tools for working with multiple data sources and for using advanced customization features. Ideal portal designs allow users to control their web environment and interact with web-based data sources. These portals are typically configured to provide targeted user communities expandable web-based work environments with the appropriate data, tools, and options to complete their jobs.

Future computational portals to geospatial remote sensing data archives will provide users with the ability to customize their interfaces to the archived data, to select their areas of interest, to select data or data product of interest, to select a specific format that ideally suits their needs, and to use computational resources to generate desired data products [3]. Furthermore, a user's preferences and session histories will be stored to allow future processing to be conducted based on data, preferences, or processes conducted in past sessions [4].

II. COMPUTATIONAL PORTALS FOR ACCESSING GEOSPATIAL IMAGE ARCHIVES

The use of computational portal technologies to provide users and user communities with access to archives of geospatial data adds user information and management, data production management, and computation resources to the architecture of the conventional interactive geospatial image archive. The development of an architectural framework for accessing geospatial image data archives will include the following component architectures (see Figure 2):

- Multi-Tier Architecture
- Data Management Architecture
- Computational Architecture: System Hardware and OS
- User Management Architecture
- Data Product Processing Architecture
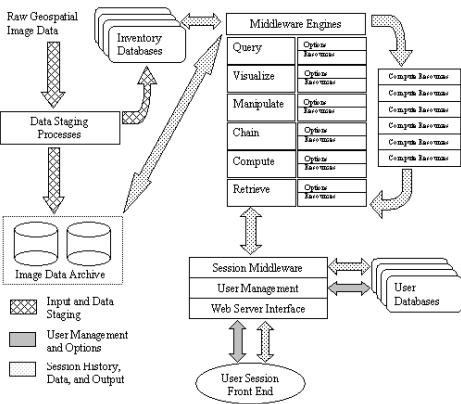- Product Delivery Architecture

Figure 2. Simplified architecture of a computational portal for accessing and processing remote sensing image data.

For each of the component architectures, consideration can be given to such factors as price vs. performance, open standards vs. proprietary formats, commodity approaches vs. limited or customized implementations. The selection of components will determine how quickly envisioned capabilities may be developed, the extensibilities, scalability, and performance of the system, interoperable aspects of the system, and the overall life-cycle cost of the system.

A. Multi-Tier Architecture

A multi-tier (or n-tier) architecture is needed to provide separation among components of the system to facilitate orderly application development, efficient processing, and overall system performance. A multi-tier architecture for a computational portal that accesses remote sensing image data would nominally include a user-interface provided by a front end, middleware that would provide flow of control and security between the user and application, computational resources to handle processing, and one or more back end databases to store system data, user data, and track system session activities. Within each of these tiers there are important choices that determine the overall nature of the application. However, current trends toward open-source architectures increasingly leverage open-source solutions that are available for use in the design and implementation of each tier in the system.

B. Data Management Architecture

In designing and implementing a data management architecture, important considerations arise in the areas of data types and formats that will be supported for staging to the geolibrary, middleware engines that manage and transform data between supported types as needed for storage and processing, and as output data types that may be implemented. Additionally, emerging areas of consideration exist for high-resolution commercial image data that are licensed rather than owned and how licensing impacts user access to access, data product generation, and distribution of derived data products. Data storage and management are also critical considerations in architectural development.

C. Computational Architecture: System Hardware and OS

For scalable commercial and scientific computing tasks, recent trends have leveraged the significant advances that have been achieved in "commodity" servers. Inexpensive "commodity" servers can be configured in compute clusters that share processing tasks and can easily scale by adding additional nodes to the cluster. Scalable storage in the geolibrary requires that performance and reliability issues be addressed in the underlying technology, while availability can be achieved through inexpensive redundancy of disk and/or array components. An open source network operating system (OS) based on Linux is a core component of the system architecture. Linux provides a highly reliable, configurable, cost effective OS that can be leveraged in the implementation and provide a customized installation that is both cost effective and easy to implement and manage. Computational middleware is required to effectively manage distributed processing tasks among portal compute nodes [2].

D. User Management Architecture

User management is a core component of the computational portal architecture. Users are provided with ability to customize the portal interface as well as the level of detail provided for defining options in operational tasks or procedures. For most if not all operations, default settings can be displayed along with configuration options that can be used to modify the operation. General interfaces to geospatial data could be crafted for user communities ranging from science researchers to the general public.

E. Data Product Processing Architecture

As a true scientific portal, the CME Portal should allow users to conduct tasks that range from simple retrieval of selected data products in desired formats to chaining tasks to conduct complex processing on a selected image or a group of images. Users may elect to share processes they have developed so that other users may test the process on their desired data. The decision to share processes can be configured to limit the use of the process by person as well as to lock the object to prevent modification of the process by others. Additionally, process owners may elect to hide and lock the configuration settings for a process to protect any intellectual property rights that the developer may wish to explore for the designed process.

F. Product Delivery Architecture

The fulfillment of data product requests should consider the bandwidth available to the end user, the size of the data product, reasonable format options to reduce the file size of the data product, and the overall resources required to deliver the data. Data compression of data products should be avoided when compression would cause data losses to data that are intended for research purposes and should be maintained in research quality. Upon generating a data product request, users are provided with an estimated time to complete the processing of the data product. Upon completion of the product, an email is sent to the user informing that the product was successfully completed and providing feedback about the data product delivery mechanism.

III. CONCLUSIONS

The demand for remote sensing image data is massive and growing at a pace that exceeds storage technology and Internet bandwidth development. Computational portals provide technology resources for reducing raw data content and creating and delivering data products to users in a timely and cost effective fashion that can be configured to meet the users needs while hiding processing complexity and minimizing stress on the users computational and networking resources. The computational portal envisioned in the CME Portal can provide a set of scalable methods and configurable processes for providing cost-effective customized data products from remote sensed data streams. The architectural components of the CME Portal build upon standard architectural components typically associated with computational clusters, multi-tier client-server applications, and scalable data storage archives. The CME Portal project at MSU is an ongoing effort to advance the science and technology of developing and implementing computational resources for managing, archiving, and processing remote sensing image data and distributing the finished data products to users and user communities for use in spatial decision support systems and activities.

IV. REFERENCES

1. Datcu, M., Seidel, K., and Schwarz, G. (1999). Information Mining in Remote Sensing Image Archives. In Kanellopoulos, I., Wilkinson, G., and Moons, T., editors, Machine Vision and Advanced Image Processing in Remote Sensing (MAVIRIC), pages 199-212.
2. Nolan, J. J., Sood, A. K., Simon, R. An Agent-based Architecture for Distributed Imagery & Geospatial Computing, in Proceedings of the 29th Applied Imagery Pattern Recognition Workshop. Washington, DC, October 16-18, 2000, pp.252-257.
3. Cai, Guoray, 2001. GeoVIBE: A Visual Interface to Geographic Digital Library, published in The 2001 ACM-IEEE JCDL Workshop on Visual Interfaces
4. Lim, E.P., Goh, D., Liu, Z., Ng, W.K., Khoo, C., Higgins, S.E. G-Portal: A Map-based Digital Library for Distributed Geospatial and Georeferenced Resources, in Proceedings Second ACM+IEEE Joint Conference on Digital Libraries (JCDL 2002), Portland, Oregon, USA, July 14-18, 2002.