University of Maryland Land Cover ESIP

Users Need Assessment

Interim report for June 1st, 1998 Milestone Report

Catherine Plaisant, HCIL
Francis Lindsay, Geography
University of Maryland - College Park
College Park, MD 20742

I. Introduction

This report reflects our initial Users Needs Assessment regarding the University of Maryland Land Cover ESIP. This document will be periodically updated to reflect our refined understanding of the users and their tasks and as the data and services offered in the ESIP evolve.

The purpose of this assessment is to provide an informed basis for the interface design and evaluation to be done in the months and years to come. To this end, we reviewed past reports discussing EOSDIS users and needs, interviewed staff working on current Land Cover ESIP datasets and studied current ESIP website usage logs. In addition, we prepared a "suggestion box" form for inclusion in several places in the early version of the web site, that will faciliate online collection of users' characteristics and needs. Finally we started a collection of representative scenarios of potential use of the ESIP. Those scenarios will be an important part of the design and evaluation process of the ESIP.

This initial phase of the users needs assessment started in March 1998. The report will be updated at the occasion of each milestone report of the 1st year.

II. Understanding Users Needs

II.1 Summary of Past Studies

We reviewed two sources of data on EOSDIS' users and their needs.

1. ECS User Characterization Methodology and Results (Sept. 1993)

This report divides the ECS user community into three main categories: EOS Science users, General Science users and Non-Science users. EOS Science Users are the investigators funded by NASA under the EOS program. The category of General Science users includes all science users. This category includes, but is not limited to, EOS-funded researchers, university researchers, and federal employees who conduct basic Earth System research. The category of Non-Science users includes the widest variety of users, ranging from commercial users to K-12 users.

An extensive literature and society membership study estimated the number of potential EOSDIS users to be as shown in Table 1. The study then continues by estimating the number of science users interested in land studies to 27% (opposed to 50% interested in ocean, athmosphere 18%, or cryosphere or

interdisciplinary 1%). This shows the relatively small number of heavy users and large number of occasional non-science users that the Land Cover ESIP can potentially attract.

The study estimates that for every 500 products accessed by Land users, approximately 300 will be Land products, 150 will be Atmospheric, 60 will be Cryospheric, 20 will be Ocean, 5 will be General, and 1 will be Miscellaneous. This shows that Land Cover ESIP users are likely to need to access datasets in other ESIPS or data centers.

	User Community	Minimum Number of Direct Users	Maximum Number of Direct Users
	EOS Science Users (access rate varies from daily, monthly or yearly)	1,900	3,200
	General Science	4,200	8,200
	Non-Science (1-2 access per year)	69,430	268,320
	TOTAL user community	**75,530**	**279,720**

Table 1: Summary of Demographics of EOSDIS User Community
(from ECS User Characterization Methodology and Results, Sept 1994)

But further analysis suggests that the Science community will spend more time with the actual data than will the non-science community. The non-science users would make more use of the descriptive information about the data (Level 4 and above data), and of samples (browse data). The study predicts a low demand for the Level 4 products by the non-science community because very few such products exist. As more descriptive information regarding Level 4 products becomes available, it is expected that the demand will increase significantly in particular from the education community.

2. EOSDIS Potential User Group Development Effort Conference (June 1995)

This report identifies twelve potential EOSDIS user groups:

1. Retrospective Research;
2. Field Campaigns and Individual Data Providers;
3. Persistent Information Production for Research;
4. Scientific Environmental Assessment;
5. Commercial Users;
6. Operational Users;
7. Resource Planners and Managers;
8. Policy Formulation and Decision Making;
9. Use of EOSDIS Data and Information Products by the Legal Community;

10. K-14 Education;
11. Collegiate and Professional Education; and
12. Libraries, the Press, and the Public.

A committee was formed for each of the twelve groups. It was asked to discuss five key characteristics of their segment of the potential user community; where possible, responses were to be given as a function of time for the period 1995 to 2001. These five key characteristics are: User population; Intensity of use; User resources; User activities; and Desired EOSDIS functionality and capabilities.

We extracted here a subset of the recommendations we felt were most relevant to the Land Cover ESIP design:

<begin extract>

From Chapter 3: retrospective research

- For an automated system to be successful, it will be important to have a flexible query system, probably using script files consisting of all likely parameters affecting the use of the data (location, time, cloud cover, data quality, existence of other satellite data within a given time span, existence of in situ data etc.).
- For many users it will be important to be able to query by relatively small regions and to acquire subsets of data of these regions rather than very large global data sets.
- Delivery of data both electronically and on various high density media will need to be maintained for the foreseeable future.
- If data are archived at different levels, which considerably affects the speed of access to data sets, then the level chosen for a particular data set should be based on frequency of usage and not on a single parameter such as age. For retrospective users access to relatively "old" data may be as important if not more important, than access to more recent data.
- We recommend that consideration be given to ensuring that initial user contact through the EOSDIS meta-data base allows use of a needs-assessment module to help new or inexperienced users to extract only the data they need and can handle.
- EOSDIS should not attempt to accommodate the limitations of all potential users, but instead should recognize and facilitate such intermediation.
- Efficient, reliable, graphically supported meta-data bases
- We recommend that EOSDIS supply data in more than one format, but the number should be kept to a small number (five or less). To ensure that EOSDIS maximizes its capabilities to receive data from other users, it is important that it is more flexible in terms of formats so far as incoming data sets are concerned.
- Mechanisms will have to be in place to decide which new products can be included in EOSDIS to avoid an explosion in the number of new products to be created and archived.
- Many users will need data sets already integrated in a co-registered form, possibly time and space sampled. Data should be provided in one of a small number of spatial grids to facilitate integrated use of different data sets. Unless EOSDIS fully commits itself to this crucial task of ensuring interuse, the use of data will be greatly restricted because of the enormous overhead placed on the retrospective users in establishing suitable long term data sets.

From Chapter 4 Campaigns and data providers

- A period of 0.5 to 2. 0 years should be allowed for data validation and quality assessment before

release of the final archived data product to the outside community. During this period, access should be restricted to the field campaign approved scientists.
- Users of field campaign data should expect complete documentation, including read routines for common computers; usable formulations; stability in the products with minimum modifications; browse products in reasonable formats such as HDF, GIF, JPEG, and TIFF; up to date data products; both on- and off-line services; and Internet interfaces for background and pointer information.
- EOSDIS should establish data links to other non-EOSDIS entities for real-time acquisition and post experiment archives. EOSDIS shall also address international software, data exchange, data restriction and copyright issues.

From Chapter 5: persistant information production for research

- The most desirable EOSDIS characteristics are simple and regular structures and complete data documentation.
- This user group requires these services: (1) rapid, interactive, network access to complete catalog and inventory information (graphical interface preferred), (2) searches for coincidences and collocations amongst many datasets, (3) data ordering via network, (4) delivery of small (< 100 - 500 Mbytes) datasets via network, (5) media delivery of large (> 1 Gbyte) datasets (compatibility with "standard" media), (6) user-defined (simple) sampling of very large (> 1 Gbyte) datasets, and (7) a human interface with a knowledgeable data scientist who can answer questions.
- Multiple, simple data formats that are properly documented are much preferred over a complex common format. If some popular formats are identified, then either standard conversion software can be supplied to the user or the dataset can be re-formatted during copying at the user's request. Two necessary elements of data documentation are a complete "bitmap" that indicates the precise location of every piece of information and a READ program written in ANSI standard FORTRAN, C or C++.

From Chapter 8: Operational users

- A tiered user interface accommodating varying level of user interest and degree of expertise (i.e., very generalized interface for general public to sophisticated interface for experts or scientists);
- Timely provision of metadata and browse images of all governmental and commercial image data including SPOT, JERS, IRS, ERS and Radarsat along with the ability to order data, determine where it is archived, and what levels or processing are included;
- The system must be understandable -- standard (motif/windows) interface and standard nomenclature -- as well as reliable and dependable;
- Reasonable cost access and use; and,
- Access to ancillary data which is often critical to the exploitation of the imagery data, including description and sourcing of such data.

(to be expanded to suggestions from other chapters)

<end extract>

The study also highlights problems of standards of data quality and recognize that virtually al users have the potential to be information providers.

II.2 Analysis of current Land Cover ESIP usage

(to be added by Frank)

II.3 Interviews

In this first phase of the users need assessment we have interviewed staff responsible for handling users requests for three of the existing ESIP datasets.

Those interviews confirmed the diversity of users and the current difficulty for the staff to assist users who lack domain expertise or computer resources and/or expertise. Many requests for special services (e.g. special output format or projection) remain unfulfilled. Researchers working on the datasets are not mandated or supported to help "ordinary users", therefore only fellow researchers from important research agencies can expect special attention. It seems reasonable to anticipate that "high-level" science users will continue to get expert help when needed, but that the development of simple tools that does not require staff intervention would open the door to the rest of the user community (thousands of users according to the ECS study.)

A common concern of the staff was many users' lack of knowledge of the difficulties of geo-reference. This is seen as a serious (but inevitable) problem when combining datasets and may require special attention in the interface in the form or training or warnings.

On the other hand the combination of the ESIP data with layers such as geopolitical, road or rivers was seen as a very common request by non science-expert users.

Those interviews suggest a simple categorization of users domain knowledge that will be of help when designing the ESIP interface. We propose 4 categories:

- Users with extensive remote sensing knowledge

Likely to have large disk space to download large files

Likely to know everything about the type of data (e.g. AVHRR channels)

Likely to know about specific datasets and come to the ESIP for them.

Those users are the current heavy users of the large datasets that they download and manipulate on their own. They will propably continue to do so until their trust in the ESIP services is sufficient. They may want to also download software.

- Users with limited remote sensing knowledge

  Likely to know of the difficulties associated with acquiring remote sensing data but they need assistance, for example to understand the format used for the data or convert it to their own format.

  Those users will immediately benefit from the processing of the data in the ESIP.

- Users with GIS knowledge (but no remote sensing)

  Likely to have smaller disk space but they expect to easily combine datasets in layers of the GIS environment. They don't necessarily know what data is available and what form it takes.

  Those users will greatly benefit from the simple subsetting of the data that will allow them to download the data on their machines. They will greatly benefit from the related datasets providing commonly used GIS layers for the related area (e.g. political boundaries, roads, rivers etc.)

- Users with no GIS or remote sensing knowledge

  Those more naïve users may have any kind of computer environment and are expecting ready-to-use imagery. They might be satisfied with sample data, or with simple geographical subsetting combined automatically with additional layers and a simple bitmap output (e.g. GIS). An resource library linking to other sites more appropriate for them will also very useful.

II.4 Online Users Needs Collection

Interviewing user support personnel only provides a partial view of users' need. Not all users call for help and only a few potential users even reach the site today. Another technique to collect the feedback, wishes or complaints of users reaching the site is to instrument the site to automatically collect comments and suggestions on line.

Appendix 1 shows the proposed suggestion box page to be used in the early stages of the ESIP. It will be available from the home page of the ESIP and each dataset, as well as shown automatically before users download data. It is not mandatory to fill the form and users can skip the page. When usage grows and we know more about users the abigatoy passage by this page will be remove or replaced by a registration page if necessary (e.g. for data processing services requiring email dialogs or storage space reservation).

(UMd seal) **Land Cover ESIP Suggestion Box**

[SUBMIT] [CONTINUE / SKIP] [EDIT OLD REGISTRATION INFO]

HELP define your needs for

- **future PRODUCTS**
- **future SERVICES,** by submitting this form, even partially filled.

How will you use this data*?*

Suggestions - Comments

Name [e.g. John Doe]	
Institution [Univ. of Maryland]	
Email [doe@cs.umd.edu]	

[X] Check if you want to receive EMAIL ANNOUNCEMENTS of major ESIP updates to that address.

[ ] Check if you would allow us to CONTACT you by email during our users need assessment.

Your name or email address will not be used for any other use without your consent.

Areas of interest (*check all that apply)*

[ ] Global change research [ ] Nat. resources [ ] Socio-economic [ ] Habitat

[ ] Hazard management [ ] Human Impact [ ] Education

[ ] Other [_____]

How did you find out about us? (*check all that apply)*

[ ] Personal contact [ ] Web Search [ ] GCMD [ ] Publication reference

[ ] Link from other site[_____]*(name or URL will help us)*

Type of community you represent or serve (*check all that apply*)

[ ] Higher education (research work) [ ] Higher education (coursework)

[ ] Government [ ] Commercial [ ] Non-profit [ ] K-12 education [ ] Other

[SUBMIT] [CONTINUE / SKIP]

Last updated 4/5/1998

For questions about the data collected in this form contact UMd-ESIP@umiacs.umd,edu

II.5 Workshops

We are planning to hold a local workshop during the summer with DC area potential users. We decided that it was preferable to implement the initial prototype of the ESIP before convening a panel of users. This will allow us to collect early feedback and involve users more actively in the design process. We plan to invite users in June for a one day workshop early September. The DC area offers the diversity of academic, government and businesses that is adequate for the initial workshop and we will avoid traveling expenses.

III. Discussion and recommendations

Users will come to the Land Cover ESIP with problems that vary according to several factors, including: how well-defined the solution to the problem is (e.g., ranging from simple facts to interpretations for complex phenomena) and how well-defined the problem is in the information seekers mind. Such information systems must accommodate different levels of experience with the content area, with the information system itself, and with information seeking in general.

III-1 Initial task taxonomy

- Finding the Land Cover ESIP
- Gaining an overview of the Land Cover ESIP (i.e. finding the gist of the site, what is there and is not there, allowing users to decide to stay or leave the site)
- Locating data of interest (or recognizing lack of) via browsing or searching
- Ordering or acquiring data of interest

- Processing data of interest with ESIP tools
- Geographical or temporal subsetting
- Alternative projection
- Alternative corrections (e.g. athmospheric corrections)
- Alternative data format
- Filtering according to certain parameters
- Understanding format of data acquired
- Locating data related to the data of primary interest (e.g. political boundaries, photo, etc.) in the ESIP, in a related ESIP, or outside the ESIPs
- Understanding if related data can be adequately processed with initial data of interest:
- Identify potential problems of geo reference, projection, corrections, temporal cooincidence, etc.
- Submitting new data to the ESIP
- Submitting data product related to existing ESIP data (either generated from or complementary to ESIP data)
- Submitting report, scientific paper, success story to ESIP
- Contacting staff or reviewing FAQs (Frequently asked questions)

III.2 Land Cover ESIP web design guideline

To provide a homogeneous environment to users of the Land Cover ESIP is will be important to define guidelines for a common organization and layout of the web pages and interfaces of the ESIP. Those guidelines will cover terminology, use of logos and graphic, consistent use of fonts and color coding, mandatory interface components and their placement. A consistent set of guidelines will describe standard look and feel for datasets and service pages.

II.3 Planning for Evaluation - Development of Metrics

How will we evaluate the success of the Land Cover Esip? Making large amount of data and tools available to the public is an important step but a better measure of the success of the ESIP should lie in the measure of the activity of users, and their ability to produce valuable scientific insight or fulfill the need of their business or occupation. Those things are difficult to measure but we will attempt to collect data and feedback from users that reflect this higher level of accomplishment in an approach combining quantitative and qualitative evaluation.

Availability of data

- Number of datasets available
- Volume of data available
- Number of algorithms or services available for download
- Number of algorithms or services provided from the ESIP site

Access to data:

- Number of users visiting the site and each dataset
- Number of data files downloaded

- Number of data products processed automatically by ESIP
- Number of software files downloaded
- Data processing speed performance

Usability of the ESIP

- Time to perform representative tasks
- User satisfaction (user the QUIS : Questionnaire for User Interface Satisfaction)
- Number of messages requesting assistance related to the user interface

Development of users community

- Number of users registering or providing personal contact information
- Number of users actively posting messages in online bulletin board
- Number of messages exchanged in online bulletin board
- Proportion of email questions/problems/positive and negative comments about:
- data itself (availability, future plans, quality etc.)
- access to data (data format, data processing, user interface)
- reference to related products
- reference to other ESIPs
- reference to related people

- Possibly: number of participants in the annual Land Cover user's group meeting

Dissemination of results by ESIP team

- Number of papers published
- Number of presentation made to EOSDIS community, conferences
- Number of mentions in newspaper articles

Outcomes

- Success stories (Scientific discovery, impact on businesses or local government, etc.)
- Data products generated by users (kept by users or re-ingested in ESIP)
- Papers written by users (published elsewhere, or available online at ESIP site)
- Adoption of similar data processing techniques by other data centers
- Adoption of similar user interfaces by other data center

II.4. Scenarios of use (to be expanded significantly)

The following descriptions represent some potential queries that the UMCP land cover ESIP activity may encounter. These are intended to foster discussion regarding the needed functionality of our database system and the underlying data products. Each situation presented begins with some task the user has formulated. We provide a brief chain of events from the original query to the resulting product. Those `stories' of use will also serve as the scenarios for upcoming demonstrations of the UMCP efforts.

1:

Employee of a state-level natural resources department needs data regarding local marsh conditions. The user:

1. downloads a map of local marsh conditions for his region;
2. wants to compare local conditions to those of other states/regions;
3. wishes to add other information layers to this map including political/jurisdictional boundaries, roads, soils, population, etc.;
4. wants to perform a time series analysis using marsh conditions taken annually for last ten years;
5. selects appropriate data layers and produce a final, high quality, map output (TIFF, PICT. EPS, etc.);
6. wishes to download raster and vector files for use in their own computer system.

2:

University Ph.D. student wishes to explore what differing types of land cover data is available for the United States. The user wishes to acquire land cover data products that will allow her to correlate changes in coastal marshes to changes in adjacent vegetation cover over some time unspecified period.

1. The user examines the browse map and notices that there is data for the whole Eastern Coast of the US. User decides to focus on the Chesapeake Bay area.
2. User wants to see detailed browse map of the Chesapeake Bay showing only areas that are completed deteriorated. User decides to examine two areas, one in Maryland and another at the border of Maryland and Delaware. User selects areas using bounding box tool. User is provided with a preview image - user decides this is satisfactory.
3. User gets full image on screen (full resolution GIF - may require combination of differing data sets - also, user may specify differing, non-web, image formats).
4. User requests an Arc/Info coverage (FTP or tape).

The student saves the query location area in virtual workspace.

1. User moves to another data set.
2. User now wishes to include AVHRR data for the areas specified.
3. User views marsh data with AVHRR data (channels 1&2 - NDVI).
4. User views the temporal coverage for Land Cover Product data for the areas selected in workspace.
5. User selects 10 years of Land Cover Product data (for every year on March 15)
6. Browser returns 10 tiled browse preview images.
7. User selects 4 images for download (FTP or tape).

3:

Member of a non-profit environmental organization wants to bolster their in-house research on tropical deforestation by including the analysis performed by the Landsat/Pathfinder project.