ECHO – An Extensible Framework Supporting the Earth Science Enterprise

Keith Wichmann
Global Science and Technology, Inc.
7855 Walker Drive, 2nd Floor
Greenbelt, MD 20770
wichmann@gst.com

Robin Pfister
Code 423
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
robin.pfister@gsfc.nasa.gov

Abstract- **The ECHO (Earth observing system Clearing-HOuse) system has been built and operated to serve as a catalog of Earth Science data and services as well as a broker. Of particular interest to this community is the extensible nature of ECHO and how it may be leveraged to provide an infusion path for information technology prototypes. One important path of extensibility is the ability to register services into the system. It is through this mechanism that other IT (Information Technology) investments can be made available to a larger audience. This paper will focus on the extension points of ECHO with some examples explaining how the overall Enterprise will benefit, with a focus on representing services through ECHO.**

I. INTRODUCTION

As NASA's Earth Science data and product holdings continue to grow and to exist in more distributed locations, ECHO is being recognized as a way to facilitate this growth while still giving the average user excellent access to all of the data they are allowed to have. In order to achieve this ambitious goal, ECHO has designed several extension points into itself. These extension points allow new capabilities to be added to the Earth Science Enterprise without requiring new releases of the clearinghouse and order broker. One such extension point is the ability to add third party provided services into the system. These services perform functions related to the Earth Science community and can exist anywhere in the world. The clearinghouse allows a service to register itself so that it can be found, and to associate itself with the type of data on which it functions. Clients of the clearinghouse can then find both data that is useful to them and services that can act on that data in one location on the internet. Furthermore, ECHO allows a client to represent the entire metadata holdings of the clearinghouse as if they were managed directly by the client without the added burden of actually doing so. A special type of service which is referred to as a metadata utility service is explored as a way of augmenting the capabilities of clients in how they access the clearinghouse with functions that were not designed into the system to begin with. It is expected that the aggregation of information about how to access Earth Science data and ser-

vices into a single location on the internet with an open API (Application Programmer's Interface) will not only allow existing applications that serve the enterprise to improve, but will also encourage the development of new ways to interact that are focused on the wide variety of users who need to use this information.

The system is envisioned to be an enabling system. Other groups can have full and open access to its holdings (with allowances made for protected metadata) allowing them to build new human interfaces or even develop automated machine-based interfaces to the system. The system also holds as a guiding principle that the information technology industry should lead, and that the system shall adapt that technology to the purposes at hand. It is also anticipated that the ultimate users of such a system would have new features of their own to contribute. So designing for change is identified as a worthwhile goal. Global Science & Technology, Inc. has built ECHO as an answer to these problems in the Earth Science domain.

While ECHO is seen as an enabler for clients that search metadata and create orders for the represented data, it is acknowledged that not every function should be provided by ECHO. For this reason, ECHO includes the concept of a search service (or metadata utility). These services extend ECHO's capabilities in some way that assists clients in their interactions with ECHO. These services allow ECHO to leverage capabilities provided by other systems without trying to replicate them. Additionally, ECHO is being augmented to allow providers of data services (those services that produced Earth Science data) to enter information into ECHO such that those services can be found and then applied to data that is (or might be) represented in ECHO.

A major feature of ECHO is that all interactions with it occur using XML (eXtensible Markup Language) as the base message format. This gives the system an extensible base upon which to build. The system uses a layered architecture combined with some code generation techniques to provide a structure that allows for rapid introduction and updating of business logic.

II. ECHO'S EXTENSION POINTS

At its simplest, ECHO is a system that understands how to receive a message, perform some action as a result, and respond to the message with a result message. However, the result message may be the result of a very simple or a very

---

complex action. As one drills deeper into ECHO's layered architecture, the core is found to be a database that stores the clearinghouse of information. That clearinghouse can be searched using traditional text based methods that web search engines use, but also using temporal and spatial types of searches that are appropriate for the geospatial data referenced here. The data model for ECHO is currently based on the EOS (Earth Observing System) data model. While ECHO is really a clearinghouse framework that has been customized to the Earth Science community, the focus of this paper is to address those extension points that are available without code changes to the system. The layered architecture allows for the framework's application in other areas as well.

The extension points of ECHO include the client interface, the provider interface, the data model and services as well as some others that are less important. Each of these extension points allows the introduction to the enterprise of new functions without a new release of ECHO.

A. Client Interface

The architects of ECHO recognized early on that a core problem of the enterprise was that a single user interface could not fulfill the needs of all the different communities in the enterprise. What ECHO could do is offer to manage the science metadata of the enterprise and broker the orders, making this available to client developers as if their client managed all of the metadata. ECHO offers APIs which are in essence XML messages sent to and received from the clearinghouse server.

Since ECHO offers all of its capabilities through this XML message passing scheme, new clients can be introduced into the enterprise to address new or more specific communities than are currently addressed. While the new client represents the potential for additional load on the clearinghouse, the functions of the system are unchanged. So, new GUIs (Graphic User Interfaces) can be developed which provide a different paradigm of accessing the metadata, or provide community specific views of the metadata, without modification to ECHO itself. ECHO supports tracking these clients so problem interactions can be traced to their origin. In the future, ECHO will provide an endorsement mechanism so that data and service providers within ECHO will be able to identify their clients of choice.

B. Data Provider Interface

As additional Earth Science data providers decide to participate in ECHO, the clearinghouse can accommodate them with no need to release a new version. This is accomplished primarily by being data driven, but also by providing proxy mechanisms that can transform ECHO's message format into other system's formats. In this way, ECHO supports the addition of new data providers, and also supports existing data providers adding new collections of data into the system.

C. Data Model

The ECHO data model leverages the data model work done for the EOSDIS (EOS Data and Information System) Core System (ECS). ECHO adopted and extended this model, but kept a key point of extensibility in place. This extension point is called Product Specific Attributes and allows a collection to define a template for new metadata to be associated with the individual granules within that collection. ECHO then allows users to query based on those items. This allows a science team to augment the metadata with a new attribute and use it as a search discriminator.

D. Services

ECHO is currently adding the capability to register and broker third party services in the system. These services can be added and removed from the system without requiring a new release of ECHO. ECHO acts as a registry or clearinghouse of the available services, allowing those services to express their relationship to the data stored in ECHO. It can also broker an order for those services in certain circumstances where it makes sense (e.g. those services that will be applied to data that has been ordered through ECHO).

E. Order Options, User Preferences

Two other extension points are mentioned here for completeness. These are order options and user preferences. Each provider that requires an order (as opposed to just providing a URL link to the data) has the option of specifying what information needs to be captured in order to ship the data (typically referred to as packaging and shipping options) which can also be used to specify additional services that should be applied by the data provider before it is shipped. User preferences allow useful state information to be added to a user's profile. One early use of this capability is to identify which un-submitted order is the current shopping cart for a user.

III. SERVICES

There are various service concepts used in ECHO. ECHO provides Clearinghouse Services. These are internal services that support the clearinghouse functions. Then, through APIs, ECHO supports various external services that are made available by service providers and shared with users of ECHO. These include metadata utility services, data and science services, and in the future, administrative services. These are described in the following paragraphs.

A. Clearinghouse Services

The Catalog Service is the mechanism used to query and retrieve results from the system. It supports multiple query languages and result formats. It supports the chunking of result sets so that large result sets can be effectively handled by clients. It also supports saving and restoring both queries and result sets by registered users of the system for later use.

- The Order Entry Service allows both registered users and guests to build, validate, quote and place orders, as well as to examine the orders that they have created.

- The Registration Service allows a guest to create a registered user account, as well as allows prospective providers to alert the operational staff of their interest.

- The User Account Service allows a registered user to maintain address, email and phone information so that a client may use that to fill in information for orders. It also supports checking a user's order history.

- The Provider Account Service provides a similar interface for ECHO providers to describe themselves and check their order history.

- The Provider Profile Service gives users information on Providers.

- The Group Management Service allows named groups of users that can then be used to grant special metadata or order access privileges to.

- The Data Management Service is used to create rules that describe what metadata in the system is protected, and who is given access to it. It supports restricting and granting access based on individual granules, collections, or sets of granules within a collection based on temporal conditions or a quality flag.

- The Provider Order Management Service allows a provider to give ECHO information about the status of an order, including the ability to cancel an order.

- The Ingest Service provides the ability to ingest/register new data, update existing data and delete.

- The Metadata Update Service supports updating certain fields of the model (and eventually, all of them).

B.   Metadata Utility Services

These services are focused on augmenting the capabilities of the clearinghouse. They can be used to add functions to the search API, or be offered as a way to interpret results that are retrieved from ECHO. Their categorization is based entirely on the fact that they work with the messages that are sent to and received from the clearinghouse.

Several existing and past ESTO prototypes fit the spirit of this description. Perhaps the best example is that of the Gazetteer. This system maintains a mapping between place names and their locations on the planet. It is reasonable to think that a user (or client developer) would want to search for geospatial data by using the name of the place they are interested in. For instance, a user may want to search for data that matches Maryland. While some states lend themselves to bounding box representations very well, Maryland is not one of them. The polygon that represents Maryland could be maintained at a Gazetteer, and the Gazetteer could offer a metadata utility service through ECHO that converts the place names which it understands into polygons which ECHO understands. In this way, ECHO's capabilities are augmented in a way that benefits clients, but ECHO is not burdened with the added responsibility of maintaining political boundaries.

Similarly, other metadata utility services can be envisioned that would assist the client's job of providing an effective interface to the user. A Thesaurus could be used to map from one community's set of keywords to the common set of keywords that ECHO provides. Similar to the Gazetteer, the mapping would be applied to a search query before it was executed by ECHO, transforming keywords that would have produced undesirable results into ones that produce the desired results. Also, a mapping could be performed on the result of the search such that the common keywords of ECHO are converted into the community specific keywords that best serve those users. There may be many instances of Thesauri, each with its own target community, but all using the same search service mechanism.

Another example would be a service that given a spatial polygon calculates the center point. This would be useful for GUIs and is illustrative of a metadata utility service that functions on the output of ECHO, rather than the input.

One final example would be Coincidence Search. In this case, the client's aim is to find a set of data in which multiple observations from different instruments of a single area within a tolerance of a given window of time. In other words, a user may want to find all the places in North America where a MISR and a TRMM observation were taken within 5 minutes of each other. The coincidence can be calculated and translated into a query to ECHO which produces results that are physically, rather than theoretically, available.

C.   Data and Science Services

Another extension point under development for ECHO is the ability to allow ECHO clients to find out about and use third-party provided services. Obviously, these services are in the Earth Science domain. ECHO refers to them as data and science services because the end result of their execution is that some piece of Earth Science data is generated. Service outputs can mimic types of data that already exist in the archive, or they can generate new information that is not archived. ECHO allows the association of these services with the granules represented in the clearinghouse. In this way, ECHO users will be able to derive what services apply to the particular data that they are working with. ECHO divides this type of service into three subtypes based on how the client and service interact with ECHO.

Advertised Data Services

Some services require no interaction with ECHO and simply present a web page that lets a user know everything they need. In this case, ECHO facilitates finding out about the service, and takes no further action. These are referred to as Advertised Data Services.

Context-Based Data Services

Other services add the need for ECHO to associate which data represented in ECHO will be used for the invocation (binding) of the service. The client continues to invoke the service directly to the service provider, but ECHO provides an additional service of pre-populating a subset of the service parameters. In this way, various types of services that are invoked via a Universal Resource Identifier (e.g. Web Mapping Service, DODS, etc.) can be invoked by a client in the same fashion without necessarily having to understand all aspects of the standard for putting together the URI. ECHO abstracts that problem, and refers to these as Context-Based Data Services.

Brokered Data Services

Services that require being ordered because of payment involved or some time delay that will be involved in invoking the service can be brokered through ECHO. ECHO will bind to the service upon request from the user for that service. Future implementations of ECHO might support the chaining of brokered services to produce new aggregate services. These are referred to as Brokered Data Services.

D. *Administrative Services*

In the future, other externally provided services will also be supported through ECHO. An example is a billing and accounting service that might be provided by a single service provider but might serve several data and service providers to allow them to charge for their data and/or service.

Another example would be a user profile service. The enterprise would benefit from a common pool of users with its management implemented in a single place. ECHO could leverage this pool of users rather than maintaining its own.

IV. CONCLUSION

ECHO is designed as a framework for building clearinghouses and order brokers. It has several extension points that allow the addition of new data providers, new data sets, new service providers, new services, new attributes of the data and new clients that represent the metadata and data of the enterprise. ECHO helps enable the enterprise to be adaptable over the long term while still providing the focused search capabilities that the client community needs.