DESIGNATING USER COMMUNITIES FOR SCIENTIFIC DATA: CHALLENGES AND SOLUTIONS

M. A. Parsons[1*] and R. Duerr[2]
National Snow and Ice Data Center/World Data Center for Glaciology, University of Colorado, Boulder, Colorado, USA
*Email: *[1]parsonsm@nsidc.org [2]rduerr@nsidc.org*

ABSTRACT

Defining a "designated user community" for a data collection is essential to good scientific data stewardship. It enables data managers to determine what information is necessary to ensure the usability of the data now and into the future. It helps managers present and enable access to the data and may determine the format of the data. However, defining a community is difficult, and it is impossible to predict how the use of a data collection may change over time. This creates a series of data management problems for data stewards that may be mitigated by a set of best practices.

Keywords: Designated User Community, Users, Uncertainty, Documentation, Long-term preservation, OAIS

1 INTRODUCTION

Scientific data preservation is pointless unless the data are used now and in the future. To ensure data usability, data managers need to understand *who* needs to use the data. The Consultative Committee for Space Data Systems (CCSDS), in its ISO standard Open Archival Information System (OAIS) reference model, defines an archive as an "organization that intends to preserve information for access and use by a Designated Community" (CCSDS, 2002, p. 1-8). In principle, the data need to be "independently understandable" by the designated community. This principle can guide scientific data managers through decisions on data documentation, formats, and presentation, *if* they have a solid understanding of the designated community.

Scientists and archivists have long targeted their efforts toward specific communities. Journals are written for primary audiences. Libraries and archives target their content to specific user communities. Yet scientific data stewardship is a relatively new discipline that faces unique challenges in defining and meeting the needs of its communities. It can be difficult to define and understand a designated community for a scientific data collection because rapidly evolving technology and scientific understanding often lead to unanticipated users and applications for the data (Hunolt, 1999).

The National Snow and Ice Data Center (NSIDC) has been archiving and managing scientific data for more than 25 years. Although we have generally targeted our data to a relatively narrow cryospheric science community, we have seen a significant increase in the use of our data by an ever-broadening user base. In this paper, we use our experience to describe a set of principles and practices to help scientific data managers better define their user communities and improve data sets and related information to ensure long-term usability by both defined and potentially unanticipated users.

2 DEFINING A DESIGNATED COMMUNITY

Often the initial user community for a data collection is self-evident. It is the participants in a particular experiment, the members of the science team for a remote sensing instrument, or a specific research group designated by a funding agency or scientific organization. Indeed, it is common for data access to be restricted to these narrow communities for a time. For example, NSIDC restricted access to data collected during the 2002/2003 Cold Land Processes Experiment (CLPX) in Colorado (NASA Cold Land Processes Mission , n.d.; NSIDC, n.d.) for one year to investigators who directly participated in the experiment. Similarly, data collected from a typical NASA Earth Observing System (EOS) satellite sensor are initially only available to the formally designated science team for six months to a year while instrument calibration and algorithm validation is ongoing.

Once data are made more broadly available, it may seem reasonable to assume that the user community will be similar to the initial users. They will be educated in a closely related scientific discipline; they will be investigating similar phenomena; they will have a similar knowledge base. In short, they will understand. But, of course, this is a dangerous assumption. At a simple level, the new broader community may not understand the jargon particular to a certain experiment or mission. They will not be familiar with the assumptions, imperatives, and compromises that led to a particular data collection approach. They may not be as similar as originally thought. They will need more information, a greater context, to use the data effectively and accurately.

Returning to the example of the Cold Land Processes Experiment, the primary focus of the experiment was to develop adequate understanding of snow pack characteristics over space and time in order to develop a satellite remote sensing method that could accurately measure the water content of snow (Cline, Elder, Davis, Hardy, Liston, Imel, et al. 2003). As such, one might assume that users of these data would be reasonably versed in the engineering and physics of microwave remote sensing. A review of our user registration logs suggests otherwise. Stated uses for the data included a variety of land surface modeling needs, atmospheric modeling, analysis of snow physics, permafrost studies, and, of course, remote sensing of snow. Moreover, most of the data have only been publicly available for a short time, so the user base is likely to continue to expand to include regional hydrologists, land and watershed managers, and others (a recent registrant plans to use the data to improve highway snow removal). So even though this started as a specialized and targeted data collection, making it publicly available made it necessary to recognize and support a broadly designated community comprised of many different user types.

In a more far-reaching example, the Special Sensor Microwave/Imager (SSM/I) has flown on U.S. Department of Defense polar orbiting satellites since 1987. The original purpose of the sensor was to support operational weather forecasting for the U.S. Air Force and Navy, but the data are used today for a broad array of global land, ocean, and atmospheric monitoring applications. Analysis of data from SSM/I and its predecessor, the Scanning Multi-channel Microwave Radiometer, has led to the creation of one of the longest satellite remote sensing time series of sea ice concentration and northern hemisphere snow cover, invaluable to global climate change studies. These applications are well beyond the scope envisioned by the original designers of the sensor. Other unanticipated uses continue today. For example, biologists have recently requested SSM/I-derived snow water equivalent data to correlate with caribou calving dates.

The OAIS model emphasizes the need to consider a broad designated community early in the archiving process (CCSDS, 2002), but as the previous examples illustrate, it is unrealistic to think one can anticipate all uses of a scientific data collection. On the other hand, it is impractical to define an overly broad designated community such as the “general public,” because it can be easily ill defined or even lead to inappropriate use of data (see Section 4). Therefore, a central operating principle is to designate a broad community but specify the community in as much detail as possible. Clearly, those two imperatives—broad but specific—are at odds, and balancing these imperatives presents a core challenge for data managers that is not explicitly recognized in the OAIS Reference Model requirement that data be independently understandable by a designated community (CCSDS, 2002). Nevertheless, data managers can prepare and maintain the data and documentation in a way that facilitates broad but appropriate use. (In this paper “documentation” includes all of what is commonly called metadata. The complete set of data and documentation is the Archive Information Package in OAIS vernacular.)

3 UNDERSTANDING KNOWLEDGE BASES — CONCEPTUAL METAPHORS AND ACCEPTED OPINION

Philosophers have long grappled with the most effective way to convey information, but there is general agreement that you cannot even begin without an understanding of the underlying assumptions of both the information provider and recipient. Lakoff and Johnson (1980) argue that people need a conceptual basis to understand something and that scientists invoke key metaphorical concepts to work observations into a coherent, consistent structure. As an example, consider the word metal. In ordinary conversation, the term metal has a distinct meaning that does not include substances such as oxygen, carbon, or complex compounds such as plastic or sand. Yet to an astronomer studying stellar evolution, metal means something entirely different. It is, in essence, a metaphor for any substance with an atomic weight heavier than hydrogen. This explicitly includes substances such as carbon and oxygen. Any description of an astronomical data set that included information about the “metal” content of stars is therefore likely to be misunderstood by people outside the core astronomical community unless the distinction is made clear.

These metaphors or common references may vary from discipline to discipline and are bound to change over time, even within a given discipline. Therefore, data managers need to be aware of the conceptual basis that led to the collection of a particular data set to effectively document its usability for a future community. Yet they need to be careful not to provide unnecessary detail. This is a central principle of classic rhetoric. Aristotle (1954, 1396a), wrote:

> … we must not carry [our] reasoning too far back, or the length of our argument will cause obscurity: nor must we put in all the steps that lead to our conclusion, or we shall waste words in saying what is manifest. It is this simplicity that makes the uneducated more effective than the educated when addressing popular audiences — makes them, as the poets tell us, "charm the crowd's ears more finely." Educated men lay down broad general principles; uneducated men argue from common knowledge and draw obvious conclusions. We must not, therefore, start from any and every accepted opinion, but only from those we have defined — those accepted by our judges or by those whose authority they recognize.

This does not mean that we should discard information that may seem tangential to the core needs of our designated community. Instead we must challenge our assumptions regarding the knowledge of our user community and provide suitable context for users to understand the data. The CCSDS (2002, p. 2-4) states "an OAIS must understand the Knowledge Base of its Designated Community to understand the minimum Representation Information that must be maintained." This knowledge base is too narrowly defined by the CCSDS. It should also include Lakoff and Johnson's "conceptual metaphors" and Aristotle's "common knowledge" and "accepted opinion." Furthermore, it is equally important to understand the knowledge base of the data creator and how it coincides with or diverges from the knowledge base of the designated community. Finally, the data documentation must strike an appropriate balance between exposition and pith to ensure that sufficient information is available to understand and use the data but not so much to overburden the user.

This is all rather abstract, but it is useful to have these principles in mind when developing data documentation. At NSIDC, the primary method we use to employ these principles is to have an educated, but non-expert, technical writer develop data set documentation in close consultation with scientists experienced with the type of data being described. We find it useful to collaborate closely with data providers when possible and we explicitly include staff scientists in our data management process. We also are able to take advantage of a formal user working group comprised of scientists and data users who meet every six to twelve months to discuss the presentation and availability of NASA data sets held at NSIDC. Others have emphasized the need for scientist involvement in data management and long-term archiving (Hunolt, 1999; Olsen, Briggs, Porter, Mah, & Stafford, 1999), and we have found this involvement invaluable in developing documentation and supporting unanticipated users.

Recently, NSIDC has begun archiving non-cryospheric data from NASA's Advanced Scanning Microwave Radiometer (AMSR-E) and Geoscience Laser Altimetry System (GLAS) and related validation data sets. We do not have subject matter experts for these data in house, so we have tried to establish formal relationships with members of the science teams to aid user support and to ensure appropriate documentation. So far this approach seems to be working well, in that we have received many questions from users about the data that we could not immediately answer in house, but we have consistently received prompt, courteous, and helpful responses from the AMSR-E and GLAS science teams. It remains unclear whether this is a viable long-term solution, however, because the data have only been broadly available for about a year.

4 CHARACTERIZING UNCERTAINTY

Writing data documentation for an unanticipated audience is a never-ending challenge, even when applying Aristotelian principles. Including certain elements in your document can help address that challenge. Chief among these required elements for scientific data is a frank and detailed explanation of the uncertainties of the data described. Scientific data creates a unique problem for data archival in that there is always some degree uncertainty about the accuracy of the data. The uncertainty may be small with a simple in-situ measurement or relatively large with a satellite remote sensing time series or even unquantifiable with certain historical global measurements, but it is always there, and it must be well described to avoid data misuse. NSIDC data on sea ice concentration derived from passive microwave remote sensing provide a good case study of how to document uncertainty and why it is necessary.

NSIDC archives 38 data sets on sea ice and 12 derived from passive microwave remote sensing. This is a confusing array even for a sea ice specialist. Yet the diversity is largely due to the complexity involved in understanding a geophysical parameter in a region where no one lives, few visit, and it is dark half the year. Passive microwave remote sensing has distinct advantages in this region since it does not require sunlight, it "sees" through clouds, and it measures over a broad area without requiring human visitation. Unfortunately, it has coarse resolution, is susceptible to contamination by weather systems, and has had problems with geolocation. Furthermore, since 1973 instruments and satellites have evolved, making it difficult to construct a consistent time series or climate data record. In other words, there is a large uncertainty associated with these data sets.

NSIDC has sought to address this uncertainty in a variety of ways. First, we try and provide an overall context for the data. We have created a web site comparing and contrasting all the sea ice products we distribute (http://nsidc.org/data/seaice/). We describe the strengths and weaknesses of each product and indicate appropriate applications. The intent is to guide scientific users to the product that most suits their application. Once a user decides on a product, the data documentation includes detail written by both data providers and NSIDC scientists on the known errors and uncertainties in the data. In particular, we describe the different attributes of the two major algorithms used to generate sea ice concentration from passive microwave and changes in sensors over time. Some of this detail has required sophisticated scientific analysis that goes beyond the scope of typical data documentation. So in addition to citing the scientific literature, we have also developed a series of special reports on data related topics. Three of these reports are devoted to deeper analysis of passive-microwave-derived sea ice products (Maslanik, Agnew, Drinkwater, Emery, Fowler, Kwok, et al., 1998; Stroeve, Li & Maslanik, 1997; Stroeve & Smith, 2001). We hope we have adequately described relevant data uncertainty for future users and have organized it well enough to avoid obscurity.

All told, we are providing hundreds of pages of data documentation. Even with the guidance provided by our sea ice web site, it is unlikely we have met the needs of many non-expert users. The necessary information is available, but it is daunting. It is, however, necessarily daunting. With these products it is actually inappropriate to try to accommodate all uses. For example, say a wildlife biologist wanted to compare sea ice concentration with polar bear migration patterns. While these data are excellent for showing trends in sea ice concentration over time and over large areas, they are very inaccurate at the detailed spatial (pixel level) and temporal (daily) scale necessary for this application (Meier, VanWoert & Bertoia. 2001). For these products it is not only difficult, but also scientifically flawed, to define the designated community too broadly.

Instead we have derived specific products geared to appropriate non-expert use of passive microwave remote sensing. Chief among these is our sea ice index (http://nsidc.org/data/seaice_index/), which provides images of average monthly ice conditions and trends and anomalies that compare recent conditions with the long-term averages (Figure 1). We specifically note the uses and limitations of the images. The idea is to provide an appropriate, high-level view of sea ice conditions and trends for people who are not experts in remote sensing of sea ice, be they other scientists or (to a lesser degree) the general public. For the general public, we also provide more general pages describing the role of sea ice in the global climate system (http://nsidc.org/cryosphere).

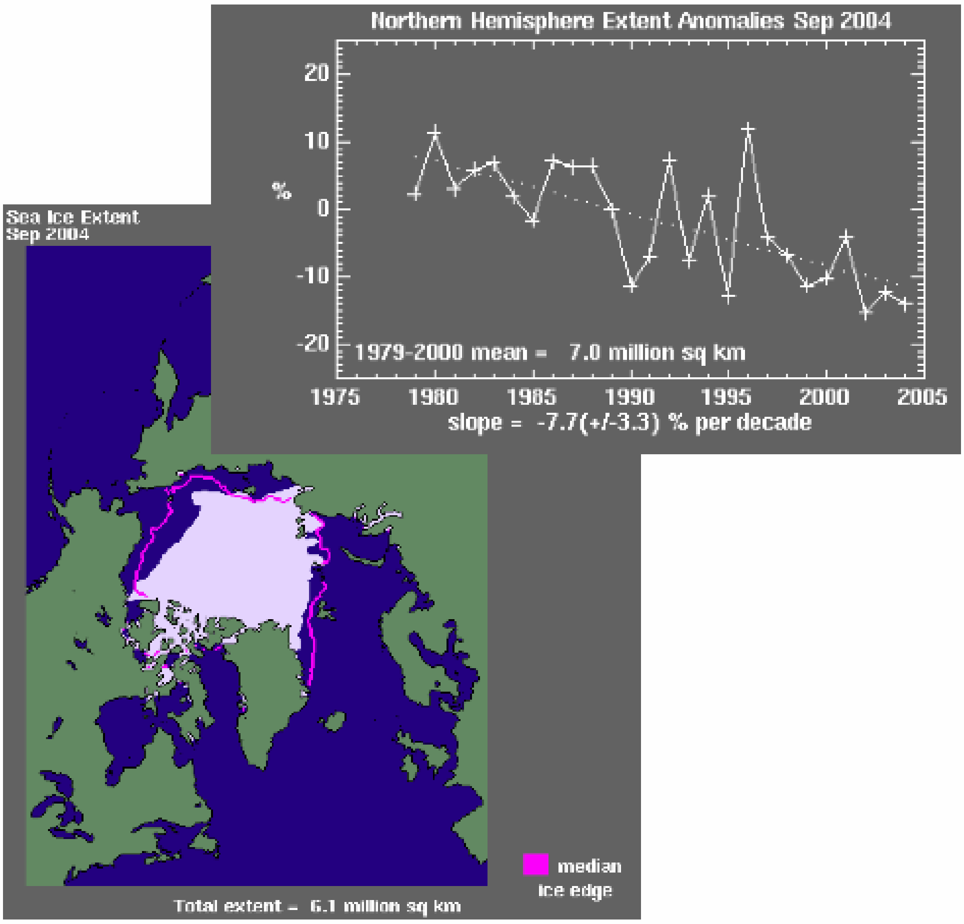

Figure 1. Examples of figures from NSIDC's Sea Ice Index showing Sept. 2004 ice extent compared to the median September extent for 1979-2000 (left) and the trend of anomalies in ice extent from 1978 to 2004 relative to the 1979-2000 mean (right) (Fetterer & Knowles, 2002).

Passive microwave remote sensing of sea ice is a complex business resulting in diverse and complex products. NSIDC's experience in presenting these products illustrates the need for scientific data managers to carefully document data uncertainty to facilitate broad and lasting data usability, but it also illustrates the importance of not defining too broad a designated community.

This passive microwave sea ice example reinforces the value of involving scientists in data management, yet we have also found it important to involve data managers directly in scientific experiment design and data collection. The Cold Land Processes Experiment included data management considerations in the design of the experiment, and data specialists participated in the actual field data collection. We found that this led to a more complete and consistent data set and potentially reduced the number of ambiguous data values by fifteen to twenty percent (Parsons, Brodzik & Rutter, 2004). In other words, by being directly involved in the experiment we reduced the uncertainty of the data collected. We also gained a much better understanding of the history and limitations the data and are, therefore, better able to describe them and ensure appropriate use (rather than misuse) in the future.

Finally, and perhaps most importantly, NSIDC staffs a dedicated User Services Office, whose primary role is to research and respond to inquiries about the data sets and their uses. No amount of documentation, no matter how cleverly presented, will be able to answer all user questions, especially from unanticipated users. It is, therefore, important to explicitly include people in the data management process. Our user services staff enhance our overall data management and help ensure continual and broad data access. Furthermore, through direct interaction with users, our staff gain additional knowledge about what kinds of data researchers are seeking and how they use the data and documentation. If we consider the sea ice example, where there are many parallel data sets with varied applications in marine operations, climate science, oceanography, biology, and other disciplines, the User Services Office can help potential data users choose the most appropriate data set, and at the same advise data teams on how to create documentation that helps researchers use the data effectively and accurately. This active human element of data management is not always recognized by funding agencies, nor is it explicit in the OAIS Reference Model, yet our user services office may be the most widely appreciated aspect of NSIDC's operation.

5 DETERMINING DATA FORMATS AND METAFORMATS

So far, we have discussed documentation as a tool to enable broad and unanticipated use of scientific data. There are also attributes of the data itself that should be considered. Central among these is data format. Determining an appropriate data format for data storage and distribution is one of the more challenging problems of data management. Raymond (2004) argues that there are four important themes in designing file formats: interoperability, transparency, extensibility, and storage or transaction economy. He argues that the general file format that best addresses these themes is text, and that the only justification for a binary protocol is with very large data sets. Raymond is arguing from a computer programming perspective, but he takes a broad perspective and probably represents the generic, unanticipated user very well. His argument for text is compelling from a long-term archival perspective, especially in terms of its transparency, i.e. human readability. Nevertheless, some scientific data, notably high-volume remote sensing images, require a binary format for effective and efficient transfer and manipulation. When you do need to use a binary format, it is especially important to consider options to aid transparency, such as sample textual representations for some of the data.

The matter is further complicated when we begin to consider the specific implementation of a file format, be it binary or textual. These specific implementations are what most generally consider the data format. Raymond (2004) calls them metaformats. Examples for text include delimiter-separated values and XML. Examples for binary include the hierarchical data format (HDF), GeoTIFF, and a variety of binary arrays. At a greater level of specificity is what the SEEDS Formulation Team (2003) calls the format profile. This is a specific implementation of a metaformat and would include machine-specific considerations such as byte order and 32 vs. 64-bit words. We will not discuss format profiles in great detail, but note there are few standards and they are subject to great variability and evolution in user requirements (SEEDS Formulation Team, 2003). With a broad user base, the ideal may be to distribute data in several formats and profiles generated on request and to preserve the data in yet another preservation-friendly format (Duerr, Parsons, Marquis, Dichtl & Mullins, 2004).

The NASA Earth Observing System (EOS) experience illustrates the format issues. Very early in the program, NASA designated HDF-EOS as the standard format for all NASA EOS data. HDF-EOS was used both as the distribution and archive format. However, in response to the growth of the Geographic Information Systems (GIS), NASA plans to convert subsets of the data to GeoTIFF. In addition, NASA will provide the capability to reformat data on demand. So although these data are only a few years old, substantial effort has been invested to provide them in a format not originally anticipated as necessary.

Profiles can be even more complex. One profile worthy of special consideration in the Earth sciences is the map projection and associated gridding scheme. Choosing an appropriate gridding and projection scheme is a perennial problem for data managers, much like choosing a data format. In contrast to data formats, though, it is often inappropriate to simply regrid data on demand, because the requested interpolation technique may not be appropriate for the data and spatial resolutions of source and target grids. There is no single "best" projection for all applications; correspondingly there is no single gridding scheme appropriate to all purposes (Knowles, 1993). We believe that one solution is to clearly define the necessary grid or grids in the experiment design and sample accordingly, while at the same time provide the swath or other lower-level data for users who may require it. This approach can, however, increase storage requirements by orders of magnitude. More sophisticated solutions such as storing remote sensing data in a geodatabase using spherical (or geoidal) geometry should also be considered.

6 SUMMARY OF PRINCIPLES AND BEST PRACTICES

Overall, our main point is that scientific data managers should maximize the value of their data by accommodating the broadest possible use including unanticipated use of the data while discouraging data misuse. In other words, they should define a broad, but not too broad, designated user community. This seemingly contradictory statement presents a core challenge for data managers. When addressing this challenge data managers need to keep the following core principles in mind:

- When developing data documentation challenge your assumptions about the knowledge of your data users and creators, and write in language and detail appropriate to your user community.

- Understand and describe data uncertainty.

- Keep data formats, metaformats, and profiles simple and flexible while addressing issues of transparency, interoperability, extensibility, and storage volume.

These principles, in turn, lead to some specific data management practices that need to be recognized by both data managers and funding agencies:

- Have educated but non-expert writers develop data documentation in consultation with scientists familiar with the data and their application.
- Engage scientists directly in the data management process, either through the use of in-house staff or through formal relationships with external experts.
- Involve data managers in the collection of scientific data to facilitate their understanding of the data and their uncertainties.
- Create formal user panels to advise on the presentation, documentation, and appropriate application of the data while recognizing that no one group can represent all interests.
- Support an active and knowledgeable user services staff to answer user queries and provide feedback on the use and understanding of the data to data managers.
- Explicitly describe uncertainties in the data.
- Provide an overall context for understanding the data and present the data in a manner appropriate to the audience. This may often involve multiple presentations of the same data to meet the needs of different audiences. Professional writers, graphic artists, and interface designers are useful in this area.
- Store data in a simple textual (i.e. human readable) data format when possible. If a binary data format is necessary, include example textual representations of the data.
- Develop data formats in response to specific user needs and be prepared to reformat data as necessary. It is often necessary to archive data in one format while distributing them in several other formats. Reformatting is an ongoing need.
- For gridded data, define the grids and projections as early as possible to facilitate intercomparison. Specifically define the interpolation methods used in any gridding schemes and preserve the ungridded data in addition to the gridded data.

7 CONCLUSION

Good scientific data stewardship requires explicit recognition and understanding of data user communities. It is often necessary to define a broad user community and still recognize that there will be unanticipated uses of a given scientific data collection. At the same time, one should recognize the risk of scientifically inappropriate use of data and be sure to carefully document data uncertainty. It may also be necessary to devise creative methods of data presentation to ensure broad but appropriate use. This requires data stewards to challenge their basic assumptions about what is understood about their data both by data users and providers, to ensure comprehensive documentation, and to provide data in transparent, interoperable formats. Ultimately, these fundamental long-term archive considerations may be summarized with the basic principle — “keep things simple but flexible.”

Industrial designers, software engineers, and even traditional archives and libraries rely heavily on studies of the use of their systems. The results of some of these studies may help guide data managers in the development of data access systems, but they provide little if any guidance on the necessary attributes of the actual data and supporting documentation. The value of scientific data increases with use. Therefore, data stewards need to better understand who will use their data and what they can do to facilitate that use. This is vital research area for the long-term preservation of scientific data.

8 REFERENCES

Aristotle. (1954) *Rhetoric*. Translated by W. Rhys Roberts. New York: Modern Library.

CCSDS (Consultative Committee for Space Data Systems). (2002) *Reference Model for an Open Archival Information System (OAIS) CCDSD 650.0-B-1. Blue Book. Issue 1*. Washington, DC: CCSDS Secretariat. [Equivalent to ISO 14721:2002].

Cline D., Elder, K., Davis, R.E., Hardy, J., Liston, G.E., Imel, D., Yueh, S.H., Gasiewski, A.J., Koh, G., Armstrong, R.L., & Parsons, M.A. (2003) Overview of the NASA Cold Land Processes Field Experiment (CLPX-2002). *Proc. SPIE Int. Soc. Opt. Eng. Volume 4894, Microwave remote Sensing of the Atmosphere & Ennvironment III (*pp.361-372). Hangzhou, China.

Duerr R., Parsons, M.A., Marquis, M., Dichtl, R. & Mullins, T. (2004) Challenges in long-term data stewardship. *Proc. 21st IEEE Conference on Mass Storage Systems and Technologies.* NASA/CP-2004-212750 (pp.47-670). College Park, MD, USA.

Fetterer, F. & Knowles, K. (2002, updated 2004) *Sea Ice Index*. Boulder, CO: National Snow and Ice Data Center. Digital media. Retrieved 13 April 2005 from the National Snow and Ice Data Center Web site: http://nsidc.org/data/seaice_index/.

Hunolt, G., (compiler) (1999) *Global Change Science Requirements for Long-Term Archiving. Report of the Workshop, Oct 28-30, 1998*. Washington, DC: USGCRP Program Office.

Knowles, K.W. (1993) Points, pixels, grids, and cells – a mapping and gridding primer. Retrieved 2 September 2004 from the Cooperative Institute of Research in the Environmental Sciences Web site: http://cires.colorado.edu/~knowlesk/ppgc.html.

Lakoff, G. & Johnson, M. (1980) *Metaphors We Live By*. Chicago: University of Chicago Press.

Maslanik, J., Agnew, T., Drinkwater, M., Emery, W., Fowler, C., Kwok, R. & Liu, A. (1998) *Summary of Ice-Motion Mapping Using Passive Microwave Data.* NSIDC Special Report 8. Boulder, CO: National Snow and Ice Data Center. Retrieved 22 October 2004 from the National Snow and Ice Data Center Web site: Http://nsidc.org/pubs/special/nsidc_special_report_8.pdf.

Meier, W.N., VanWoert, M.L. & Bertoia, C. (2001) Evaluation of operational SSM/I algorithms. *Annals of Glaciology 33,* 102-108.

NASA Cold Land Processes Mission (n.d.) Homepage of NASA Cold Land Processes Mission. Available from; http://www.nohrsc.nws.gov/~cline/clp.html

NSIDC (n.d.) Home page of Cold Land Process Experiment. Available from: http://nsidc.org/data/clpx for

Olson, R.J., Briggs, J.M., Porter, J.H., Mah, G.R. & Stafford, S.G. (1999) Managing data from multiple disciplines, scales, and sites to support synthesis and modeling. *Remote Sensing of Environment. 70***,** 99-107. DOI: 10.1016/S0034-4257(99)00060-7

Parsons, M.A., Brodzik, M.J. & Rutter, N.J. (2004) Data management for the Cold Land Processes Experiment: improving hydrological science. *Hydrological Processes. 18,* 3637 - 3653. DOI: 10.1002/hyp.5801.

Raymond, E. S. (2004) *The Art of UNIX Programming.* Boston, MA: Addison-Wesley.

SEEDS Formulation Team. (2003) Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS) Formulation Team Final Recommendations Report. Retrieved 15 January 2004 from the World Wide Web: Http://lennier.gsfc.nasa.gov/seeds/FinRec.htm.

Stroeve, J., Li, X. & Maslanik, J. (1997) *An Intercomparison of DMSP F11- and F13-derived Sea Ice Products.* NSIDC Special Report 5. Boulder, CO: National Snow and Ice Data Center. Retrieved 22 October 2004 from the from the National Snow and Ice Data Center Web site: Http://nsidc.org/pubs/special/5/index.html.

Stroeve, J. & Smith, J. (2001) *Comparison of Near Real-Time DMSP SSM/I Daily Polar Gridded Products and SSM/I Polar Gridded Products.* NSIDC Special Report 10. Boulder, CO: National Snow and Ice Data Center. Retrieved 22 October 2004 from the from the National Snow and Ice Data Center Web site: Http://nsidc.org/pubs/special/nsidc_special_report_10.pdf.