161-TP-001-001

EOSDIS Product Use Survey

Technical paper - Not intended for formal review or Government approval.

September 1995

Prepared Under Contract NAS 5-60000

RESPONSIBLE SCIENTIST

Giulietta S. Fargion /s/ 9/27/95

Dr. G. S. Fargion, Senior Scientist ESSi Date
EOSDIS Core System Project

SUBMITTED BY

Mark Settle /s/ 9/29/95

Dr. J. A. Colucci, Manger Science Office Date
EOSDIS Core System Project

Hughes Information Technology Corporation
Upper Marlboro, Maryland

This page intentionally left blank.

Abstract

The EOSDIS Product Use Survey is designed to understand the need for, and use of, data products that will become available through EOSDIS in the years 1998-2000. This information will be used to understand on-line access to products (i.e., ordering and browsing). This knowledge permits system developers to determine the required size of individual data servers and communication links.

The EOSDIS user community includes scientists, students, state and federal agency personnel, policy-makers, and commercial users. The current ECS design focus is on the science users; for this reason, the survey was developed with science users as the main audience. A message inviting potential EOSDIS users to complete the Survey was e-mailed to about 4700 earth scientists. In addition, the survey was announced on electronic bulletin boards and in two scientific journals. The survey was administered electronically via the World Wide Web (WWW) and responses were received from 595 users. This paper describes the history of the EOSDIS Product Use Survey, the response data, quantitative data analyses and potential applications. The results indicate that the pull of a product is related to the dynamic nature of the phenomena studied, which, in turn, is closely related to the temporal resolution of the product.

Keywords **:** survey, products, user, characterization, EOSDIS, product, use, earth, scientists, science.

This page intentionally left blank.

Contents

Abstract

1. Introduction

2. Survey Development and Analysis Methods

3. Results and Discussion

4. Conclusion and Applications

5. References

Appendix A

Appendix B

Abbreviations and Acronyms

Figures

Tables

1. Introduction

1.1 Purpose

In order to meet the needs of the potential EOSDIS user community, ECS developers need to understand the users' relative demand for the various data products. This information allows the system developers to determine the required size of individual data servers, size of communication links, and other related parameters. To collect this data, a survey was developed that queried scientists about their future needs for browsing and ordering data products that will become available through EOSDIS during the period 1998-2000. The intended audience of the survey was the general earth science community including, but not limited to, the EOSDIS NASA-funded scientists.

The present paper is intended to inform ECS developers, NASA, and the user community about the development and results of the EOSDIS Product Use Survey. The information provided from this survey was used for the purpose of statistical analyses only in order to assist system developers in gauging interest in each data product.

1.2 Organization

This paper is organized as follows:

Section 1 provides the purpose and general organization of this paper and delineates the procedures for its review and approval.

Section 2 describes the development of the survey and methods used on the survey data.

Section 3 presents results and discussion.

Section 4 presents conclusions and applications.

1.3 Review and Approval

This Technical Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

The ideas expressed in this Technical Paper are valid for six months from the approval date. Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Information Technology Corporation
1616 McCormick Drive
Upper Marlboro, Maryland 20774-5372

1.4 Acknowledgments

The entire staff of the User Characterization Group worked as a team to make the EOSDIS Product Use Survey possible and the analysis of this data a success; their contributions are gratefully acknowledged. Special thanks are due to:

Tess Wingo for the endless work on the survey outline and content;

Tim Pruss for the WWW implementation;

Vijay Hariharan for the survey's database development; and

Joe Miller, Lori Tyahla, Joy Colucci, Larry McGoldrick and Yun-Chi Lu for appropriate criticisms and useful suggestions.

2. Survey Development and Analysis Methods

2.1 Introduction

To completely understand the outcome of the EOSDIS Product Use Survey, it is necessary to describe the development and review process that the survey underwent. This section has been divided into survey development methods and data analysis methods. Analysis methods will cover sections on procedures to ensure data quality, consistency of the results and methods related to product pull analysis. To manage the large output of the survey a *FoxPro* data-base was developed.

2.2 Survey Overview

The EOSDIS Product Use Survey was designed to understand the need for, and use of, data products that will become available through EOSDIS in the years 1998-2000. The survey included products that will be available in the Release B time frame as well as DAO, SAGE III, ERS, JERS, and RADARSAT. Products migrating from Version 0 and other heritage systems were not included in the survey (Release A).

The EOSDIS Product Use Survey development included four major review and modification phases. The survey was developed and reviewed internally by the Hughes Team; reviewed and tested by over 30 external reviewers including members of the science community, ESDIS, DAAC representatives, and survey experts; reviewed by ECS instrument team representatives; and reviewed by DAAC managers, user services representatives and science community representatives. At each stage of review all comments were tracked, evaluated and used to develop the next version of the survey. Reviewers were also apprised of how the survey had been revised in response to their inputs. This rigorous review process, though time consuming and labor intensive, proved valuable, and resulted in an overwhelmingly positive response from the science community.

2.2.1 Survey Development

The survey was originally developed using a paper prototype (Shneiderman 1992), and was reviewed extensively by several Hughes Team representatives for concept, useability and content. Multiple revisions of the survey were made in response to products listed, questions asked, and concern over the amount of time which would be required of respondents. After a final paper prototype was developed it was reviewed by the University of Maryland Survey Research Center for validity and reliability. Comments were incorporated into the next version which was implemented on the World Wide Web (WWW).

After the initial WWW version of the survey was developed, it was internally reviewed and tested by the ECS Science Office User Characterization Team. Modifications were made and an extensive external test was conducted. Thirty individuals selected from the science community, ESDIS, the DAACs, and the ECS Science Office were asked to review the survey and provide detailed suggestions on how to improve the survey to make it more user friendly and precise. An electronic Useability Exit Survey was provided at the end of the survey as well as a "comment" or "suggestion" field (see Appendix A, Table 1). Each comment from the review process was reviewed and addressed appropriately. Most comments were addressed in the subsequent version, and only a few could not be addressed due to the limitations of WWW technology, or policy/political considerations.

External test participants included members of the science community (David Glover, Bruce Barkstrom, and 8 professors and graduate students at the University of Maryland); ESDIS representatives (Chris Daly, Yun-Chi Lu, Marti Szczur, Frank Rockwell, Steve Wharton, Greg Hunolt); DAAC Science Liaisons (8 representatives); User Services Working Group (12 representatives); ECS Science Office Personnel (15 representatives); University of Maryland Survey Research Center; and many others including graduate students from the University of Maryland Geography Department, and students from George Washington University's graduate level class on Human Factors in User Interface Development. In addition, the Chairman of the Data Panel was given the survey URL on 1/10/95 and asked to review the WWW implementation. Most comments could be addressed, however responses to some comments were limited by resources, World Wide Web architecture or policy considerations. Some of the changes that were made as a result of the review process are as follows:

- products were re-ordered within sub-disciplines;
- added a summary screen with only the product name, level & instrument for each sub-discipline;
- spelled out all abbreviations;
- enhanced consistency of display of products, level, etc.;
- increased clarity of directions;
- constructed more descriptive pull down menus;
- reviewed data product lists;
- verified Level and Instrument descriptions; and
- added more detailed product descriptions.

After incorporating all of the comments from the external test and review described above, a third development and review process was initiated that focused on including detailed descriptions of all data products, checking the technical accuracy of the data products, incorporating all of the specific products supported by Release B, as well as verifying the grouping of products into disciplines. This process was both internal to the Science Office and external with instrument teams as necessary. The Science Office has representatives that work closely with most instrument teams and are involved in meetings, included in documentation distributions, and are kept up to date on changes in the products. These individuals were asked to review the product descriptions and provide the most current information available. New data products were added (DAO, SAGE

III, ERS, JERS, and RADARSAT), descriptions for all products were collected and included, and survey revisions were reviewed by those who contributed new data.

The final version of the EOSDIS Product Use Survey was reviewed and tested by the User Characterization staff (5 representatives); the ECS Science Office Manager, and the ESDIS Science Office Chief. In addition, the survey was also reviewed by User Services representatives at each of the DAACs, and was forwarded to Bill Emery and George Emmitt for review. Only a few questions were received from the User Services Working Group representatives; and final minor text revisions were made.

2.2.2. Survey Effectiveness

An ad hoc measurement of interest for future survey development efforts concerns the time that respondents needed to complete the EOSDIS Product Use survey using Mosaic or Netscape. The electronic Useability Exit Survey during the development survey phases requested the users to estimate completion time. The mean completion time was then calculated from the responses and a distribution of response time was developed. The time chosen for completion in the final version of EOSDIS Product Use survey was estimated at 30 minutes.

2.2.3 Administration of the Questionnaire

Most surveys randomly sample a small subset of a large population of potential respondents. Such samples are carefully crafted to avoid the cost of surveying the entire population while producing conclusions that are representative of the population. It was the intent of this survey to obtain a good statistical sample of potential ECS science users in order to assess the relative demand for products. ECS sought earth scientists' names and internet addresses from EOS/Transactions AGU Meetings for the 1994 and 1995 years, from the DAAC user lists, from home pages of university departments, from the OMNET scientist list and from the most recent EOS Science Directory (1995). The list of people who were e-mailed messages was adjusted to represent the appropriate proportion of users in each discipline according to previous studies about the earth science population (Tyahla, 1994). This controlled population bias due to responses from users of just one discipline.

In addition to this outreach, an announcement was posted on several news groups regarding the EOSDIS Product Use Survey (Table 2.1). Announcements were also posted on DAAC home pages and in two journals: EOS/Transactions AGU and The Earth Observer (EOS publication).

Beginning on April 24, 1995, ECS sent e-mail messages to the scientists inviting them to complete the survey and posted the EOSDIS Product Use Survey on the WWW (http://observer.gsfc.nasa.gov/egsus/intro.html). ECS sent out a total of 4700 e-mail messages, and received 449 responses during the first phase (5/24/95) and an additional 146 (595 in total) during the second phase (7/24/95).

Table 2.1. News Groups

sci.aeronautics	sci. environment	sci.image processing	bionet.general	
sci.astro	sci.geo.eos	sci.space tech	nasa.infosystems.www	
sci.astro research	sci.geo. geology	sci.space news	comp.human-factors	
sci.biology	sci.geo.hydrology	sci.space science	comp.infosystems gis	
sci.bio ecology	sci.geo.meteor.	sci.system	comp.specification	
sci.education		sci.geo.ocean.	alt.sci. planetary	comp.software-Eng.

2.3 Data Analysis Overview

The survey was open to all WWW users and some responses were given by surfers, browsers and other non-science people. Since the survey was primarily directed towards the general earth science community, those responses from non-science people not included in any analysis. In addition, the database was sorted into two categories:

A. Responses which contained at least one data product request;

B. Responses which did not request any data product.

In our final analysis from a population of 595 user hits, 375 responses fell into category A and 220 responses into category B. In our analysis, only category A was used which had a total of 6212 data requests.

2.3.1 Preparation of Responses

The data were checked for single and unique inputs from each respondent. When different users used the same machine (with the same internet [IP] address), data were cross-checked using their name and other user details. When a user responded more than once, the latest version was considered correct. In addition, some users had no entries in some fields, such as discipline and title. The missing information was searched for in the Conferences Abstracts for the years 1994 and 1995 (AGU-Spring and Fall, IUGG and TOS), Membership Handbook of ASLO, and WWW home pages of Universities and Laboratories. Also, some were contacted by email to acquire missing information.

2.3.1.1 Correction for Temporal Resolution

It was assumed that a person would not order the same data (by space and time coordinates) more than once. Some data requests did not match the possible time availability of the product and were adjusted. For example, if a product is ordered daily (i.e., maximum of 250 days/year = 365-holidays - weekends) but is produced only every 15 days, the user cannot order data daily but only once in 15 days. In this case, the product order frequency was adjusted from 250 days/year to 24 times/year (250 days per year/15 days). In addition, some products had more than one temporal resolution. In these cases, the lowest temporal resolution of a product was taken to be the minimum order frequency and the highest temporal resolution of a product was taken to be the maximum order frequency (Table 2.2). The temporal adjustment was applied only to the order frequencies and not to browse frequency because a user can browse the same data more than once.

Table 2.2. Temporally Adjusted Data

Product	Order Freq. Response	AvailableTemporal Resol.	Min. Order Frequency Days/year	Max. Order Frequency Days/year	Adjusted Min. Order Frequency Days/year	Adjusted Max. Order Frequency Days/year
AST13	Monthly	1/16 days	11	24	Min(365/16 and 11) = 11	Min (365/16 and 24) = 23
MOD28	Monthly	1/day, 1/week, 1/month	11	24	Min(365/12 and 11) = 11	Min (365x1, 24) = 24
CER05	Rarely	1/hour	0	1	Min(365x24 and 0) = 0	Min (365x24 and 1) = 1
MOD28	Daily	1/day, 1/week, 1/month	101	250	Min(365/12, 101) = 30	Min (365x1, 250) = 250
...	...	...	...	...	...	...
Totals			$\Sigma_n(x)_{min\_or}$	$\Sigma_n(x)_{max\_or}$	$\Sigma_n(y)_{min\_or}$	$\Sigma_n(y)_{max\_or}$

where y1, y2,..yn are the new estimates.

2.3.1.2 Science Population Evaluation

Complete survey responses, from category A (375) were further examined to evaluate possible survey science discipline bias. It is possible that respondents in some disciplines were quicker in answering the survey and were able to complete the survey in less time. When 4700 e-mail notifications were sent to science users after completing the survey, the proportion of users in each discipline was kept in mind, and this may have helped to control this type of bias. To determine whether a discipline bias did occur, the proportion of respondents in each discipline was calculated and compared to previous discipline studies of the earth science community (Tyahla, 1994). Since the resulting discipline proportions were similar to estimates from earlier earth science population studies (Tyahla, 1994; and McGoldrick, personal communication) a population "weighting" factor was not applied.

2.3.1.3 Characterization of Respondents

Complete survey responses from category A (375) were further examined to characterize the respondents in terms of discipline, home institution, profession, academic status, and research interest by spatial and temporal scales. With the exclusion of the personal information and research fields, which were individually examined, other characterizing information such as response certainty, system usage etc. were directly extracted from the *FoxPro* database, and proportions in each category calculated.

To evaluate the academic status of each respondent, additional information was located on the WWW home pages of Universities and Laboratories. Respondents were grouped into academic status according to classes. The classes included associate scientist, professor, student, technical staff and other. The "associate scientist" category included associate scientists, researchers and

Post Doctoral appointees. The "professor" category included assistant and full professors. The "student" category included graduate and undergraduate students. The "technical staff" category included support staff, programmers and data managers. The "other" category included other, program managers, geologists, and naturalists. Global Change Fellows and IDS team members were grouped separately. The percentage of respondents was calculated for all the above classes [(no in each class / total no of respondents) x 100].

Respondents' temporal scale of research interest were individually examined and grouped by classes. The classes for temporal coverage were as follows: event 0-11 months, 1 year, 2-9 years and decades. The classes for spatial coverage were directly pre-determined from the survey and were as follows: smallest (under < 1,000 km^2), small (from 1,000 to 10,000 km^2), medium (from 10,000 to 1,000,000 km^2), and largest (> 10,000,000 km^2).

2.4 Product Pull Analysis

Before analyzing the relative pull on individual products, an evaluation was done on response quality to discover hidden biases. Several approaches were taken. For example, individual answers were examined for internal consistency and particular groups were identified, such as EOS investigators, and examined to determine if they might have a product pull different from that of the general science community.

In our sampled population we had twenty-eight responses out of 375 from EOS investigators in the following categories: Interdisciplinary Investigation Teams (9), Instrument Teams (7) and Global Change Fellows (8) (Appendix A, see Table 2 and 3). This EOS group was statistically compared with the rest of the responses. No introduced bias or "ballet box stuffing" was detected.

2.4.1. Browse and Ordering Frequency

The EOSDIS Product Use Survey used time ranges as a measure of demand frequency (Table 2.3). A detailed inspection of individual answers revealed that some responses did not match the time availability of the product . A temporal adjustment was than applied only to the order frequencies and not to browse frequencies (see section 2.3.1). Browse frequency results should be considered with caution. We received the following comment from one respondent: "I suggest that you add a supplementary question regarding the extent of browse activity". It seems that many respondents were confused as to what browsing referred to, whether they would browse examples using the product or the actual product granules. For the above reasons, browse data was not analyzed according to individual disciplines.

Table 2.3. Order and Browse Frequencies in the Survey

	Min. Days/Year	Max. Days/Year
Daily	101	250
Weekly	25	100
Monthly	11	24
Quarterly	3	10
Annually	1	2
Rarely	0	1

2.4.2 RPAF Computation

The Relative Product Access Frequency (RPAF) for each product was computed as a ratio of the number of days a product is accessed to the total number of product accesses. It was observed that the minimum and maximum days/year values are not distributed in the same proportion for the time ranges. As an example, for a daily request, the minimum number of days/year is 101 while the maximum is 250 days/year, their ratio being 1:2.475; for a weekly request, the minimum number of days/year is 25 days/year and the maximum is 100 days/year, a ratio of 1:4. This variability in the ratio made the minimum RPAF for some products greater than their maximum estimate and vice versa. For computing the RPAFs for the products, the total number of requests for each product was computed as the sum of the days/year for all requests of that product. The ratio of this total to the total number of requests for all products resulted in the RPAF for a product. The process is described below:

Let $x_1, x_2, ..., x_n$ be the n products. Let $\Sigma_n(x)_{min}$ be the sum of the number of days per year for all products with minimum values substituted. Similarly, let $\Sigma_n(x)_{max}$ be the sum of the number of days per year for all products with maximum values substituted (Table 2.4).

Table 2.4. Computing RPAFs

Product	Browse Freq.	Order Freq.	Browse Freq. Min. Days/year	Browse Freq. Max. Days/year	Order Freq. Min. Days/year	Order Freq. Max. Days/year
AST13	Daily	Monthly	101	250	11	24
MOD28	Monthly	Monthly	11	24	11	24
CER05	Weekly	Rarely	25	100	0	1
MOD28	Daily	Daily	101	250	101	250
...	...	...	...	...	...	...
Totals			$\Sigma_n(x)_{min\_br}$	$\Sigma_n(x)_{max\_br}$	$\Sigma_n(x)_{min\_or}$	$\Sigma_n(x)_{max\_or}$

Now, we have two RPAFs, a maximum and a minimum. They will be :

$RPAF_{min-br}(x_t) = \Sigma(x_i)_{min\_br} / \Sigma_n(x)_{min\_br}$

$RPAF_{max-br}(x_t) = \Sigma(x_i)_{max\_br} / \Sigma_n(x)_{max\_br}$

$RPAF_{min-or}(x_t) = \Sigma(x_i)_{min\_or} / \Sigma_n(x)_{min\_or}$

$RPAF_{max-or}(x_t) = \Sigma(x_i)_{max\_or} / \Sigma_n(x)_{max\_or}$

where: x_i is the i^{th} product,

$\Sigma(x_i)$ = No. of requests per year for product x_i and

$\Sigma_n(x)$ = Total No. of Requests per year for all products.

Since the $RPAF_{min-or}(x_i)$ for a individual product can be greater than the $RPAF_{max-or}(x_i)$ and vice versa, to estimate the average RPAF for product x_i for ordering, the Max.($RPAF_{min-or}(x_i)$, $RPAF_{max-or}(x_i)$) was computed. Obtaining this for all the products, the RPAFs were normalized so that comparisons could be made between the products. The same process is repeated for browse using $RPAF_{min-br}(x_i)$ and $RPAF_{max-br}(x_i)$ instead of $RPAF_{min-or}(x_i)$ and $RPAF_{max-or}(x_i)$.

2.4.3 Ordering Factor

The number of products from each instrument varies. The range of numbers of products in the EOSDIS Product Use Survey varied from a minimum of one product on Landsat-7 ETM to a maximum of 40 products from the MODIS instrument. Because of this variation in the number of products, the demand for an instrument (or DAAC) can be misinterpreted when viewed by product demand. This bias can be removed by introducing an Order Factor as a measure of popularity of different levels that takes into account the difference in number of product levels from each instrument. The Order Factor is the ratio of the RPAF for a product level, to the percent of products from that level. It is clear that the break-even factor has a value of 100%.

3. Results and Discussion

3.1 Overview

The first section of the results will focus on the representativeness of the survey responses and will discuss the survey effectiveness by response time, by characteristic of the respondents such as home institution affiliation, respondent status, discipline distribution and research interest. The second section will then discuss the product pull results.

3.2 Response Times

The time estimate was voluntarily provided by the respondents in the last question of the survey, only 287 respondents (out of 375) provided this value. The actual mean time that respondents needed to complete the survey using the WWW was 20.6 minutes (standard deviation 11.2), with 90% of respondents taking 30 minutes or less to answer the survey and 70% taking 20 minutes or less (Figure 3-1). This suggests that we succeeded in fielding a survey that can quickly be answered. Furthermore, the data indicate that most of the respondents must have selected a period of time without interruptions in order to complete the survey. Only 2% of users took over sixty minutes for completion (maximum being 90 minutes); this time included periods in which the respondent was apparently distracted from the survey.

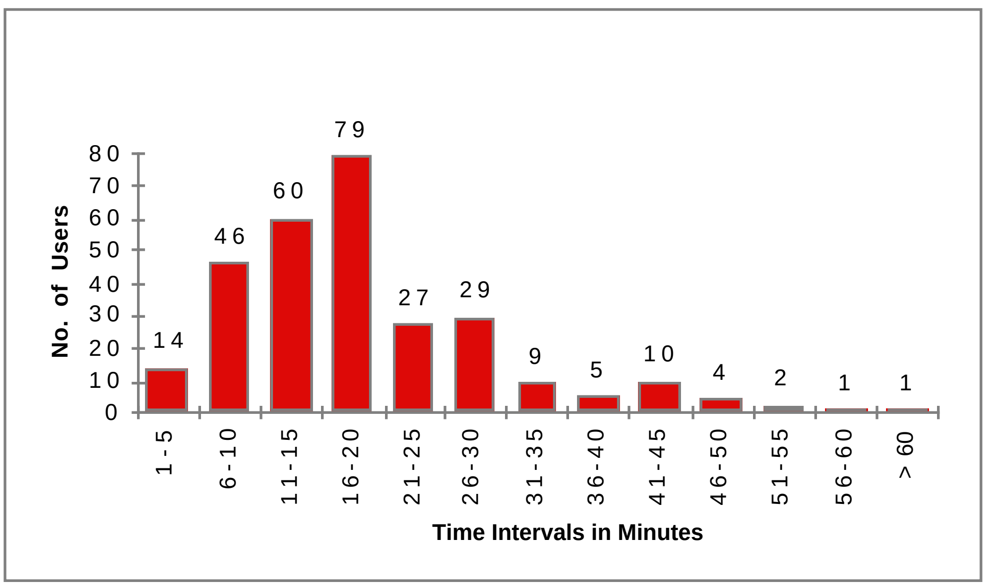


Figure 3-1. Time Taken for Survey Completion

3.3 Characterization of Respondents

One of the goals of the EOSDIS Product Use survey was to obtain a representative sample of the science community which consists of scientists; students; state and federal agency scientists; and commercial corporation scientists. The characterization of respondents is presented below.

3.3.1 Distribution of Responses Across Discipline

The community sampled was well distributed among disciplines. This balanced result is in part due to the WWW interface, outreach e-mail messages and announcements. Of 375 responses we found 49% of users performed interdisciplinary research, 40% single disciplinary research and 11% provided no information on their discipline (i.e., null). After computing the survey results relative to discipline size, the survey showed discipline proportions very close to the actual proportions claimed by previous studies of science discipline demographics (Tyahla, 1994; and McGoldrick, personal communication, 1995). The distribution of science users to various disciplines is shown in Table 3-1.

Table 3-1. Distribution of Users by Disciplines

User Disciplines	Number of hits Final Data 7/24/95	EOSDIS Survey %	McGoldrick Data,1995 %	Tyahla Data,1994 %
Atmosphere	111	33%	35%	50%
Land	83	25%	22%	27%
Ocean	90	27%	25%	18%
Cryosphere	16	5%	8%	1%
Public Policy	5	1%	2%	
Other	29	9%	8%	
Total	375			

3.3.2 Home Institution of Respondents

In addition to the community being well distributed among disciplines, the responses received were from prestigious home institutions. Our data base had 59 US universities, 28 foreign universities, 47 USA federal and government laboratories, 13 foreign laboratories and 11 commercial corporations (see Appendix B for detailed list, Tables 1, 2 and 3).

Among the US respondents were scientists from Massachusetts Institute of Technology, Scripps Institution of Oceanography, California Institute of Technology, L-DGO of Columbia University, Princeton University, Cornell University, Stanford University of California, University of California at Berkeley, University of Miami, Colorado University, Texas A&M University and others. The non-US respondents were scientists from the Max Plank Institute for Meteorologie, World Meteorological Organization, British Geological Survey, Canadian Atmospheric Environment Service, University of Dundee, European Commission's Joint Research Centre and others. Commercial corporation responses included scientists from Mobil Oil and IBM, among others.

The international proportion was 11% of the total respondents, although some of these institutions had multiple users, and so this proportion could be higher. System usage by foreign institutions is expected to grow in the future because of the increased connectivity abroad. Some DAACs are already showing an increase in international internet traffic (see statistics at P.O. DAAC), most likely facilitated by WWW technology.

3.3.3 Distribution of Respondents Across Profession

The responses originated mainly from an academic population of users (Figure 3-2). Our sample is quite different from previous surveys (Wingo, 1994). First, the detailed information we received in such a short span of time gives us a representative "slice" of the academic community. Second, our responses are originated mainly by active researchers and so the data under-represents the student population (higher faculty to student ratio). And third, our respondents have high academic status. The distribution is as follows: associate scientists (46%), professors (11%), students (8%), technical support staff (6%), PIs and co-Is of Interdisciplinary Investigation Teams and Instrument Teams (5%), Global Change Fellows (2%), engineers (4%), other users (13%) and null (5%). Given the above proportions, the data is a good representation of the academic- and research-oriented earth science community that should be responsible for a high percentage of the product pull.

3.3.4 EOS vs Non-EOS Scientist

Seven percent of the population sampled was made up of EOS investigators; quite close to the expected predicted proportion of the total population (4%) (Jarvis, personal communication; Tyahla 1994). The EOS group presented similar percentage distributions by level, product, instrument or DAAC when compared with the survey analysis. No introduced bias or ballet box stuffing was detected (see section 3.4.2) and the EOS group was statistically similar with the rest of responses.

3.3.5 Research Interest by Spatial and Temporal Scales

The heterogeneity in the sampled earth science community is also reflected by the selected science interests in space and time scales. Research interests span areas smaller than 1000 km^2 to global scales with time spans from discrete events to several decades. Figure 3-3 presents research interests by space and time for all the surveyed disciplines. It should be noted that any given discipline contains researchers with common interests. In addition, some disciplines may concentrate interest in particular scales (e.g., global researchers tend to require data spanning decades, which has been monthly averaged). Further analyses of research interest by discipline reveal and confirm trends in global , regional and small scale research typical of some disciplines.

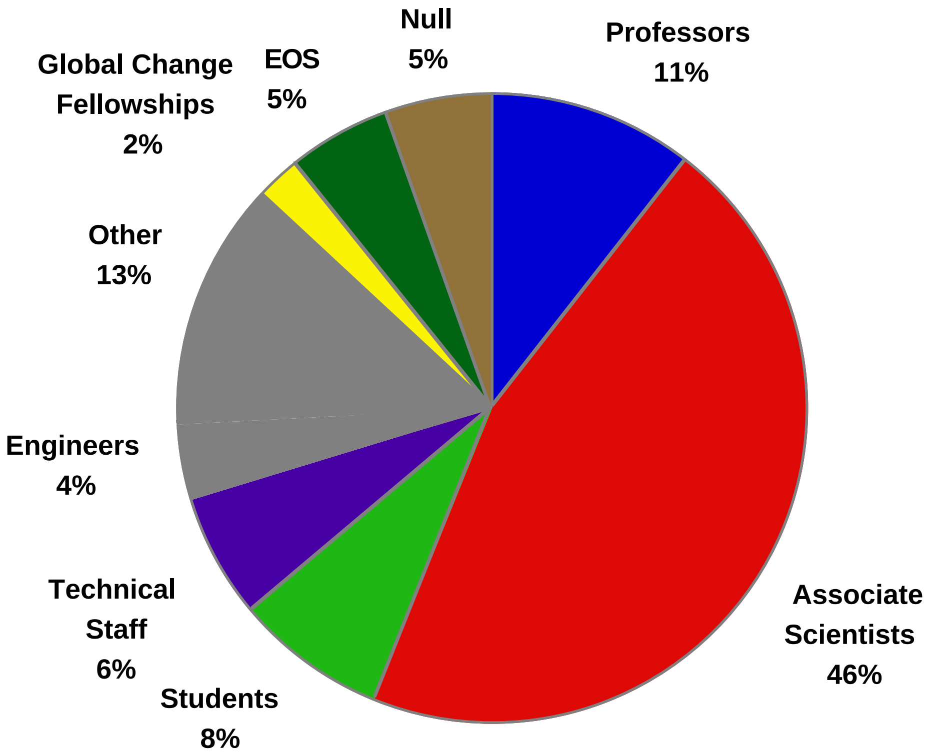


Figure 3-2. Distribution of Users by Line of Work

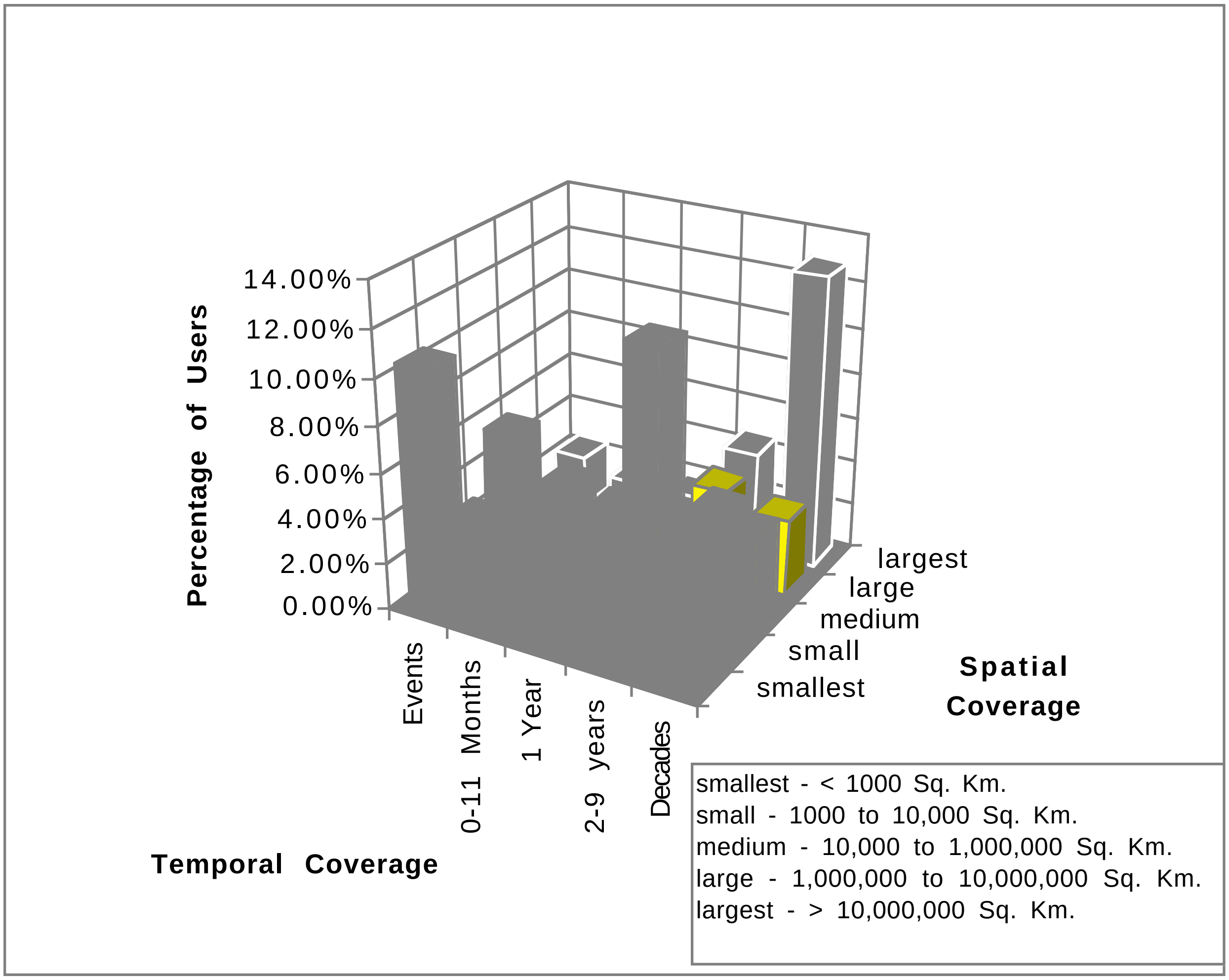


Figure 3-3. Research Interest by Spatial & Temporal Coverage

Figures 3-4 and 3-5 illustrate the spatial coverage and temporal resolution results. Survey results on spatial area indicate that 49% of surveyed users are interested in areas smaller than 10,000 km^2 (Figure 3-4). This clearly illustrates the importance of subsetting satellite data collected over large areas.

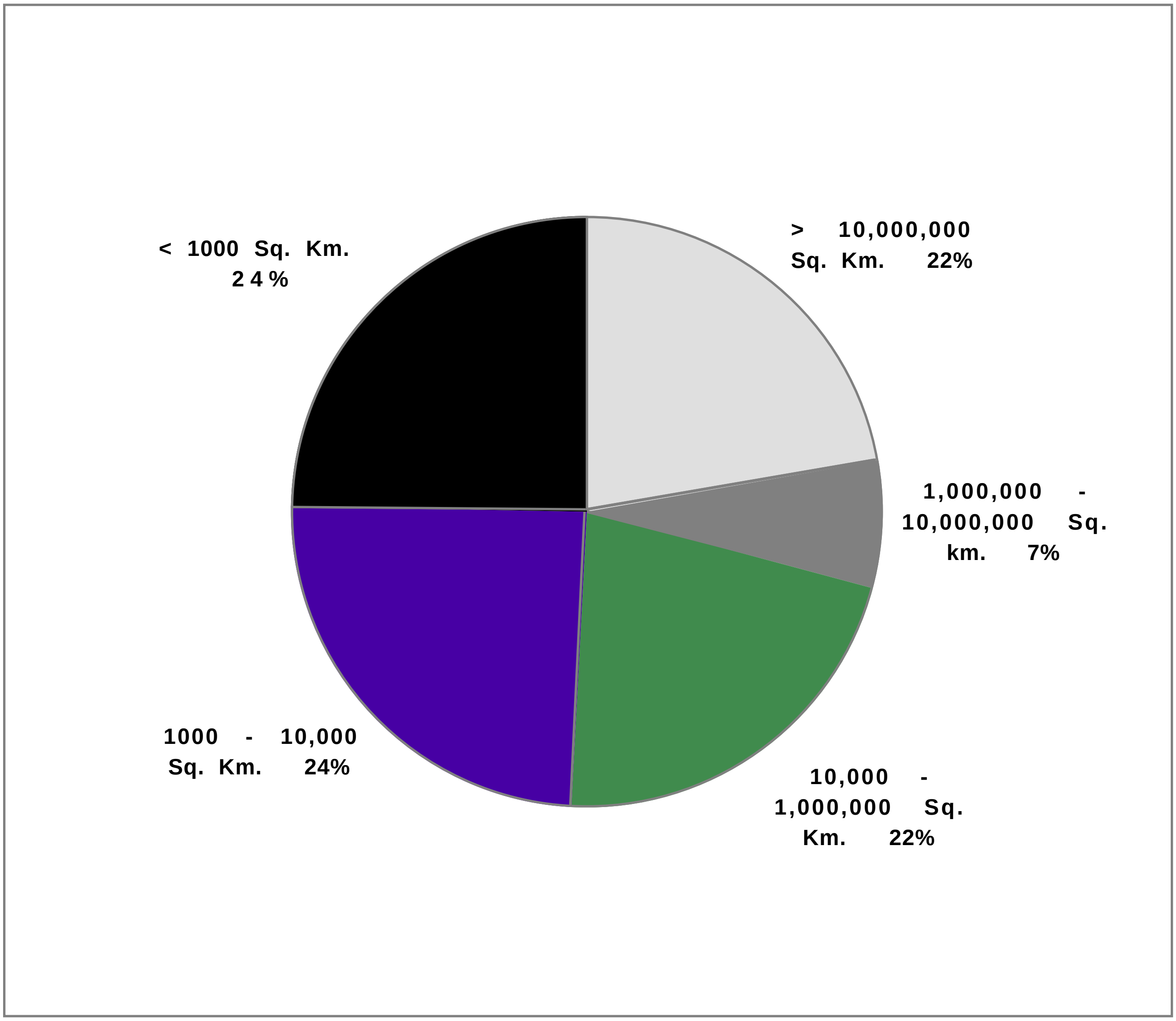


Figure 3-4 . Distribution of Required Spatial Coverage

Survey results on temporal resolution indicate that 55% of respondents require data at a high frequency (hourly 12%, daily 29% and weekly 14%) while there is less access demand at monthly (22%), seasonally (13%) and yearly (10%) frequencies. These results could be due to the dynamic nature of the process being studied or that scientists collect data at a higher frequency to avoid sampling bias, and prefer to average the data later.

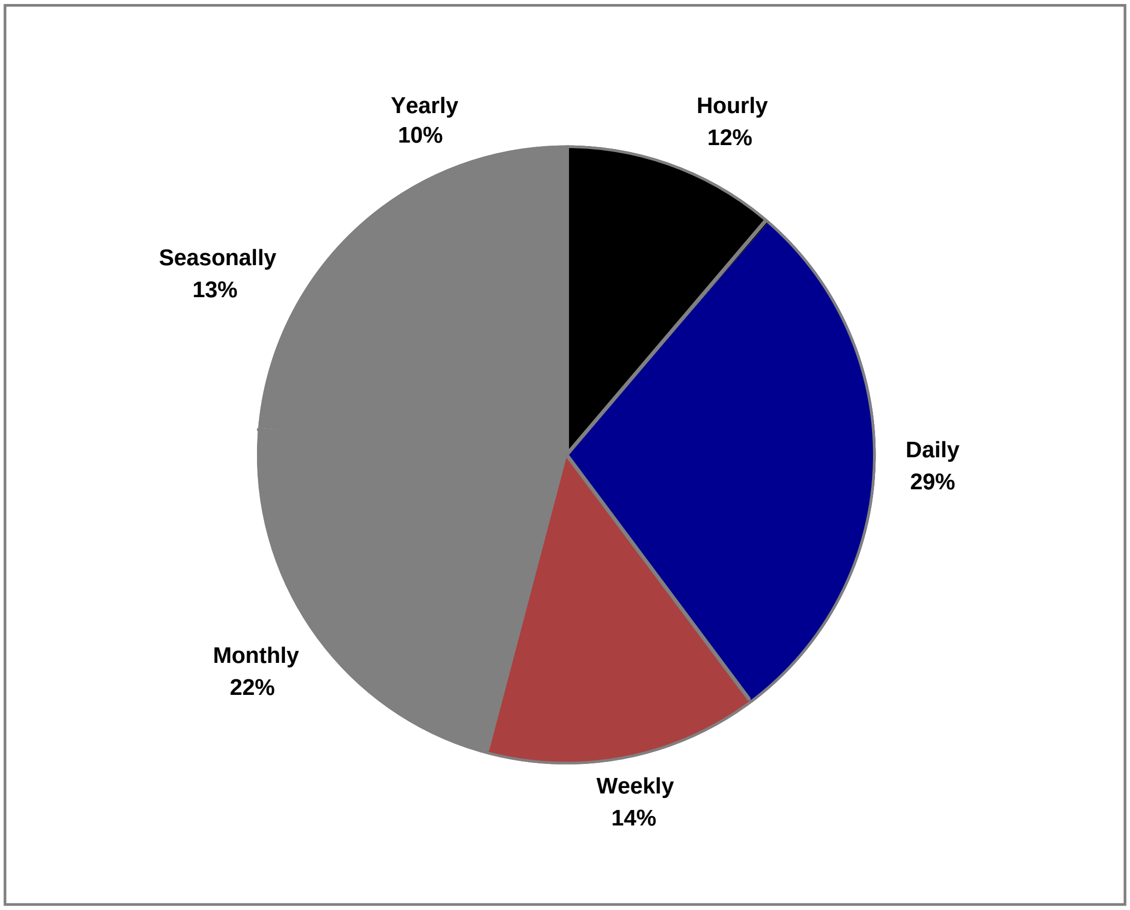


Figure 3-5. Distribution of Required Temporal Resolution

3.3.6 Distribution of users by System Usage

Figure 3-6 indicates that 58% of users will locate and visually inspect data (16% perform analysis, 20% request data and 6% produce). This can be interpreted that a large fraction of users' time will be spent on inspecting data (presumably through browsing) in order to test hypotheses, acquire new ideas and compare data.

Users will access EOSDIS in the following ways: 2% by phone, 63% by user interface and 35% by automated process (direct machine-to-machine). The direct machine-to-machine percentage is quite high relative to a previous study (Tyahla, 1995). Several factors could have played a role in this discrepancy: one or both of the questions were not clearly formulated, or the advances in technology and a more computer-literate population make this type of access easier. Unfortunately, the reason for this discrepancy was not entirely clear.

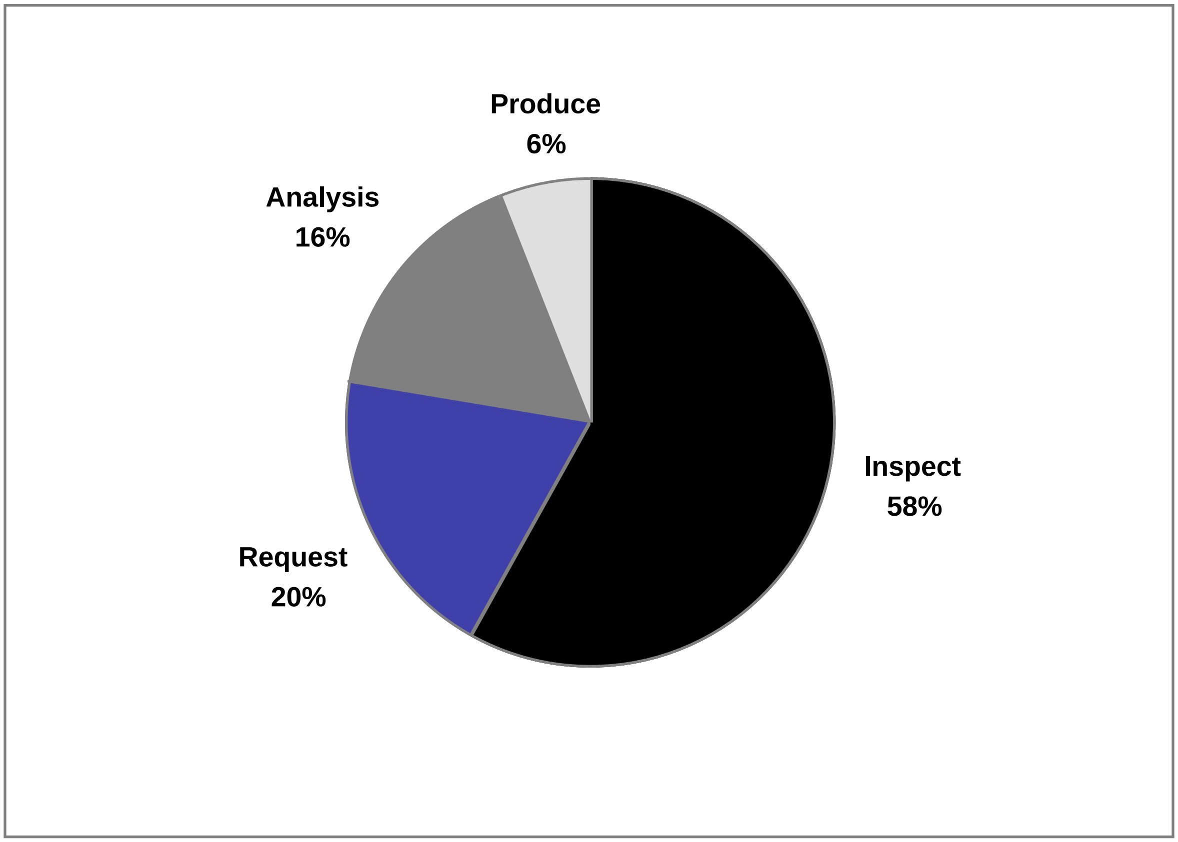


Figure 3-6. Distribution of Users by System Usage

3.3.7 Certainty of User Projections

To evaluate the confidence of the responses received, the survey included a question relating to the users' certainty of the correctness of their responses. Figure 3-7 shows that over 50% of respondents used educated guesses and 12% based their answers on certain funding, either current or future. The distribution of certainty reflects the high confidence level of the population sampled (and the reputations of the universities surveyed).

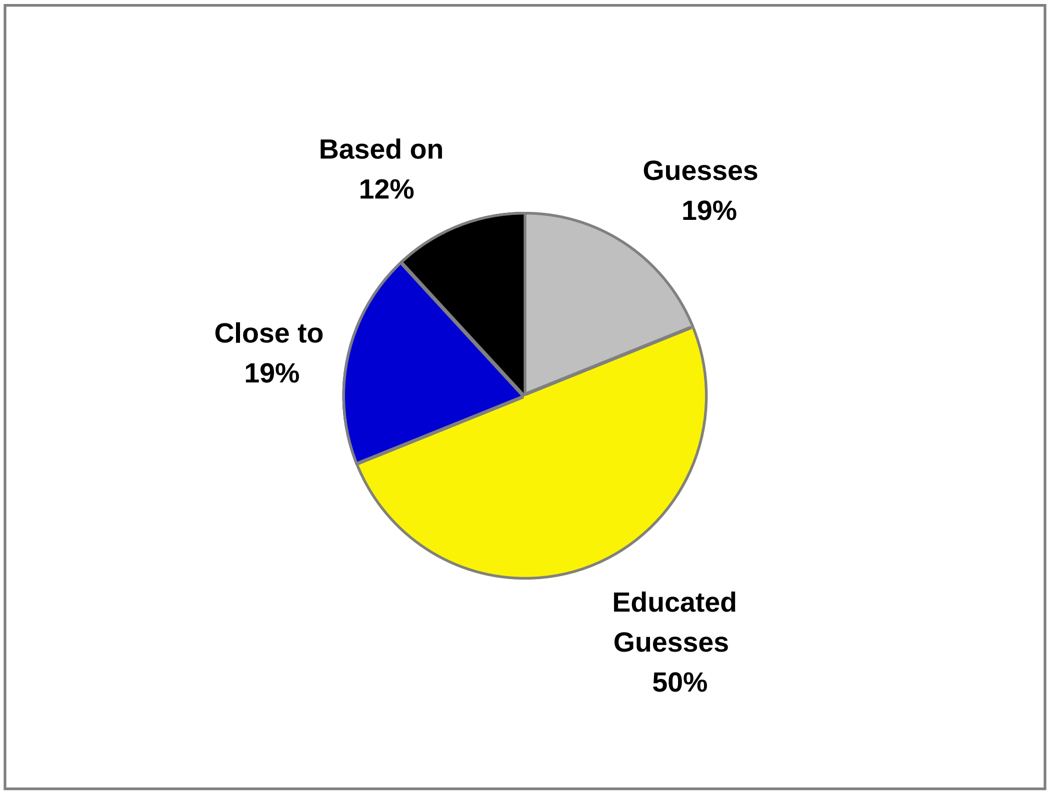


Figure 3-7. Certainty of Estimated Demand Projection

3.4 Product Pull Analysis

3.4.1 Browse and Ordering Frequencies Overview

Results of questions concerning browse frequency should be considered with caution (see method section 2.6.1) Figure 3-8 indicates that 73% of the respondents expect to browse infrequently: 25% monthly (11-24 times/yr.), 24% quarterly (3-10 times/yr.), 24% annually (1-2 times/yr.), 15% rarely (0-1 times/year). "Heavy browsers" represent 27% of the respondents: 11% daily (>100 times/yr.) and 17% weekly (25-100 times/yr.). The data indicate that users of the most dynamic products are the major contributors to the daily and weekly categories.

Ordering responses were analyzed as a whole and for each separate discipline. As expected, as a whole, the percentage of respondents ordering products at a higher frequency is lower than the percentage ordering at a lower frequency. For example, the percentage of users ordering products daily and weekly is 14%, and those ordering quarterly and annually is 42% (Figure 3-8). In general, users, as expected, order data less frequently than they browse products. Figures 3-9a and 3-9b illustrate the different patterns of ordering by discipline. It is noted that more dynamic products tend to have higher ordering frequencies.

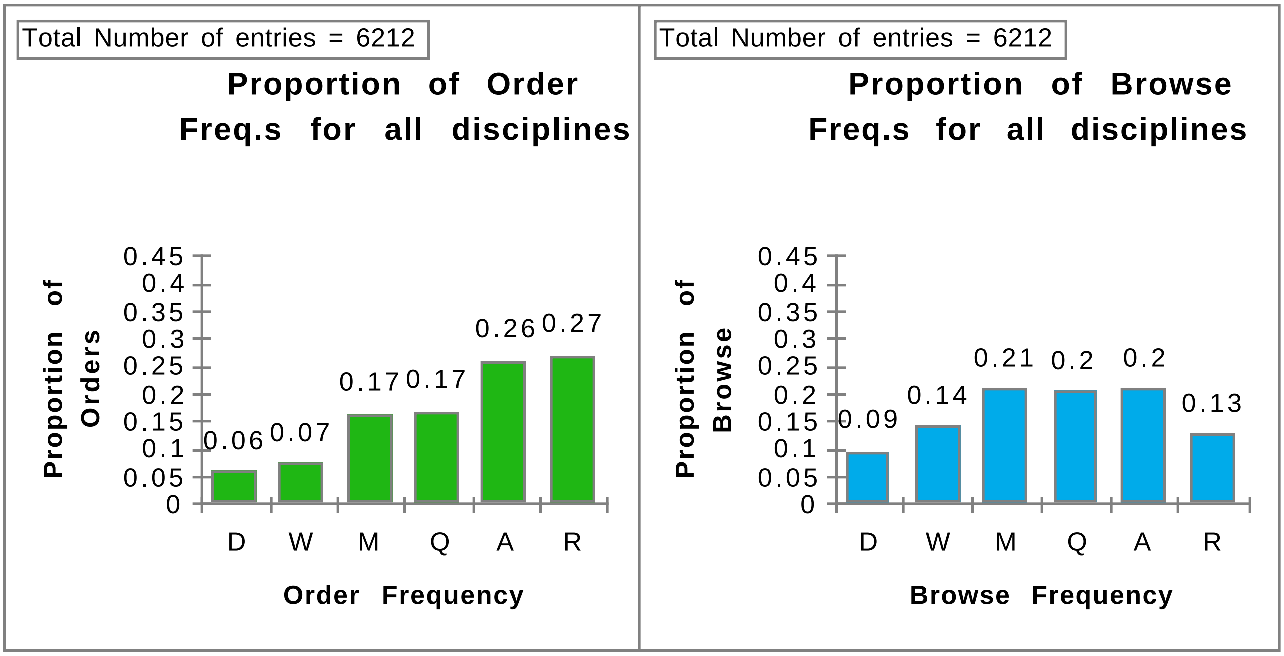


Figure 3-8. Percentage of Order and Browse Frequencies by Users

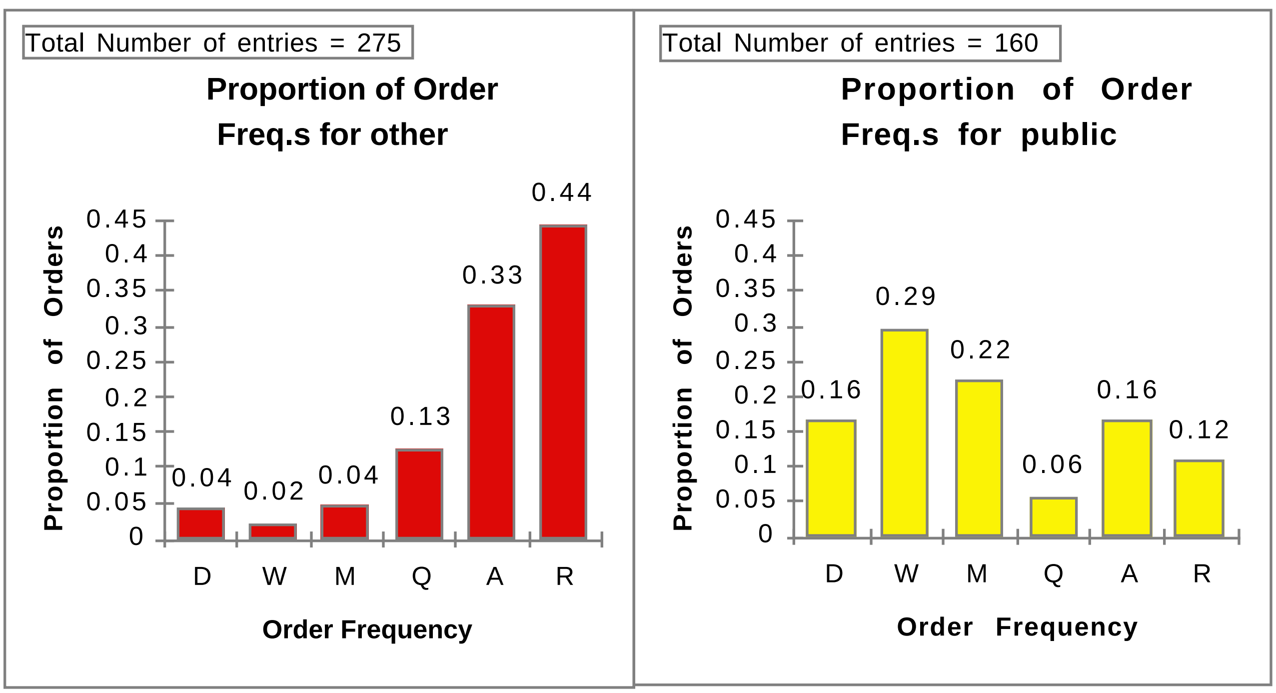


Figure 3-9a. Percentage of Order Frequencies by Discipline

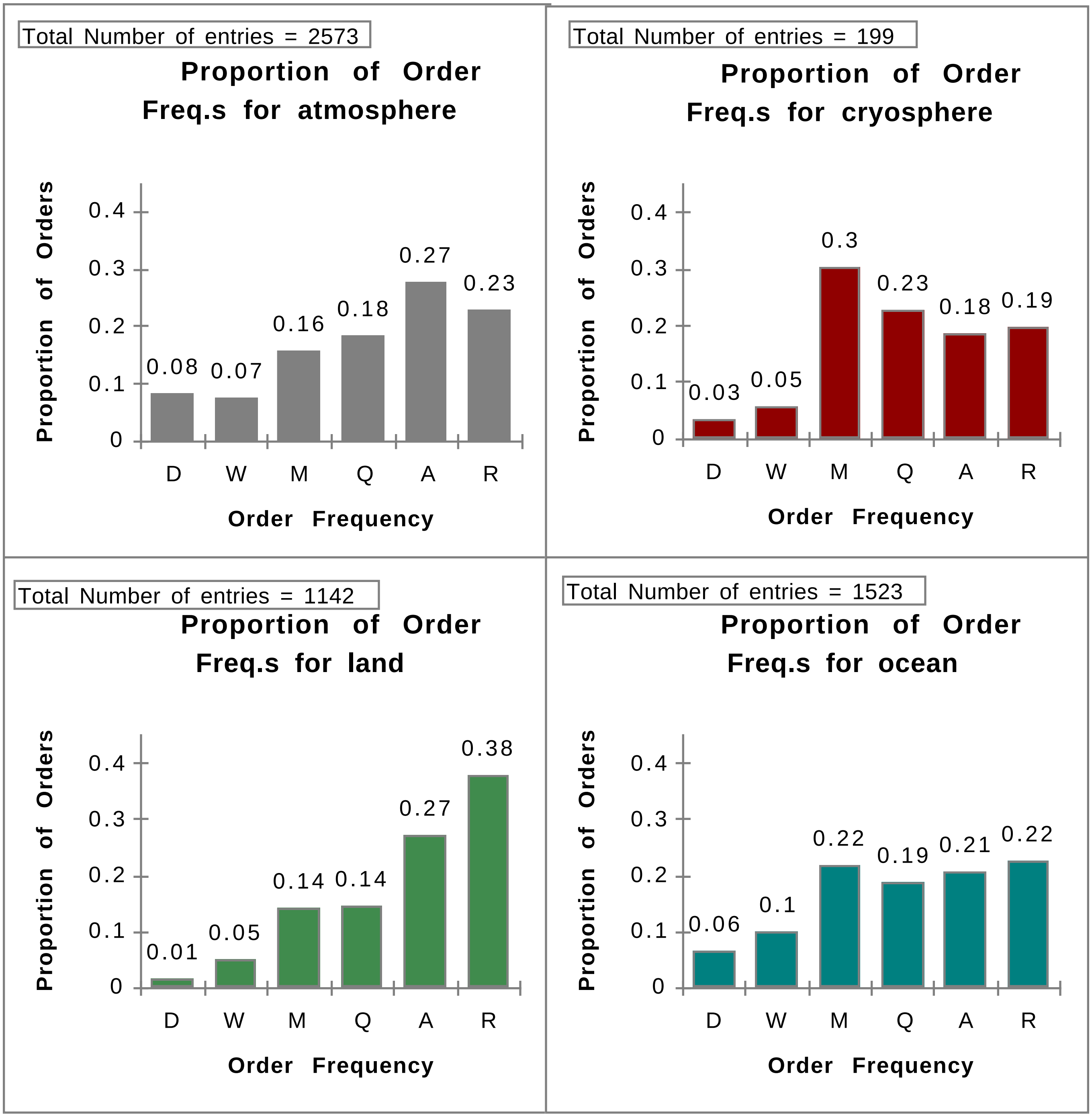


Figure 3-9b. Percentage of Order Frequencies by Discipline

Figure 3-9a illustrates the public policy users and "other" disciplines. Other disciplines seem to be ordering annually or rarely (77%) while the public policy group prefers the higher frequencies of daily and weekly (45%). In both of these categories, the sample size is small enough that statistical values should be considered with caution. Disciplines studying the land and cryosphere have low ordering frequencies, indicating relatively static, or slowly changing processes. The ordering pull is six times larger for the land than the cryosphere (1142 orders for land vs. 199 for cryosphere). Land orders tend to concentrate at annual frequencies (27%), followed by monthly and quarterly orders (roughly 14% each) while the cryosphere orders tend to be monthly (30%) followed by quarterly (23%) and annually (19%). The ocean and atmosphere disciplines make up a relatively large number of entries (greater than 1500 each) and have a more uniform distribution of ordering frequencies. These disciplines represent a large variety of research interests (from small to large scales and spanning from events to decades) which could, in part, explain why no particular frequency was preferred. Furthermore, in comparison to other disciplines, including land and cryosphere, the daily and weekly frequencies are higher. This could be explained by the dynamic nature of phenomena studied in the ocean and atmosphere.

3.4.2 Relative Product Access Frequency by Level

In general, the respondents preferred products with a moderate amount of processing (levels 2, 3 and 2&3) which indicates a high confidence in the processing procedures and teams (Figure 3-10). The interest in level four products may be misleading because there were only 13 level four products in the survey. In addition some products have been classified (in the survey) under a separate level "2&3". These products contain parameters that are available in different levels (2 or 3). Because of a significant number of entries (1061 out of 6790) for Level 2&3 products a decision was made not to distribute level 2&3 between levels 2 and 3. Also, some products did not have any levels specified; these products have been grouped into "to be determined" (TBD).

The NASA EOS Funded group was analyzed and presented similar percentage distributions by level (Figure 3-11). In general, lower level products require more storage space, time, programming and expertise, which tend to restrict the number of users, in all groups of respondents, interested in them.

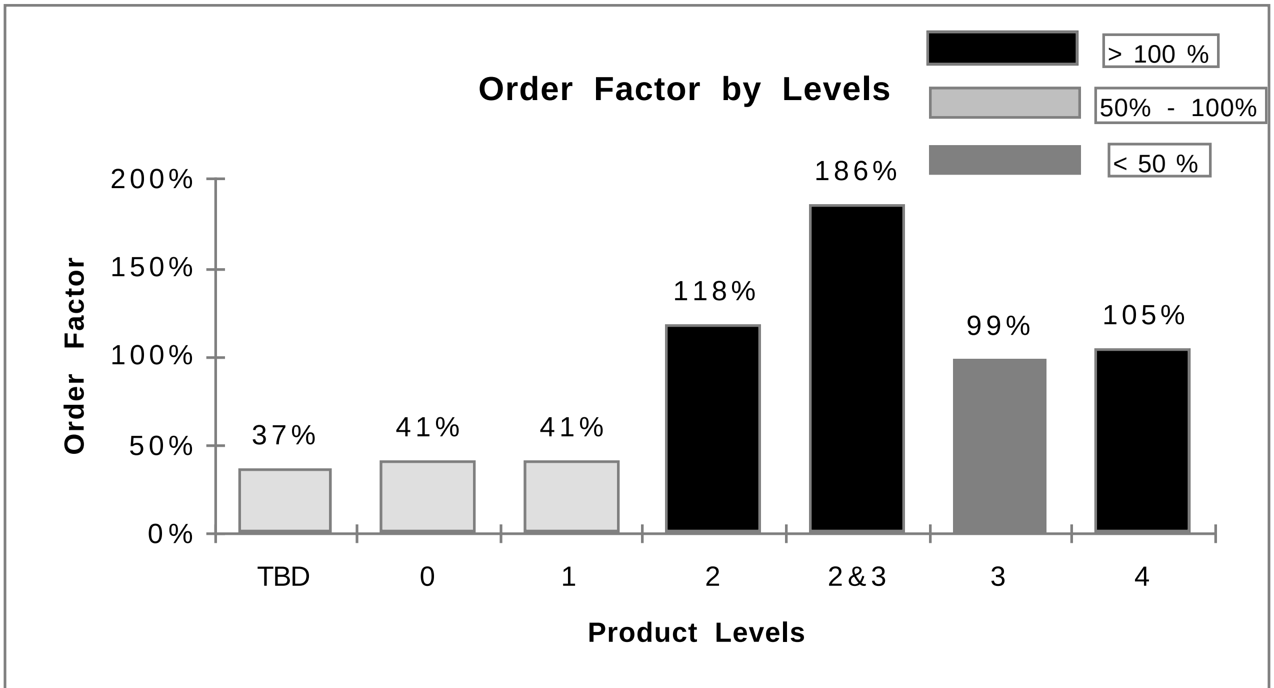
3.4.3 RPAFs by DAAC, Instrument and Product

There were concerns that respondents would view the survey as a "popularity contest" used to eliminate products and instruments. In order to prevent this occurrence, the following statements were made to the respondents about the use of the survey:

1) the Survey will be used: for statistical analysis purposes ONLY in order to assist system developers in gauging interest in each data product.

2) the Survey will be used: will NOT be used to change the list of planned data products.

The ECS project is aware that the survey results are accurate only to a certain degree and caution will be employed when extrapolating the results. Due to the sensitive nature of this material, the Science Office has determined some information to be confidential. Any interested parties should contact the author or Joy Colucci for authorization.



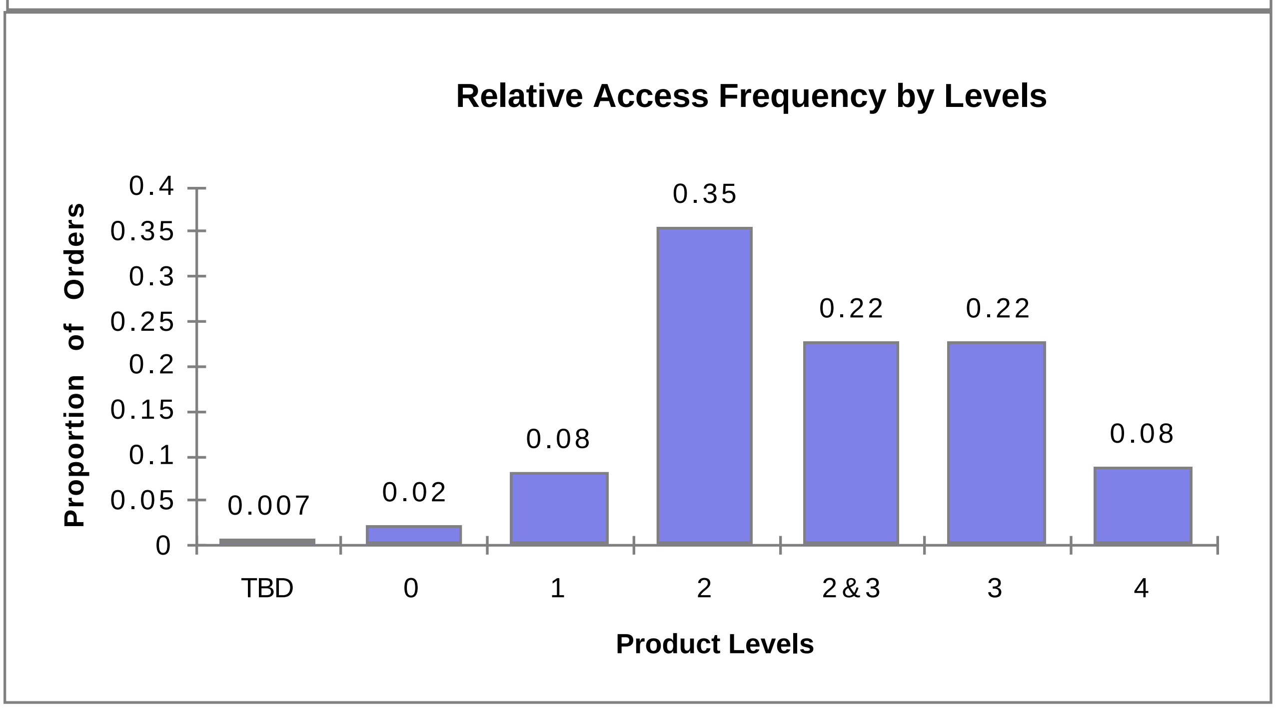


Figure 3-10. Ordering Factor and Relative Access Frequency by Level

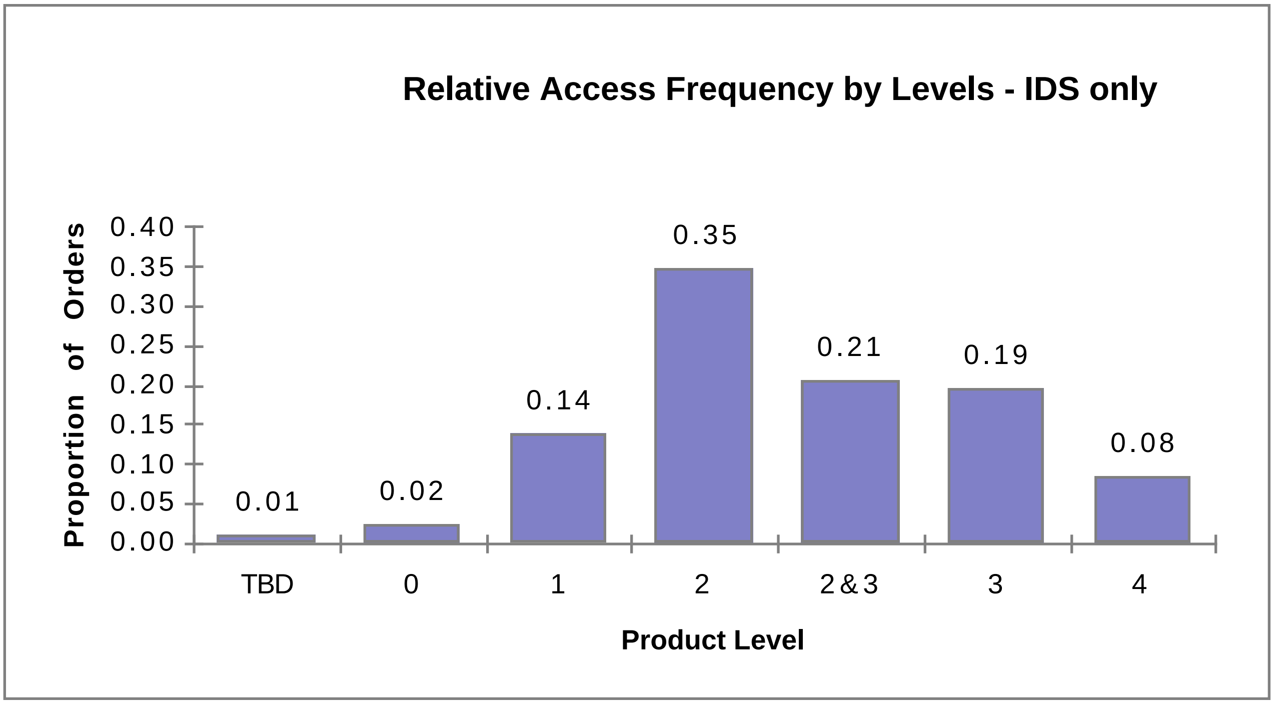


Figure 3-11. Relative Access Frequency by Level for NASA EOS Funded Respondents

4. Conclusion and Applications

4.1 Conclusions

The large response and the high quality of data collected are a result of the WWW interface and the broad advertisement. The data is consistent and very stable; marginal differences were seen between the preliminary results and the final results. Major points suggested by the results are as follows:

1) The time and space resolutions are tightly connected with the process studied (Stommel, 1963) and to the time-space resolution of an instrument. Some events can change more rapidly than others (e.g., a phytoplankton bloom can cause dramatic changes in a day, whereas years may be necessary to see the effects of coastal erosion). Therefore, some disciplines are more dynamic than others, requiring a higher frequency of data set access and ordering. High ordering frequency should not, therefore, be mistaken for popularity, and certainly not for importance or quality, of a product.

2) Given our limited sample size, the EOS-funded scientist did not present statistical differences in the way data was used or pulled. The percentage distribution by level is similar in shape and frequency to the science community at-large. These results indicate that EOS-funded scientists operate and use data in a fashion similar to the general scientific community.

3) More than 70% of the potential users prefer products with moderate levels of processing (2, 3 and 2&3), which indicates a high confidence in the processing procedures and teams.

4) About 49% of the potential users are interested in areas smaller than 10,000 km^2 and 25% are interested in areas smaller than 1,000 km^2. Subsetting seems to be required if volume of deliveries are to be reduced.

5) About 58% of potential users will locate and visually inspect data. This indicates that a large fraction of users' time will be spent on inspecting data (presumably through browsing) in order to test hypotheses, acquire new ideas and compare data.

4.2 Applications of Survey Results

The results of the survey are used in several ways. Often, the ECS developers approach the User Characterization Team with specific questions about the user community for which they need answers. Collectively, these questions are referred to as "developer questions"; each question , the analysis performed, and the "answer" to the question are documented and maintained by the User Characterization Team. The Team also maintains an "Information Catalog" that contains an abstract of each question asked by developers (Miller, 1995). The information catalog is distributed to developers periodically. Developers can peruse the catalog to see if their current question has been asked by others; in this case, the answer is already available. If the question is a new one, a developer can submit a request to the User Characterization Team for the information.

Table 4.1 lists the developer questions where the results of the Product Use Survey were used in the analysis. Also included in the table is the area of the design that is affected by each question.

Table 4.1. Design Applications of Survey Results

Subject of Question	Design Component Impacted
Frequency and Distribution of Order Requests	Data distribution segment
Subsetting of subintervals of data	Data server processing requirements
User pull on data pyramid levels	Data Model - demands placed on data objects
Relative demand for data products	Data server sizing; communications links (see section 4.2.1)
Volume of InterDAAC traffic	DAAC-to-DAAC network sizing

Section 4.2.1 describes, in detail, how the survey results were used to answer a particular developer question. This example is included in order to explicitly show how the survey results are used in an analysis and how the results directly affect the design of an ECS system component. Other analyses proceeded in a similar fashion and the interested reader is referred to the Information Catalog (Miller, 1995) for more details on additional analyses performed by the User Characterization Team.

4.2.1 Impact of the Survey Results on the Data Server Design

The results of the EOSDIS Product Use Survey affect the ECS design in many areas. The primary use of the survey data is to estimate the relative pull on the individual data products. This knowledge, when combined with the 27 science user scenarios and demographics estimates, provides the ECS developers with parameters necessary to determine the size of individual data servers. This process is explained below.

Although the user scenarios were collected from individual users, an estimate was made of the fraction of the total user community that would use the EOSDIS in a manner similar to each scenario. Multiplying these proportions by the total number of expected users provides a number of users for each scenario, N(i) where *i* is the *i*th scenario and N is the number of users. Within each scenario, one can obtain a total for the number of times per year the user accesses data, regardless of the type of data. Let this number be AF(i) where AF is the access frequency per year. The final parameter is the relative product access frequency for Product X, RPAF(X), which comes directly from the survey.

These three parameters can be used to estimate the number of times per year that Product X is accessed, NA(X), as NA(X) = N(i) x AF(i) x RPAF(X). Since the ECS developers have mapped each data product to a specific data server, the number of accesses to a data server will be the sum of all of the accesses to all of the products associated with that server.

The number of accesses to a data server is a parameter that affects several system components. First, it will determine the size of the data server DBMS cache because a larger number of accesses requires a larger cache. Second, if a data server has a large number of accesses, this will generally cause more traffic on the I/O lines between the data server and the archive, requiring a larger communications bandwidth at the site in question. In addition, if a group of data servers at a particular site is accessed more frequently than a group at another site, the data staging area must be larger at the site with more accesses. Finally, a site with a larger number of accesses to its group of data servers requires a larger external network bandwidth than sites with fewer accesses.

This page intentionally left blank.

5. References

5.1 References

Miller J. (1995). User Characterization Information Catalog, Internal manuscript.

Shneiderman B. (1992). Design the User Interface: Strategies for Effective Human-Computer Interaction, Addison-Wesley Publishing Co., Inc. Reading Mass., 1- 573pp.

Stommel H. (1963).Varieties of Oceanographic Experience, Science vol.139, 572-576 pp.

Tyahla L. (1994). ECS User Characterization Methodology and Results, White Paper # 194-00313 TPW, Hughes Applied Information System, Inc., Landover (MD), 1-53 pp.

Wingo T. (1994). User Characterization and Requirements Analysis, White Paper # 194-00312 TPW, Hughes Applied Information System, Inc., Landover (MD), 1-30 pp.

This page intentionally left blank.

Appendix A

Table 1. Useability Exit Survey

1. The Survey was easy to understand.
2. Navigation through the Survey was easy.
3. This On-Line Survey was easier to complete than other paper or e-mail surveys I have participated in.
4. Were you able to complete the survey in the estimated time range.
5. The data product information provided with the survey was adequate.
6. My experience using the Survey was positive.
7. Comments or suggestions

Table 2. EOS Investigators

Investigator	EOS Directory Page	Project Type
Jeff Dozier	223	I.I: Hydrology, Hydrochemical Modeling, & Remote Sensing in Seasonally Snow-Covered Alpine Drainage Basin
Joy Crisp	235	I.I.: A Global Assessment of Active Volcanism, Volcanic Hazard, & Volcanic Inputs to the Atmosphere from EOS
Michael D. King	245 203	I.I.: Interd. Inv. of Clouds & Earth's Radiant Energy System CERES
Richard Eckman	226	I.I.: Observational & Modeling Studies of Radiative, Chemical, & Dynamical Inter. in the Earth's Atmosphere
L. Walter	235	I.I.: A Global Assessment Active Volcanism, Volcanic Hazard, & Volcanic Inputs to the Atmosphere from EOS
Mark Abbott	217 & 211	I.I: Coupled Atm./Ocean Processes & PP in the Southern Ocean; MODIS
Dennis L. Hartmann	228	I.I.: Climate Processes Over the Ocean
Robert H. Haskins	222	I.I.: NCAR Project to Interface Modeling on Global & Regional Scales with EOS Obs
Anne Kahle	235	I.I: A Global Ass. of Active Volcanism, Volcanic Hazards & Volcanic Inputs
Eric J. Fielding	229	I.I: Climate, Erosion, & Tectonics in the Andes & Other Mountain Systems

Table 2. EOS Investigators

Investigator	EOS Directory Page	Project Type
Warren Wiscombe	231	I.I: Global Hydrologic Processes & Climate
Charles R. Bentley	205	GLAS
Thomas A. Herring	205	GLAS
Bob Wells	206	HIRDLS
William Mankin	206	HIRDLS
Paul Bailey	206	HIRDLS
W.L. Barnes	211	MODIS
Hugh H. Kieffer	202	ASTER
Larry D. Travis	204	EOSP
Peter Cornillon	215	SeaWINDS
TOTAL 3 PI and17 co-Is		TOTAL 7 Instrument Teams and 9 I.I

Table 3. Global Fellowships

Global Change Fellowships	EOS Directory Page and Advisor name	University
Greg Tucker	291, Advisor Slingerland	Pennsylvania State Univ.
John Albertson	290, Advisor Parlange	Univ. of California Davis
Scott Greene	292, Advisor Willmott	Univ. of Delaware
David Early	289, Advisor Long	Brigham Young Univ.
Jeff McCollum	289, Advisor Krajewski	Univ. of Iowa
Drew Pilant	291, Advisor Rose	Michigan Tech. Univ.
Paul R. Houser	292, Advisor Sorooshian	Univ. of Arizona
Steve Reising	288, Advisor Inan	Stanford University

Appendix B

Table 1. Universities

Aston University	Boston University
Brigham Young University	Brown University, Dept. Geological Sciences
California State University, Monterey Bay	Caltech, California
Center for Space Research UT- Austin	Colorado State University
Columbia University, NY	Cornell Center for the Environment
Cornell Theory Center	National Ctr for Geo. Info & Analysis-SUNY Buffalo
Univ. of Wisconsin, Madison	Florida State University
George Mason University	Harvard University
Iowa Institute of Hydraulic Research	Iowa State University
Irvine, University of California	Johns Hopkins University
Kansas State university	Lamont Doherty Earth Observatory, Columbia Univ., NY
LSU Southern Regional Climate Center	Michigan Technological University
Millersville University PA	MIT Center for Space Research
MIT Earth Resources Lab	MIT Lincoln Laboratory
Dartmouth College	New Mexico Institute of Mining and Technology
NUTIS, University of Reading	Oregon State University
Penn State University	Princeton University
Purdue University	Scripps Institution of Oceanography, UCSD
St. Cloud State University	St. Olaf College
Stanford University	Texas A&M University, College Station
The University of Kansas	The University of Rhode Island
UBC Oceanography	UC Berkeley
University at Albany, ASRC, SUNY	University California Santa Barbara
University of Alabama in Huntsville	University of Arizona
University of California San Diego	University of Chicago
University of Colorado	Dept. Geography, Birkbeck College, Univ. of London
University of Maryland	University of Miami, FL
University of San Francisco, Marine Science	University of South Florida

University of Wisconsin-Milwaukee	Washington University
Yale University, Geology and Geophysics	Antartic CRC, University of Tasmania
Copenhagen University Observatory	Dalhouise University
Dept. of Surveying, Univ. of Cape Town	Dept. de Ciencias de la Atmosfera, Universitad de Buenos Aires
University of Hawaii	Dept. Meteorology, University of Edinburgh
Dept. Oceanography, Univ. of Southampton	Dept. Physical Sciences, Embry-Riddle Aeronautical University
Dept. of Survey & Mapping, Univ. of Natal	Dipartimento Scienze, University D'Annunzio, Italy
EC's Joint Research Centre-Italy	Fraunhofer Institute for Atmospheric Env. Research
Free University of Berlin	Ice Service Environment of Canada
Istanbul Technical University	Instituto del Mar del Peru, IMARPE
Max-Plank-Institut fuer Meteorologie	Nagaoka University of Technology
Old Dominion Univ.	School of Marine Science & Tech, Tokai Univ.
Politecnico di Milano, Univ. Milan, Italy	Royal Netherlands Meteorological Institute
Oxford University	Tech Univ. Munich, Inst. Astro. & Phys. Geodesy
Universities Space Research Association	University of Dundee
University of Hamburg	University of Sao Paulo/Dept. Oceanography

Table 2. Federal and Government Lab.

Alaska SAR Facility	ARA/LMD/CNRS
ARL/SORD	Bedford Inst. of Oceanography
Bermuda Biological Station for Research	Brookhaven National Laboratory
Bureau of Meteorology	Chesapeake Biological Lab.
CIRES/NOAA	CONAE
CSC/Langley Research Center	CSIRO, Division of Applied Physics
EOSDIS V0 IMS	GAULT Geological Services
GFDL/NOAA	JHU Applied Physics Lab
JPL	Los Alamos National Lab.
LUW Air Quality Dept.	MAPS Development, NOAA Forecast Sys. Lab
Marine Research Institute	NASA GSFC
NASA Langley Research Center	NASA MSFC
NASA/MSF/HSTX/GCIP	National Weather Service
Naval Postgraduate School	Naval Research Laboratory (NRL)

NCAR	Swiss Federal Institute for Forest, Snow and Landscape Research
NOAA Climate Diagnostic Center	NOAA Pacific Marine Environmental Lab
NOAA/ Sanctuaries and Reserves Division	NOAA/Colorado
NOAA/ERl/ETL	NOAA/NESDIS
NOAA/NMFS/SWFSC	NSF
Oklahoma Climate Survey	Pacific Northwest Laboratory
PMEL/NOAA	Rutherford Appleton Laboratory
US DOC/NOAA/NOS/ORCA/Sea Division	US EPA
US EPA Region VII	US. Geological Survey
USDI-USGS-WRD	British Geological Survey
Canadian Atmospheric Environment Service	Canadian Forest Service
Cemagref-Engref Remote Sensing	Fraunhofer Inst. for Atmosph. Envi. Research
Geological Survey of Canada	Hungarian Meteorological service
James Rennell Centre for Ocean Circulation	Max-Plack Institut fur Meteorologie
Mote Marine Lab.	Natural Env. Research Council, UK
World Meteorological Organization	NMC/NOAA

Table 3. Commercial Corporations

Alerta Ltd.	Batelle PNL
Centre for Tropical Pest Management	GATS Inc.
General Science Corporation	IBM Thomas J. Watson Research Center
IMFUA,RUC	Institute of Computer Science
Lockeed Engineering & Science Corporation	Research and Data System Corporation
RSDAS	SAIC
The MITRE Corp.	

This page intentionally left blank.

Abbreviations and Acronyms

DAAC	Distributed Active Archive Center
ECS	EOSDIS Core System
EOS	Earth Observing System
EOSDIS	EOS Data Information System
NASA	National Aeronautics and Space Administration
RPAF	Relative Product Access Frequency
URL	Universal Resource Locator
WWW	World Wide Web