EXPERIENCES WITH NASA EARTH SCIENCE DATA INFORMATION SYSTEMS AND SUGGESTIONS FOR IMPROVEMENTS FROM A SCIENTIST USER PERSPECTIVE

K.-S. Kuo

Goddard Earth Science Technology Center/Caelum, Greenbelt, MD 20771
Email: Kwo-Sen.Kuo@nasa.gov

1. INTRODUCTION

For planning the continuous evolution of the NASA Earth Observing System Data Information System (EOSDIS) [1] the Earth Science Data Information System (ESDIS) project requested the author, from an Earth science data user point of view, to explore potential improvements for the future system. Observations of the weaknesses and inefficiencies of the current modes of operation are made based on (1) user scenarios produced by users of the EOSDIS data centers, (2) a survey conducted by the author on search-and-order clients operated by ESDIS affiliates, and (3) the author's personal experience with helper applications provided by NASA Earth Science data systems.

In this abstract we first introduce briefly the data considered in the study. Observations obtained from the data are summarized next. Insights drawn from an in-depth analysis of these observations, taking also into account other relevant socioeconomic factors, lead to the strategic recommendation for the creation of an *integrated data environment*, in which all data related tasks can potentially be carried out just as an Integrated Development Environment is able to handle all code-development related tasks.

2. DATA

2.1 User Scenarios

To gauge the performance of and to find areas of improvements for its data systems, the data centers were asked by the ESDIS project to produce user scenario reports prepared either by the users themselves or with the assistance of data center user services staff. Table 1 lists the User Scenarios that were received as of January 2009. These scenarios cover a range of topics, including studies on wild fires, sea ice, land surface, precipitation, and carbon dioxide emission. The format used for reporting these scenarios is designed to capture users' experiences on three aspects of data access[1] through the data centers coordinated by ESDIS: *finding data*, *gathering data*, and *extracting data*.

Table 1 User scenarios.

User(s)	Scenario Title	Sponsor
Chuvieco *et al.*	Global characterization of fire activity	SEDAC
Fetterer and Meier	Investigation of the Beaufort Sea polynya in August and September 2006	NSIDC
Hulley *et al.*	ASTER land surface emissivity data base of North America	LP DAAC
Kuo *et al.*	Coincident data for precipitation retrieval	ESDIS
Luyssarert *et al.*	CO_2 balance of boreal, temperate, and tropical forests derived from a global database	ORNL

2.2 Search-and-Order Capability Survey

ESDIS also supported a survey of the capabilities of currently available search-and-order clients (or, some prefer to call them web services). The primary purpose of the survey is to discover strengths of various clients, which may then be used as a basis for improving other clients or developing future clients. Sixteen (16) clients from fourteen (14) providers are surveyed. These providers include data centers, metadata repository, middleware provider, and data-and-service directory service, all of which are now ESDIS-funded activities. Table 2 lists the acronyms of these providers and the clients surveyed with corresponding providers.

Table 2 Search-and-order providers and their corresponding clients.

Provider	Description	Client Acronym	Client Long Name
ECHO	Earth observing system ClearingHOuse	WIST	Warehouse Inventory Search Tool
ASF DAAC	Alaska Satellite Facility DAAC[2]	URSA	User Remote Sensing Access
CDDIS	Crustal Dynamics Data Information System	(None)	
GCMD	Global Change Master Directory	GCMD	Global Change Master Directory
GES DISC	GSFC Earth Sciences Data and Information Services Center	Mirador	(None)
		Giovanni	GES-DISC Interactive Online Visulization ANd aNalysis Infrastructure
GHRC	Global Hydrology Resource Center	HyDRO	Hydrologic Data Search
LaRC ASDC	Langley Research Center Atmosphereic Science Data Center	MOCT	MISR Order and Customization Tool
		AOT	ASDC Order Tool
LP DAAC	Land Processes DAAC	GloVis	USGS Global Visualization Viewer
MODAPS	Level 1 Atmosphere Archive and Distribution System (LAADS)	LAADS	Level 1 and Atmosphere Archive and Distribution System
NSIDC DAAC	National Snow and Ice Data Center DAAC	ADS	Advanced Data Search
NSIDC	National Snow and Ice Data Center	GISMO	Graphical Interface for Subsetting, Mapping, and Ordering
ORNL DAAC	Oak Ridge National Laboratory DAAC	MSS	Metadata Search System (aka Mercury)
OBPG	Ocean Biology Processing Group	L12B	Level 1 and 2 Browser

[1] Here "data access" is defined with a scientist-centric view, in which it includes accessing the data values after data granules have been downloaded from the data center archives.
[2] DAAC: Distributed Active Archive Center.

PO.DAAC	Physical Oceanography (PO) DAAC	POET	PO.DAAC Ocean ESIP[3] Tool
SEDAC	Socioeconomic Data and Applications Data Center	SG	SEDAC Gateway

As indicated by the survey's title, it consists of two parts: search capabilities and order capabilities. The former have to do with finding the data granules of interest. In addition to order processing, the latter also include related capabilities such as on-demand subsetting, user account management, and order delivery management, which are considered integral parts of the ordering process. Search capabilities are further separated into two categories: *Conventional Capabilities* and *Advanced Capabilities*.

2.3 Other Tools and Services

Due to the proliferation of open source software, when there is a need many people are now in the habit of searching the Open Forum first before purchasing commercial ones or writing their own codes. This essentially constitutes software reuse, which usually results in increased productivity. GCMD, as well as the open web, has thus been explored by the author to find services and/or helper applications for data exploration, visualization, and extraction (specifically, subsetting).

3. OBSERVATIONS AND ANALYSES

3.1 Observations

Based on the User Scenarios, the Capability Survey, and the experience on GCMD Data Services, the following observations are made:

- Although the search-and-order clients are developed by different organizations, those surveyed here are rather similar in features and functionalities.
- There is considerable room for improvements in services after an order has been placed. When a user orders a large number of data granules from disparate data products it becomes difficult and tedious to manage what follows after the order.
- Almost all of the search-and-order clients and data service tools are stand-alone (i.e., not interoperable). The users (*i.e.* researchers) cannot count on consistency of design among these similar or related tools.
- Those tools or services developed with different but complementary purposes do not interface with each other.
- Despite the absence of some useful capabilities mentioned above, there exist currently many other help applications or services that could make using NASA Earth Science data easier and more productive. However, they suffer similar deficiencies as the search-and-order clients, *i.e.*

[3] ESIP: Earth Science Information Partners.

duplication in basic functionalities but lack of sophistication and advanced capabilities. In addition, their discovery could be made easier.

- It is usually not obvious as to how up-to-date the helper applications or software packages are, when one encounters them for the first time. Then, there is the issue of keeping up-to-date once a package has been installed.

3.2 Analysis

The results from an analysis of the observations are summarized below:

- From an external user's point of view, the multiplicity of basic clients and tools with much the same capabilities implies that a good portion of the collective effort has been spent on duplication while it could have been better utilized to create advanced features.
- Without good order management services, sorting out a large order becomes labor-intensive and time-consuming and graceful fault-recovery is impossible.
- Adopting new technology and learning new services and tools are made easier if they have consistent design in operational procedures and look-and-feel.
- Interoperability helps realize the full potential of tools and services but it can only be achieved through standardized interface protocols.
- Software reuse depends on a central repository that encourages and implements sound software engineering principles and practices.

4. RECOMMENDATION FOR STRATEGY

In addition to the observations and analysis discussed above, the socioeconomic environment of NASA funding practices and project management need to be considered as well. As a result the author believes the best strategy is probably the creation of a common integrated data environment for all data related processes, similar to the Integrated Development Environments (such as Eclipse and NetBeans) for computer code developments. The rationale for such a recommendation will be provided at the conference presentation; the requirements for the common integrated data environment will be detailed there also.

5. REFERENCES

[1] M. Esfandiari, H. Ramapriyan, J. Behnke, E. Sofinowski, "Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS)", Proceedings of the IGARSS 2006, Denver, CO, July 31-August 4, 2006.