

Scientific Knowledge Mobilization: Co-Evolution of Data Products and Designated Communities

Karen S. Baker

Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Ruth E. Duerr

National Snow and Ice Data Centre
University of Colorado at Boulder

Mark A. Parsons

Research Data Alliance
Rensselaer Polytechnic Institute

Abstract

Digital data are accumulating rapidly, yet issues relating to data production remain unexamined. Data sharing efforts in particular are nascent, disunited and incomplete. We investigate the development of data products tailored for diverse communities with differing knowledge bases. We explore not the technical aspects of how, why, or where data are made available, but rather the socio-scientific aspects influencing what data products are created and made available for use. These products differ from compact data summaries often published in journals. We report on development by a national data center of two data collections describing the changing polar environment. One collection characterizes sea ice products derived from satellite remote sensing data and development unfolds over three decades. The second collection characterizes the Greenland Ice Sheet melt where development of an initial collection of data products over a period of several months was informed by insights gained from earlier experience. In documenting the generation of these two collections, a data product development cycle supported by a data product team is identified as key to mobilizing scientific knowledge. The collections reveal a co-evolution of data products and designated communities where community interest may be triggered by events such as environmental disturbance and new modes of communication. These examples of data product development in practice illustrate knowledge mobilization in the earth sciences; the collections create a bridge between data producers and a growing number of audiences interested in making evidence-based decisions.

Received 17 February 2015 | *Revision Received* 25 November 2015 | *Accepted* 25 November 2015

Correspondence should be addressed to Karen Baker, Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, IL. US. Email: karensbaker@gmail.com

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/> Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



International Journal of Digital Curation
2015, Vol. 10, Iss. 2, 110–135

Introduction

It is not just metadata that makes research data useful. We find an ongoing data product development cycle supported by a team with multiple perspectives can be a key to mobilizing scientific knowledge. We provide an historical investigation of data product generation spanning three decades that reveals co-evolution of data products and designated communities. The role of a data center, originally conceived as delivering multi-year satellite data to a specialized scientific community, evolves to include tailoring data and findings for multiple audiences resulting in a variety of derived and interpreted data products.

Several decades of professional experience with data systems and data repositories are providing insight into the complexities of data (e.g. Borgman, 2015). In addition, recognition of responsibilities relating to the production and sharing of digital data is growing, prompted by agendas of change (e.g. ICS, 2004; NSDA, 2014; Berman, 2014; NRC, 2012a; NSF, 2013; Genova et al., 2014) and funding agency mandates (OSTP, 2013; NSF, 2011).

We investigate the development of a collection of data products at the National Snow and Ice Data Center (NSIDC), a data center with the technical capacity to handle large volumes of observational data. Two cases of data product development in the earth sciences illustrate an expanding scope of data services. By creating data products with scientific community engagement and review, the center facilitates data sharing and provides access to scientific research by broader audiences. Despite advances in the description of data objects via metadata, it remains a challenge to ensure that data are not only discoverable but also understandable and useful. The NSIDC cases illustrate an approach capable of conveying scientific knowledge about complex environmental phenomena to diverse communities.

Background

The conduct of science by individual researchers or small groups in academic institutions contrasts with the role played by a national data center. A center has a well-defined mission to support research efforts whose size, duration, equipment, and technical specificity are beyond the scope of a single research enterprise. National data or science centers are formal organizations that institutionalize the management and creation of research data (Mayernik, 2015). National data centers were originally conceived as serving particular designated user communities. Increasing demand for data support in the 21st century prompted a National Research Council study of world and national data centers (NRC, 2003). The report describes data availability, standards, hardware, software, and metadata management but stops short of discussing the development of data products.

Today, many kinds of data facilities are emergent, though the use of terms such as centers, archives and repositories is muddled in practice and ambiguous in the scholarly literature. For simplicity, we adopt ‘data repository’ as a general term referring to an organization or an organizational unit that supports aggregation, processing, packaging, and purposeful curation of data. Data repositories take many forms that may or may not include archive services with certified standards for managing, preserving, and

describing data for long-term use. Repository functions ultimately depend upon some combination of goals, resources, and stakeholders. Two major issues for repositories that aim to share data are the production of data products and identification of their designated communities.

What are the Data Products?

In this paper, data refers to observational data recorded by those interested in environmental and earth systems in particular. In considering the objectivity of data as evidence, scientists recognize that a multiplicity of factors influence observations and their interpretation, such as instrument characteristics as well as sampling design, processing and analysis biases that delimit the data (Longino, 1990; Leitzell and Meier, 2009). Data are context-laden, subject to observational bias and other unintentional biases introduced by field logistics and cultural norms as well as processing, analysis, modeling, and publication procedures (Longino, 1990; Galison and Stump, 1996; Hilborn and Mangel, 1997).

Available data are referred to by a variety of terms with overlapping meaning, such as data sets (Maggio et al., 2015; Renear et al., 2010), data packages (Penev et al., 2011), data collections (Cragin, 2009), and information packages (CCSDS, 2012). The term ‘data product’ is similar to the term ‘information package’ used in the Open Archival Information System (OAIS) model (CCSDS, 2012) while the term ‘data set’ is used in practice for scientific observations processed and stored in one or more files. Patel (2012) provides a broad definition of ‘data product’ by referring to ‘use’ as inclusive of both primary and secondary information about an observed system: “a data product is a product that facilitates an end goal through the use of data”. The role of metadata is to document a data product’s capture, completeness, calibration, compositing, selection, integration, compaction, and visualization.

In what follows, we are explicit in recognizing the agency involved in producing data products inclusive of processed, derived, and interpreted data. Just as data recordings are representations of selected aspects of the real world, all subsequent forms of the data produced are representations or re-presentations of observations. Data products may be matrices or streams of numbers but they may also take alternative forms, such as static graphs and interactive displays. Further, sets of data products may be grouped into collections (Palmer, 2004). We build from the work of Parsons and Duerr (2005) to describe below the development of a particular kind of collection, a number of related data products from a single data source.

Who are the Designated Communities?

Data sharing often depends upon with whom the data is shared. The concept of sharing data is limited when understood solely technically in terms of defined variables and documented procedures. Data product providers must make informed decisions about levels of detail, language and form of presentation in conveying contextual information about the data and data taking. The OAIS reference model (CCSDS, 2012; ISO 14721:2012) defines a ‘designated community’ as “an identified group of potential consumers who should be able to understand a particular set of information. A designated community may be composed of multiple user communities.

Parsons and Duerr (2005) made clear that to be assured of what CCSDS (2012) refers to as “understanding a particular set of information”, data providers must not only document the data and its transformation but also be able to carry out the difficult tasks

of identifying a designated community and becoming familiar with both its knowledge base as well as its concerns. They emphasize that knowledge bases include conceptual metaphors and accepted opinion. The OAIS model uses the phrase ‘knowledge base’ to refer to shared community knowledge, such as domain-specific understandings that develop within a community. Within traditional academic disciplines, a shared ‘knowledge base’ is developed through exposure to similar educational materials and perspectives (e.g. schools and curricula). Collective experience centered on a geographic site, a platform, an instrument, an event or a program can also contribute to a shared knowledge base. A designated community is characterized as having members who have an established common ground that minimizes miscommunication. Traditionally, designated communities of data users are domain literate and have some familiarity with the scientific context, data generation, or intended data use. However, with the increasing availability of data today, the existence of interested audiences with a variety of scientific backgrounds outside the domain of data collection must be taken into account.

We use ‘audience’ as a term to include all groups and communities having interest in data products. Some audiences are not familiar with the sciences; they have a cultural knowledge base that differs significantly from that of a scientific community. This must be taken into consideration in order for scientific knowledge to be widely conveyed and understood. There are a variety of models for communicating science to a public audience (e.g. Brossard and Lewenstein, 2010), but few that take into account the need for communication both across the sciences as well as with non-scientific audiences. Despite periodic reminders to scientists that communicating with broader audiences is part of their responsibility (e.g. The Royal Society, 1985; NRC, 2014), the typical investment in knowledge delivery is a small fraction of that spent on research. This is not surprising when the primary role of research institutions is seen as discovery rather than communication or knowledge transfer. A recent NRC report (2014) discusses the need for sustainable infrastructure for communication since “advances in sciences have profound implications for the well-being of society and the natural world.” Data products may serve as a communication mechanism, a common ground persisting over time and providing a shared foundation from which to build shared understandings.

The Research Site: NSIDC 1978-2012

The National Snow and Ice Data Center (NSIDC) collects and disseminates snow and ice data in support of polar and cryospheric research.¹ It is a service-oriented center dedicated to managing and delivering data of interest to a variety of community partners and more recently to new audiences. Historically, within its purview is the processing,

¹ Established in 1957 as a World Data Center (WDC) for Glaciology and relocated to the University of Colorado in 1976 with NOAA sponsorship, the center was designated as the NSIDC in 1982 (NRC, 1999). Such a national data center is administratively distinct from federally funded research and development centers (FFRDC). Data centers are long-term facilities that may serve as repositories for data and data products. As such, they are ‘centers of calculation’ for particular varieties of data (Latour, 1999). The NSIDC is configured to encompass multiple types of facilities, services, and activities including: the World Data Service for Glaciology - University of Colorado; part of NOAA’s National Geophysical Data Center; management and support of polar data for NSF (Antarctic Glaciological Data Center (AGDC) and Advanced Cooperative Arctic Data and Information Service (ACADIS)); a NASA Distributed Active Archive Center (DAAC) working closely with research teams on NASA Earth Observing System (EOS) satellite and field data for frozen regions; and leadership and participation in academic, national and international efforts. As a member of the University of Colorado Cooperative Institute for Research in Environmental Sciences (CIRES), NSIDC is hosted at University of Colorado in Boulder, Colorado.

archiving, maintaining and disseminating of sea ice data, as well as support for selected research communities interested in data on snow cover, ice (freshwater, sea, and ground), and glaciers (NRC, 1999). From early on, work with data included supporting observational studies, aggregating data, developing data sets, networking with other facilities, and interacting with research partners. The center's mission states:

‘NSIDC will make fundamental contributions to cryospheric science and will excel in managing data and disseminating information in order to advance understanding of the Earth system.’ (NSIDC Mission Statement)

Recently, the number of communities interested in understanding the earth system has increased rapidly. The business culture of the center specifically emphasizes community engagement and rapid response to queries. In addition, there is formal recognition of a core responsibility to interact with data generators and scientific data users in order to identify needs and products. Close and active relationships with those familiar with generation of primary data ensure data integrity. An NRC (1999) review of the Distributed Active Archive Center (DAAC) reported “The NSIDC DAAC provides an outstanding example of how good data management practices and a close relationship with researchers can help lead to scientific advances.” An overview of their interactions with scientific communities was noted:

‘The NSIDC-WDC-DAAC complex has a long and impressive history of responding to the needs of snow and ice researchers. Active involvement on the part of technical personnel in the acquisition and development of data products, and the close juxtaposition of the external support function with active faculty and student in-house research, have resulted in an understanding of the *modus operandi* of scientific research on the part of the technical staff and in a proactive attitude. The panel notes that this cooperative and proactive attitude is a strong positive attribute and that the in-house scientific competence adds value to the data sets.’

As a data repository with the capacity to transform data, NSIDC brings together a diverse set of expertise based on deep knowledge of data processing procedures and product development.

Data Product Teams

A data product team provides one example of a contemporary approach to open science communication and data product development. The team is an organizational arrangement at NSIDC that includes research scientists, data scientists, and other specialists including programmers, analysts, documentation writers, and user services staff skilled as communicators. The initial vision for the teams was developed in an all-staff ‘data jam’. Dialog within a team creates communication paths across different understandings of the meaning and usefulness of a data product, thereby ensuring human mediation as one component of a knowledge infrastructure. While staff address questions sent to the User Services Office within 48 hours, data product teams establish longer-term communications that bring together a local team and data requesters from outside the center. As requests work their way across the full extent of the team, they inform design and development of potential future products. The presence of researchers on the team further enables dynamic feedback and a direct connection to

scientists and their information needs. This approach contrasts with that of data preparation carried out by an ‘editorial team’ focused on submission to a repository (Kansa et al., 2014) where less priority may be given to checking on community interests or considering design of new data products. Models that begin with reference to ‘data ingestion’ without mention of pre-ingestion activities typically fail to incorporate the conceptualization of new data products.

Data product teams at NSIDC generally consist of four to six individuals who may be members of several data product teams. Collaboration often extends beyond the center to engage data providers, algorithm developers, instrument engineers, and science teams. Having a multiplicity of perspectives renders data product teams sensitive to potential ramifications of processing options and product design decisions. The team makes possible the tailoring of data products that bridge the data needs of various communities with the realities of the data. Product teams, formed over the past decades to support collection and dissemination of sea ice data, have crafted a community-oriented approach that addresses the data product needs of sea ice researchers as well as a wider set of research, education, policy, and planning communities. Interestingly, over the years since the establishment of data product teams, NSIDC has moved away from scientist led product teams to teams led by data managers, primarily to broaden the team’s perspective on data beyond initial use to include future reuse.

Data product teams provide a human connection tying together data, the scientific community, and diverse audiences. The teams illustrate a configuration that supports the view of Edwards et al. (2013) that, “preserving the meaning of data is a human affair, requiring continuous curation.” Data processing, analysis, and communication expertise, together with close ties to designated scientific communities, have been discussed as factors that put data repositories in a position to add value by delivering data products that align with practices and technologies of domain communities (Ember and Hanisch, 2013). Indeed we add to the Ember and Hanisch description of the role of domain repositories to emphasize the addition of value via the creation of new derived and transformed data products.

The engagement of the center with the community and the presence of researchers on data product teams facilitate review and vetting to ensure a consistent *series* of data products. Vetting, a time-consuming activity central to the scientific process, ensures the interpretation and context of data is agreed upon and understood collectively. Developing vetted, shared products establishes community-recognized departure or reference points for research. Assembling and describing the provenance or lineage of data products enhances the ability of researchers to review and revisit various stages of the data processing and analysis. Documentation of provenance ensures any particular data product can be referenced as a recognized starting point for subsequent work. For the first case discussed below, the initial data product (a single time series of data) and the scientific community (cryospheric scientists) are well defined. However, the development of derived data products can continue for some time, as shown in Table 1. Vetting is a process that must accompany the development of derived products no matter the audience.

Two Examples of Data Product Development

We describe the development of two extensive collections of data products created for multiple audiences by NSIDC. One collection characterizes sea ice products derived from satellite remote sensing data and unfolds over three decades. The second collection characterizes Greenland Ice Sheet melt, where development of an initial suite of data products over a period of several months was informed by previous experience.

Sea Ice Data Products

Satellite-based, passive microwave remote sensing has been a boon to polar research because of its ability to provide surface information from sea ice covered regions with relative insensitivity to the cloud cover and darkness covering the polar regions much of the year. Starting in 1972, remotely sensed sea ice data from a satellite was generated with a global spatial coverage astonishing to those familiar with earth-bound sampling efforts (Parkinson, 1989). The use of passive microwave remote sensing marks the beginning of an era of data product evolution.²

Sea ice algorithms over time

The archiving of sea ice data from satellites at NSIDC began in conjunction with a single designated community. Early in the satellite era, the primarily NASA-funded research community interested in broad-scale, sea ice processes recognized the potential of polar orbiting satellites to greatly expand their ability to observe sea ice and monitor its dynamics. The community developed a set of algorithms for producing sea ice extent and sea ice concentration values from the measured passive microwave brightness temperatures. Each algorithm performed well when compared with validations, but also showed limitations under some conditions; no single algorithm was found to be clearly superior and all yielded similar trends and anomalies. NSIDC initially produced two products designated NASA Team and Bootstrap, but other products were also available from other centers. Over time, NSIDC began producing its own versions of the NASA Team and Bootstrap products to provide more timely updates. Since no algorithm was clearly superior, there were fraught discussions among researchers and within the NSIDC science advisory group PoDAG (Polar DAAC User Working Group) trying to discern and assert which was the more “authoritative” product given differing approaches and political sensitivities. In addition, PoDAG was concerned about the added costs to NSIDC of distributing multiple products. In time, NSIDC developed into a neutral broker able to speak authoritatively from a data management perspective by describing each product in terms of their relative strengths and weaknesses without passing scientific (or political) judgment. Disparate products could be appreciated for their differences as they were archived and made accessible at NSIDC and elsewhere. Ready access to data in a shared digital forum and attention to documentation facilitated the turn from competing interests to acceptance of a number of community-recognized data products originating from a single source but processed using different algorithms (Meier et al., 2001; Parkinson, 1992), but tensions and confusion continued.

Though the need for consistent assembly and documentation of data products was recognized, cryospheric researchers were faced also with instrumentation that changed subtly over time (Cavalieri et al., 1999; Parkinson et al., 1999). They worked independently and informally over the years, managing a growing number of data

² See: <http://nsidc.org/data/seaice/pm.html>

products. With increasing interest in understanding climate systems and climate change, the research community recognized the importance and difficulties of creating a consistent time-series of sea ice data from the variety of instruments deployed over the years on different satellites. Earlier experience prepared them somewhat for the challenges of conveying the intricacies of data processing and analysis, a topic that continues to command considerable attention (Meier et al., 2009; Eisenmann et al., 2014).

Figure 1 illustrates the complexity of the satellite sea ice data product production process. Note, this overview was first created in 2010, decades after the production of data products began. A number of independent research activities were carried out before such an integrative view was conceived as an internal aid and subsequently recognized as significant enough to publish.

Sea ice data products over time

Over time, the passive microwave remote sensing segment of the sea ice community began to develop an understanding of the differences among the myriad sea ice data products, but researchers outside this community were less sanguine and still confused. They frequently asked which product would be best suited to their purposes. The idea of designating a winner or ‘best’ algorithm again surfaced. Ultimately, NSIDC responded by developing a large amount of explanatory material. A timeline of data product development that occurred over several decades is summarized in Table 1, along with external events, audience, and new audience needs. Development continues today. There are likely to be new products for new audiences, including operational concerns such as marine transportation and shipping (NRC, 2012b).

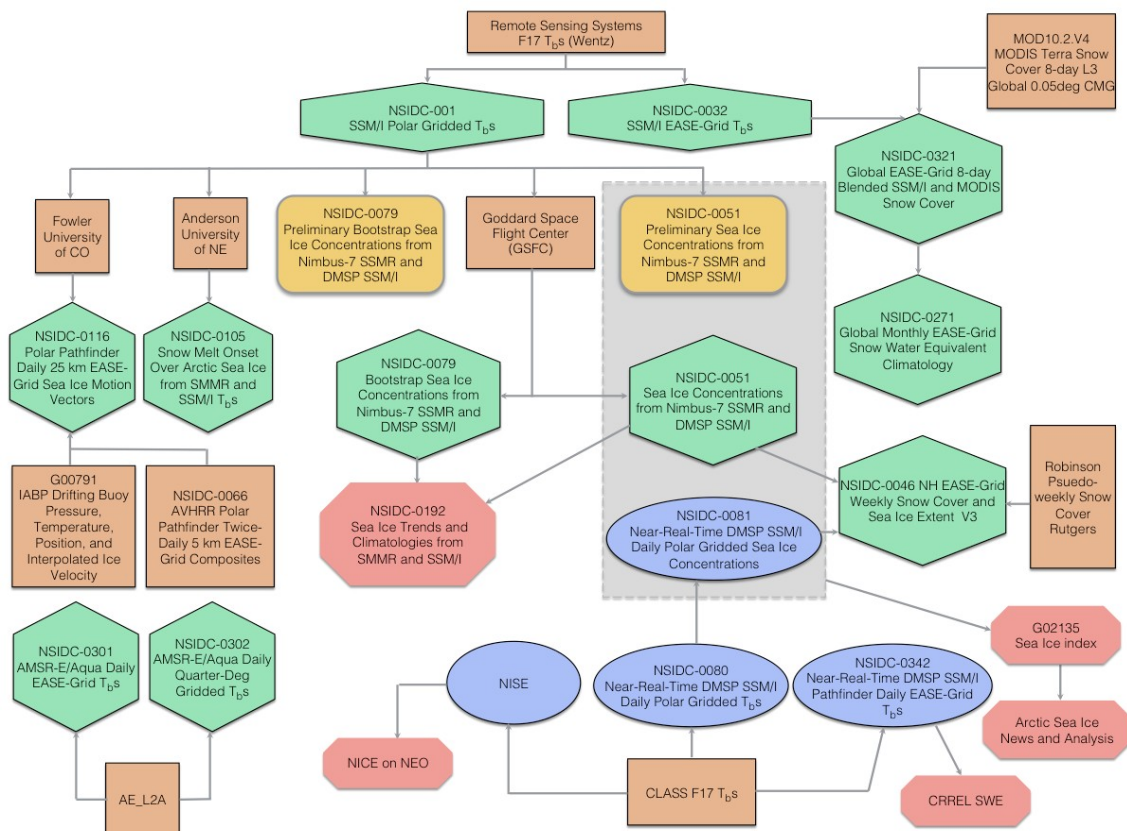


Figure 1. Passive Microwave Production Data Workflow circa 2010, where source data are

shown as boxes (brown), value added products as octagons (red), near real-time products as ovals (blue), preliminary products as boxes with rounded corners (gold), and final products as hexagons (green). A dashed light gray box contains three data products in the same product line referenced in Table 1. Brightness temperature is abbreviated T_b s. Redrawn from original work by Donna Scott, who managed the NSIDC Passive Microwave Product Team.

In the early 2000s, a Frequently Asked Questions (FAQ) website was created that compared and contrasted the characteristics of the data produced by each algorithm. It detailed what types of uses were most appropriate for each product. This information was initially somewhat controversial among NSIDC scientists, who found themselves in a position of judging their colleagues work outside normal scientific channels. As with most scientific communities, assessment was typically carried out in academic publications. The idea that a data center might recommend particular products for particular uses would traditionally be seen as outside their purview. Nevertheless, information about the data products proved to be valuable to researchers. The development of guidance materials continues today, with full support from both science teams providing the data and communities using the data.

Table 1. Summary of events, audience, needs, and products developed for sea ice.

Dates	External Event	New Audiences	New Needs	New Products
1970s to present	Availability of Passive Microwave Brightness Temperatures from satellites	Passive Microwave sea ice experts	Sea ice products covering the whole Arctic aggregated from multiple instruments	A variety of products (see NSIDC-0001 ³ , NSIDC-0080 ⁴) including gridded sea ice concentration products (see NSIDC-0081 ⁵ , NSIDC-0051 ⁶)
Mid 1990s	Availability of many different sea ice time series products	Sea ice researchers	Sea ice researchers that weren't passive microwave experts need advice as to which product to use for their purposes	FAQs, Technical reports, algorithms descriptions, sea ice related web pages ⁷

³ NSIDC-0001: <http://nsidc.org/data/nsidc-0001>

⁴ NSIDC-0080: <http://nsidc.org/data/nsidc-0080>

⁵ NSIDC-0081: <http://nsidc.org/data/nsidc-0081>

⁶ NSIDC-0051: <http://nsidc.org/data/nsidc-0051>

⁷ See <http://web.archive.org/web/20020225101044/http://nsidc.org/data/seaiice/data.html> for an example listing of the products available from NSIDC in 2002.

Table 1. Summary of events, audience, needs, and products developed for sea ice (continued)

Dates	External Event	New Audiences	New Needs	New Products
2000-2004	Statistically significant decreases in Arctic sea ice extent observed in the time series	Climatologists	Climatologies, monthly mean values, anomaly maps, etc.	Sea Ice Index (G02135) ⁸ , Sea Ice Trends and Climatologies (NSIDC-0192)
2005-2006	Arctic sea ice minimum observed in 2005	Biologists and other research communities; general news reporters; public awareness grows with book “An Inconvenient Truth”	Analyses of Arctic sea ice conditions and general information about sea ice and the Arctic	Arctic Sea Ice News and Analysis (ASINA) ⁹ including its links to general information about sea ice and the Arctic
2007	Record minimum Arctic sea ice extent and extensive media coverage	Biologists and other research communities outside the sea ice polar climate community	Advice about which products to use for their purposes	Updated sea ice related web pages, product FAQs, etc.
2007	Record minimum Arctic sea ice extent and extensive media coverage	Wide-spread public	Answers to public questions about sea ice and to correct misinformation spreading through the social media	Icelights: Your Burning Questions About Ice and Climate ¹⁰ ; Updated educational materials about sea ice (e.g., All about sea ice ¹¹)

By 2005, the satellite record was long enough to show what was beginning to be recognized as a statistically meaningful declining trend in sea ice extent (Parkinson, 2006). This made these data, which heretofore had primarily been of interest only to the sea ice community, interesting to the broader climate change community. And these users had different needs than the original users. In particular, climatologists were primarily interested in derived products: monthly mean values; climatologies or the average sea ice extent over the period of record; and anomaly maps depicting how much

⁸ Sea Ice Index (G02135): <http://nsidc.org/data/g02135>

⁹ ASINA: <http://nsidc.org/arcticseaicenews/>

¹⁰ Icelights: Your Burning Questions About Ice and Climate: <http://nsidc.org/icelights/>

¹¹ All about sea ice: <http://nsidc.org/cryosphere/seaice/index.html>

any particular measurement varied from the average value for that area for that day. This prompted development of the sea ice index¹², an innovation that conveys current conditions to specialist and non-specialist alike. The index provides a quantitative representation of sea ice conditions for Polar Regions over long periods of time. It is a proxy that gives a quick but consistent overview of sea ice extent and concentration in image and tabular form.

Due to increased interest, NSIDC also began to produce periodic press releases discussing Arctic sea ice conditions at the end of the summer melt period. Initially the audience for these releases was science reporters. However, as the sea ice extent continued to decline, attention from the general media rose dramatically, leading to a barrage of queries for basic information about the Arctic (e.g., Where is it?) and sea ice (e.g. Is Arctic sea ice disappearing?). To provide these users the information they needed without drastically increasing the outreach staff at NSIDC, the Arctic Sea Ice News and Analysis (ASINA) web site was launched in the fall of 2006. Begun as a simple news site with graphics and links to general information about sea ice, the site quickly became NSIDC's most popular web page, particularly for the few week period around the time when the sea ice reaches its minimum each year in September.

In 2007 the minimum sea ice extent shattered its previous record by 27%, leading to a renewed barrage of media interest. Some 150 media contacts were made, including front-page coverage in The New York Times and reporting in most major media. The coverage broadened awareness of the daily data underlying the ASINA site and increased the number of researchers using the data. This event and its associated coverage also led to the general public becoming a constant and vocal audience. NSIDC was contacted by increasing numbers of people who had heard reports of declining sea ice. These individuals often lacked basic scientific backgrounds but wanted to look at the data themselves. In addition, upon investigating a number of reports that incorrectly cited the data or analysis, it became apparent that with the great public interest in sea ice there was some misinformation proliferating on the web through social media.

To prevent user support staff from becoming overwhelmed by the need to answer the same or similar questions being raised by large numbers of readers, NSIDC investigated ways of addressing their questions as a group. This led NSIDC to develop a new website, *Icelights: Your Burning Questions About Ice and Climate*, in 2011. Monitoring the online media for topics of active interest, NSIDC formulates responses in *Icelights* as part of the public conversation. The site aims to present “what’s hot in the news around climate and sea ice, provide a behind-the-scenes view of what scientists are talking about, and answer the questions you and your fellow readers send us.” The site also includes general information about sea ice and data, provides pointers to more information about sea ice and climate change for people with the time or inclination to delve deeper, and allows users easy access to like/share/tweet the posts. Each post on the *Icelights* site is based on interviews with researchers, includes photos or graphics that help explain the topic, undergoes scientific review prior to publication, and includes references to published papers and research web sites associated with its content.

Greenland Ice Sheet Melt Data Products

Data from the Greenland ice sheet provides a second example of data product evolution. Development efforts for this project were made easier by insights gained from experience with the sea ice data products.

In 2006, Dr. W. Abdalati gave NSIDC roughly 30 years of Greenland ice sheet melt data derived from the same passive microwave sensors used to derive the sea ice

¹² Sea ice index: <http://nsidc.org/data/g02135>

products. The data, as it originated from the investigator, consisted of a suite of daily melt files in ASCII format and included only the pixels in a defined subset of the standard Northern Hemisphere polar stereographic grid that were melting on that day. Each pixel where melting was detected was given its own line in the file. This provided an easy-to-use data format for researchers studying how melting of the Greenland ice sheet was changing over time. For example, a file with three rows indicated that three pixels showed signs of melting on that day.

While changes in the sea ice data products developed slowly over time, for Greenland the product development occurred as the source data was ingested into the NSIDC data repository. It was recognized that the data would also be of interest to a wide variety of audiences, including the climate community and potentially the public given its implications for sea level rise. Drawing on experience with sea ice data, an assessment of the analytic potential (Palmer et al., 2011) of the data was incorporated in the data accession process. A variety of additional data products were generated, including gridded daily melt status files, gridded annual files where each pixel contained a count of the number of days of melt that year, and a climatology of the entire time series from which anomaly maps could be generated. Moreover, the data was made available as a KML file so that it could be displayed using Google Earth. Further, the data was added to a number of NSIDC data visualization and analysis tools, e.g. the Greenland surface melt layer was created for the North Pole view in the Atlas of the Cryosphere (see NSIDC-Atlas¹³).

In 2012 the issue of Greenland ice melting hit the public's consciousness when, for a few days in July, the entire ice sheet indicated surface melt. Data product team activities took into account use by multiple audiences and support for a continuing conversation with research scientists. The melting event prompted a rapid response; NSIDC built the Greenland Ice Sheet Today website, which launched in January 2013, using near-real-time data (Mote, 2007; Tedesco et al., 2008), data that as part of an active ongoing research program had not yet been ingested into the NSIDC archive. As with the ASINA site, the Greenland Today site provides expert analysis of environmental conditions, with links to basic information about ice sheets and the data used. While it is too early to tell whether the Greenland site will achieve the consistently high readership of ASINA (ASINA consistently has over a 100,000 unique page views per month with peak readership of over 220,000 in the summers of 2012-2014), it should be noted that Greenland Today had a peak of 36,600 unique visitors in July 2014 with users generally staying on the site for longer than three minutes.

Discussion

The NSIDC catalog currently lists 199 sea ice data products of which 24 are derived from passive microwave sensors. Some products are intended to support scientific research and others are for data use by non-expert, general audiences. In this section factors that broaden audiences and the mobilization of knowledge are discussed.

Expanding Audience

Parsons and Duerr (2005) describe several reasons for an ever-broadening user base for data products. One is development of new understandings and unanticipated creative use of increasingly available data; another is rapidly evolving technology. Additional

¹³ NSIDC-Atlas: <https://nsidc.org/data/atlas/>

factors that stimulate interest in data products by new audiences are suggested by the Arctic product development case examples: 1) external events as triggers for data product development, 2) access to diverse data products at differing levels of generality and interpretation, and 3) delivery of products using new modes of communication.

Data product development: Multi-cycle trajectory

The development of a collection of data products over time summarized in Table 1 reveals data product development as part of a multi-cycle trajectory. While vastly oversimplifying what is actually a complex series of interactions between audiences, funding sources, and repositories, this cycle is shown in Figure 2. The occurrence of an event is portrayed as triggering interest of a new audience with particular needs that are met by new products. In such cases, development responds to what Wallis et al. (2014) describe as “opportunities for known reuse.” New events may then trigger the product development cycle anew. The availability of passive microwave brightness temperatures from satellites prompted passive microwave sea ice experts to recognize the need for an agreed upon time-series of sea ice products, which led to gridded sea ice concentration products. Recognition of a statistically significant decrease in Arctic sea ice extent, together with the desire for more attractive, dynamic, and reusable graphics, stimulated the interest of climatologists. This led in turn to recognition of their need for climatologies, monthly mean values and anomaly maps that resulted in the development of the sea ice index. Event triggers may take many forms – an unanticipated environmental disturbance, widespread media coverage, availability of data products, a new collaborative forum, an organizational realignment, or perhaps the timely arrival of a new leader. Table 1 shows early triggers related to availability of data products, while the last two entries document large-scale environmental events that prompted interest from research communities as well as the public.

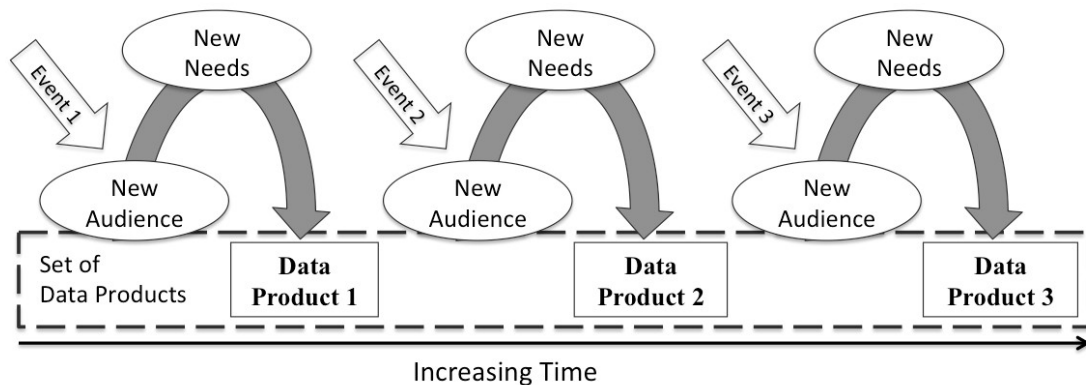


Figure 2. A simplified view of the continuing development of scientific data products. Each cycle is initiated by one or more events that create a new audience that leads to generation of a new data product in response to the needs of a recently identified designated user community.

Data product diversity: Multi-level collection

Sea ice data product dependencies are outlined in Table 2. The multiple products originate from a single data source but differ in important ways. Earth science data products have a range of temporal resolutions as well as spatial extents and resolutions,

so frequently convey varying degrees of breadth and/or specificity. Taken together, the products provide a multi-dimensional understanding of the system observed and create a scaffold for learning for multiple audiences. In addition to detailing data product interdependencies, Table 2 classifies the products using a data processing levels framework (see NSIDC-Drift¹⁴) detailed below.

In work with satellite data processing, NASA defined a five level data processing framework for the earth observing system program (NASA, 1986). Data moves through multiple stages of processing and analysis from the initially recorded data to derived data. The originally recorded data is designated Level 0. Level 1 data is reconstructed, unprocessed, full-resolution sensor units (e.g. voltages) with temporal and geospatial calibration. Level 2 data consists of geophysical variables derived from the sensor units (e.g. sea ice concentrations), and Level 3 data provides variables interpolated and/or merged to standard structure (e.g. spatial grid, time grid). Level 4 denotes derived data transformed or modeled based on analyses of lower level data in response to research needs (e.g. sea ice indexes).

Designed for practical use, the levels are a mix of processing and interpretation categories. Levels 1-3 are progressive enhancements, with these levels situating the data within spatial, temporal reference systems. In the recent design of the National Ecological Observatory Network (NEON) for continental scale analysis and forecasting, use of this empirical classification scheme continues (Keller, 2010; Keller et al., 2010). Data products for NEON Levels 0-3 are typically derived from a single instrument, observer, or a specific sampling location. Given the value of the levels as a shared concept that supports community-wide recognition of individual data products, some researchers have explored use of a level framework for humanities and social science data (Renear et al., 2009).

The NASA Level 4 category identifies derived data products interpreted for specific scientific communities. While NASA levels pertain to data processing targeting a scientific community, the sea ice products suggest additional levels to describe higher-level 'interpretive' products. In Table 2, then, Interpretive Level 1 indicates data products derived from lower level data that are interpreted for diverse scientific audiences. Data products at this level may be accompanied by visuals and text using various delivery mechanisms, including contextual information in recognition of the potential non-scientific-oriented knowledge bases of the audiences. Beyond the first interpretive level, higher levels involve alternative forms for conveying products (e.g. annual rather than hourly averages; regional rather than square kilometer measures) as well as for delivery methods (e.g. newsletters and blogs). Interpretive Level 2 indicates derived data products presented periodically that are packaged with continuing expert data analysis using a variety of delivery mechanisms and may be accompanied by relevant lower level data. Interpretive Level 3 also involves packaging and presentation of derived data products but uses general language for responses to public audiences concerned about higher-level products. This level represents a loosely structured dialogue where data interpretation, analysis and synthesis are presented in response to public discourses monitored in a variety of social media. At a time when levels are being reinterpreted in new domains as data provenance descriptors (Bose and Frew, 2005) and as text encoding levels (Renear et al., 2009), the interpretive levels 1-3 have been created as a mechanism for exploring data products generated as part of a particular scientific collection. Further research is needed on interpretive categories and inclusion of other types of data products from other contexts.

¹⁴ NSIDC-Drift: <http://nsidc.org/the-drift/2013/08/is-it-1b-2-or-3-definitions-of-data-processing-levels/>

Table 2. Sea ice data product dependencies and levels.

Precursor Product or Shortname	Precursor Product	Precursor Product Level	Product Shortname	Product	Product Level
	Antenna counts	L0	Wentz ATs	Wentz Antenna Temperatures	L1
	Antenna counts	L0	NOAA/CLASS ATs	NOAA/CLASS Antenna Temperatures	L1
Wentz ATs	Wentz Antenna Temperatures	L1	Wentz Tbs	Remote Sensing Systems Tbs (Wentz)	L2
NOAA/CLASS ATs	NOAA/CLASS Antenna Temperatures	L1	NOAA/CLASS Tbs	NOAA/CLASS Brightness Temperatures (Tbs)	L2
Wentz Tbs	Remote Sensing Systems Tbs (Wentz)	L2	NSIDC-0001	SSM/I Polar Gridded Tbs	L3
NOAA/CLASS Tbs	NOAA/CLASS Brightness Temperatures	L2	NSIDC-0080	Near Real-time DMSP SSM/I Daily Polar Gridded Tbs	L3
NSIDC-0001	SSM/I Polar Gridded Tbs	L3	NSIDC-0051	Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data	L4
NSIDC-0080	Near-Real time DMSP SSM/I Daily Polar Gridded Tbs	L3	NSIDC-0081	Near-Real-time DMSP SSM/I Daily Polar Gridded Sea Ice Concentrations	L4

Table 2. Sea ice data product dependencies and levels (continued)

Precursor Product or Shortname	Precursor Product	Precursor Product Level	Product Shortname	Product	Product Level
NSIDC-0051 and NSIDC-0081	Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data and Near-Real-time DMSP SSM/I Daily Polar Gridded Sea Ice Concentrations	L4	G02135	Sea Ice Index	I1
G02135	Sea Ice Index	I1	ASINA	Arctic Sea Ice News and Analysis	I2
ASINA	Arctic Sea Ice News and Analysis	I2	Icelights	Icelights: Your Burning Questions about Ice and Climate	I3

Data product delivery: Multi-mode communication

To capture the context of data sufficiently, the degree of specificity or minimum representation information depends a great deal upon the knowledge base of the audience (Parsons and Duerr, 2005). Documentation of a data product targeting a community associated with a particular scientific discipline may be briefer and more technical than that for a community who need some basic concepts of the discipline's knowledge base to be described. For example, sea ice researchers will be able to interpret data measurements labeled as 'reflectance' whereas individuals unfamiliar with analysis of satellite imagery will not understand the relationship of reflectance to sea ice without further explanation. The ramifications of various transformations, as well as the context of data, are difficult and time consuming to document in their entirety even within a particular scientific community (Schuurman and Balka, 2009).

The data product teams and scientific partners who vet data products and their documentation are aware of the choices being made in the analysis and packaging of data. The existence of multiple sea ice products assembled together provides the opportunity to reflect on the collection as a whole. Communication about the products facilitates the generation of scientific consensus as well as community-recognized responses to questions by various interested parties. Some descriptive detail, however, requires sophisticated scientific analysis that goes beyond the scope of typical data documentation. As a result, in addition to citing the scientific literature, there are special reports on data related topics. In an NSIDC report series, three reports are devoted to deeper analysis of passive microwave-derived sea ice products (Maslanik et al., 1998; Stroeve, Li and Maslanik, 1997; Stroeve and Smith, 2001). In another satellite case, an

NRC (2009) report argued for the value of sea ice derived products and outlined details of proposed dissemination.

The traditional aims with documentation are to adequately describe data and its uncertainty for future users, as well as to organize data and its description for discovery in order to avoid obscurity. We found that more than added context is needed for multiple audiences. Different kinds of documentation for the various data products are needed to reach multiple audiences. For instance, science blogs provide an example of new forms for communicating science in public spaces. However, within these informal arenas there may be inadequate analysis of complex systems. Presentations of incomplete, non-peer-reviewed analysis and commentary may be driven by interest in what is understood within scientific arenas but may also be due to interest in creating confusion and distraction (e.g. Watts, 2008a, 2008b). In one of a series of posters¹⁵ presented at the American Geophysical Union annual meeting that document new roles in communicating not only with science writers but also with the public, the visits to the 2007 ASINA website showed the public beginning to compete with the media as a “vocal and constant audience” (Renfrow et al., 2008). As a result, communicating science became a priority at NSIDC. Despite being an activity rife with pitfalls, the Arctic Sea Ice News and Analysis website was developed with the aim to explain climate science to the general public (Leitzell and Meier, 2009, 2010).

Data Teams

The data product team approach at NSIDC represents a relatively stable configuration that has proven suitable for supporting the development of derived data products in close collaboration with a variety of communities and in response to a variety of audiences. Development of useful collections of data products requires ‘situational awareness.’ Situational awareness or familiarity with community knowledge bases and community group activities foregrounds the importance of taking into account diverse perspectives (Palmius, 2005; Suchman et al., 1991; Haraway, 1988). Though data systems are able to aggregate and provide access to data and information for a community, Dourish and Bellotti (1992) point out some dangers of information systems separated from shared community workspaces: they often suffer from presuppositions of relevance to users and have difficulties with access. With the two case examples from a data center that is not co-located with its communities, situational awareness is generated via engagement with partners, a commitment called for in the NSIDC service mission. The role of NSIDC as a design center is anchored by working relationships with community partners in conjunction with the data team’s ready access to the data and familiarity with the data. While NSIDC aggregates community data products and generates new products, community group participants provide input on identified needs as well as peer review of products.

Mobilizing Knowledge

The concept of knowledge mobilization emerged in the late 1990s to describe “the flow of knowledge among multiple agents leading to intellectual, social and/or economic impact” (SSHRC, 2009). During a period with growing interest in big picture or systems view of kinds of research use (Weiss, 1979; NAP, 2012) and with kinds of research users (Nutley et al., 2007), knowledge mobilization was used to refer to

¹⁵ See the NSIDC news on events: <http://nsidc.org/news/events>

activities that increase the use of research outcomes. There are a variety of related terms such as knowledge management, translation diffusion, knowledge-to-action and implementation science (Levin, 2008; SSHRC, 2008). Some distinguishing features of knowledge mobility are an orientation toward change, situated awareness, and use of persuasive methods that depend upon research-based evidence (SSHRC, 2008).

Though often used in areas concerned with learning, such as education, knowledge mobilization is useful in considering data sharing arrangements within science where researchers aim to learn from the data, that is, from evidence produced by other scientists. A collection of data products represents a mechanism for mobilizing knowledge. Knowledge mobilization occurs through activities such as data dissemination, transformation, brokering, transfer, and co-production (SSHRC, 2009). Data repositories with the technical capacity for requisite data processing, analysis, and presentation ensure data can be found and accessed. Presentation, whether in the form of online collections, technical reports or social media, contributes to product dissemination.

The polar data product development examples demonstrate the kind of scientific reach and integration possible when mediation is recognized as a core activity that involves liaison work. Production of data in short timeframes and over a number of longer periods is evident in the Arctic examples, as data products were developed for a particular project in the 1970s and 30 years later for newly identified designated communities requiring higher levels of interpretation (see Table 1). NSIDC illustrates an approach to supporting the diverse, often interdependent stages of data work via their mission to support community data needs. Between research and uptake is a ‘knowledge-translation’ process, a boundary area populated by mediators. A data repository that changes data into a form appropriate for a designated community can be described as doing the work of transformation. They carry out collective work where understandings from earlier data products are reviewed, restated, translated, and transformed so as to facilitate uptake and improve understanding. Levin (2008) reports that third party organizations of all kinds – sometimes called brokers or mediators – play a critical role in the spread and impact of research but their nature and need for neutrality have not been fully explored. At NSIDC the role of the data product team with its situated awareness has evolved to carry out liaison or mediation work that contributes to mobilization of knowledge.

The Cost of Mobilizing Knowledge

It is difficult to counter institutional momentum and conservative tendencies relating to preservation efforts in order to respond to emergent information needs. As a result, a great deal of the sea ice data products work was initially ad hoc, a ‘skunk works project’¹⁶ primarily supported by the NOAA@NSIDC program that produced the products and automatically generated graphical displays of the results for the NSIDC website (Fetterer, 2002, 2003; Fetterer and Knowles, 2004). The products continue to be updated today.¹⁷ Development was made easier for subsequent projects, such as the Greenland Ice Sheet Melt, by experience gained with sea ice efforts. Experience and attention to planning are still lacking for prototyping and proposals supporting exploration of potential data products that arise in response to emergent data needs.

¹⁶ A skunk works project is a designation for an informal study carried out by a self-identified, loosely structured group that has relative independence within organizational management.

¹⁷ NSIDC Sea ice: https://nsidc.org/data/docs/noaa/g02135_seaice_index/

Establishing and maintaining reciprocal relations is time-consuming so recognizing, articulating, and planning communication must be established explicitly as a priority.

Support for internal skunk work projects, however, is tenuous at best. Unanswered questions remain about prioritization and support for the evolution of data production efforts. What time will repositories devote to new, incoming data versus time for transformation of existing data for new audiences? Will subject-based repositories evolve to play a role in development of data products? How are management perceptions of grant responsibilities to be balanced with researchers' perceptions of what constitutes scientific work?

In addition to the practical matters of managing the data, as well as producing the products and the technology required, NSIDC currently is developing new procedures and language for cost issues. A layered approach called 'levels of service' has been designed and is undergoing review and testing (Duerr et al., 2009). The aim is to provide a menu of cost options for those planning to submit data while establishing a sustainable approach to the work of ingestion and maintenance from the repository perspective. If such a cost model were to become part of the data culture, it provides a higher-level stability that Ribes and Finholt (2009) refer to as institutionalizing the work in terms of longer-term support and policy development. In contrast to top-down institutional change, this is a repository-specific attempt to address current unsustainable costs by establishing a new norm for the data culture. Piloting of a pay-as-you-go cost model is planned in an attempt to establish a new norm given our contemporary data culture that is struggling with issues of data support (Ember and Hanisch, 2013).

Conclusion

Though development of approaches to data sharing in the earth and environmental sciences is ongoing, efforts are nascent, disunited and incomplete. In the examples presented above, data work focuses initially on aggregating and delivering observational satellite data. Subsequent activities involving transformation, interpretation, visualization, and communication of scientific results resulted in creation of new data products. Data work in the Arctic sea ice and Greenland Ice Sheet cases reveals the importance of arrangements that support the development of new data products on an ongoing basis. They underscore how the concept of completion and categorization as 'done' does not apply to a collection of scientific data products in the same way it is understood for a published journal article.

A 'continuing development' perspective requires broad interpretation of mission mandates, especially when unanticipated events serve as prompts for innovation. The availability of tailored data products at differing levels of processing and interpretation makes scientific knowledge more readily available to a variety of audiences and facilities. Each product release may be seen as one step in a multi-cycle trajectory of data product development that is spurred by events that create new user communities with new needs that in turn catalyze the creation of new products. The concept of a multi-cycle data product development trajectory requires an infrastructure to support the observed co-evolution of data products and designated communities.

Discussion of data product development contributes to awareness of a potential role for subject-based data repositories. It is not just metadata for individual datasets that facilitates data use but also a diversity of data products. A collection of data products

crafted for various designated communities provides multiple interdependent windows into the phenomena observed and serves as a substrate from which others learn and build. A multiplicity of data products for a multiplicity of designated communities represents both a strategy for data dissemination and an approach to transfer of data, documentation, and knowledge.

The two earth science cases illustrate an advanced set of services ensuring mobilization of knowledge that facilitates data use and enhances the impact of scientific research. A collection of data products can reach beyond scholarly discourse to inform planning, mitigation, policy-making and other areas. Collections of data products are an information resource that provides a way for both experts and non-experts to be better informed about the state of the environment. As we begin to acknowledge the need to manage the earth as a whole, a collection of data products for multiple designated communities represents one approach to mobilizing knowledge about earth's dynamic systems.

Acknowledgements

A DCERC program (Institute of Museum and Library Services RE-02-10-0004-10) provided fellowship support to K.S. Baker for work at the National Center for Atmospheric Research that informed this study. The authors would like to thank Walt Meier and the many NSIDC staff members who provided helpful comments, especially Roger Barry, a research scientist with considerable insight into data issues and NSIDC's director during much of the time period discussed. Ken Knowles and Mark Parsons led the NSIDC 'data jam'.

References

- Berman, F. (2014). Building global infrastructure for data sharing and exchange through the research data alliance. *D-Lib Magazine*, 20(1/2). doi:10.1045/january2014-berman
- Borgman, C.L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Bose, R., & Frew, J. (2005). Lineage retrieval for scientific data processing: A survey. *ACM Computing Surveys*, 37(1), 1–28. doi:10.1145/1057977.1057978
- Brossard, D., & Lewenstein, B.V. (2010). A critical appraisal of models of public understanding of science: Using practice to inform theory. In L. Kahlor & P.A. Stout (Eds.), *Communicating science: New agendas in communication*. New York: Routledge.
- Cavaliere, D. J., Parkinson, C.L., Gloersen, P., Comiso, J.C., & Zwally, H.J. (1999). Deriving long-term time series of sea ice cover from satellite passive-microwave multisensory data sets. *Journal of Geophysical Research*, 104, 15803-15814. doi:10.1029/1999JC900081

- CCSDS. (2012). *Reference model for an Open Archival Information System (OAIS)*. Washington DC: CCSDS 650.0-M-2, Magenta Book. Issue 2. June 2012. Retrieved from the Consultative Committee for Space Data Systems website: <http://public.ccsds.org/publications/RefModel.aspx>
- Cragin, M.H. (2009). Scientific data collections: Use in scholarly communication and implications for data curation. PhD, University of Illinois at Urbana-Champaign.
- Dourish, P., & Bellotti, V. (1992). *Awareness and coordination in shared workspaces*. Paper presented at the 1992 ACM Conference on Computer-Supported Cooperative Work, Toronto, Canada.
- Duerr, R., Leon, A., Miller, D., & Scott, D.J. (2009). Levels of service: National Snow and Ice Data Center (NSIDC). National Snow and Ice Data Center, online resource. Retrieved from [https://nsidc.org/sites/nsidc.org/files/files/NSIDCLevelsOfService-V2_0a\(2\).pdf](https://nsidc.org/sites/nsidc.org/files/files/NSIDCLevelsOfService-V2_0a(2).pdf)
- Edwards, P.N., Jackson, S.J., Chalmers, M.K., Bowker, G.C., Borgman, C.L., Ribes, D., Burton, M., & Calvert, S. (2013). Knowledge infrastructures: Intellectual frameworks and research challenges. Ann Arbor: Report of a workshop sponsored by the National Science Foundation and the Sloan Foundation, Deep Blue.
- Eisenman, I., Meier, W.N., & Norris, J.R. (2014). A spurious jump in the satellite record: Has Antarctic sea ice expansion been overestimated? *The Cryosphere*, 8(4), 1289-1296. doi:10.5194/tc-8-1289-2014
- Ember, C., & Hanisch, R. (2013). Sustaining domain repositories for digital data: A white paper. Paper presented at the Sustaining Domain Repositories for Digital Data Workshop, University of Michigan. Retrieved from http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRD_D_121113.pdf
- Fetterer, F. (2002). Sea ice index: Interpretation resources for sea ice trends and anomalies. NSIDC Informal Technical Report. Retrieved from http://nsidc.org/data/docs/noaa/g02135_seaice_index/interpretation.html
- Fetterer, F. (2003). Recent Arctic ice extent minima observed with the sea ice index. Paper presented at the ARCUS study of environmental Arctic change (SEARCH) Open Science Meeting, Seattle, Washington. Retrieved from https://nsidc.org/data/docs/noaa/g02135_seaice_index/docs/SEARCH_2003_iceindex_poster.pdf
- Fetterer, F., & Knowles, K. (2004). Sea ice index monitors polar ice extent. *Eos: Transactions of the American Geophysical Society*, 85, 163. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1029/2004EO160007/full>
- Galison, P.L., & Stump, D.J. (1996). *The disunity of science: Boundaries, contexts, and power*. Palo Alto, CA: Stanford University Press.

- Genova, F, Hanahoe, H., Laaksonen, L., Morais-Pires, C., Wittenburg, P., & Wood, J. (2014). The data harvest: How sharing research data can yield knowledge, jobs and growth. In R.L. Hudson & N. Moran (Eds.), *The data harvest report*. RDA Europe. Retrieved from https://europe.rd-alliance.org/sites/default/files/repository/files/TheDataHarvestReport_%20Final.pdf
- Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575-599. doi:10.2307/3178066
- Hilborn, R. & Mangel, M. (1997). *The ecological detective: Confronting models with data*. Princeton, NJ: Princeton University Press.
- International Council for Science. (2004). ICSU report of the CSPR assessment panel on scientific data and information. Paris: ICSU.
- Kansa, E.C., Kansa, S.W., & Arbuckle, B. (2014). Publishing and pushing: Mixing models for communicating research data in archaeology. *International Journal of Digital Curation*, 9(1), 57-70. doi:10.2218/ijdc.v9i1.301
- Keller, M. (2010). NEON level 1-3 data products. Technical Report NEON.MGMT.DPS.005004.REQ. National Ecological Observatory Network, Boulder, CO.
- Keller, M., Alves, A., Aulenbach, S., Johnson, B., Kampe, T., Kao, R., Kuester, M., Loescher, H., McKenzie, V., & Powell, H. (2010). NEON scientific data products catalog. Technical Report NEON.MGMT.DPS.005003.REQ. National Ecological Observatory Network, Boulder, CO.
- Latour, B. (1999). Circulating reference: Sampling the soil in the amazon forest. In B. Latour (Ed.), *Pandora's Hope: Essays on the Reality of Science Studies* (pp. 24-79). Cambridge: Harvard University Press.
- Leitzell, K., & Meier, W.N. (2009). Uncloaking the scientific process. *Eos Transactions of the American Geophysical Union*, 90(52), Abstract presented at 2009 Fall Meeting. Retrieved from <http://abstractsearch.agu.org/meetings/2009/FM/PA21A-1298.html>
- Leitzell, K., & Meier, W.N. (2010). Propaganda, News, or Education: Reporting Sea Ice Conditions. Abstract presented at AGU 2010 Fall Meeting Session Priorities and Pitfalls: Pathways for Effective Communication, San Francisco, CA. Retrieved from <http://abstractsearch.agu.org/meetings/2010/FM/PA21C-1643.html>
- Levin, B. (2008). Thinking about knowledge mobilization. Canadian Council on Learning, Toronto, Ontario.
- Longino, H. (1990). Science as social knowledge: Values and objectivity in scientific inquiry. *Good science, bad science*. Princeton: Princeton University Press.

- Maggio, I., Molch, K., & Albani, M. (2015). Long term preservation of Earth observation space data, preservation workflow (Draft v1.4). http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Best_Practices/Preservation%20Workflow_v1.0.pdf
- 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 from the National Snow and Ice Data Center Web site: https://nsidc.org/pubs/documents/special/nsidc_special_report_8.pdf
- Mayernik, M.S. (2015). Research data and metadata curation as institutional issues. *Journal of the American Society of Information Science and Technology*. doi:10.1002/asi.23425
- Meier, W.N., VanWoert, M.L., & Bertoina, C. (2001). Evaluation of operational SSM/I algorithms. *Annals of Glaciology*, 33 102-108. Retrieved from https://www.researchgate.net/publication/233582617_Evaluation_of_operational_SSMI_ice-concentration_algorithms
- Meier, W., Stroeve, J., Duerr, R., & Fetterer, F. (2009). Help, I don't know which sea ice algorithm to use?!: Developing an authoritative sea ice climate data record. Poster presented at AGU Fall Meeting. Retrieved from <http://adsabs.harvard.edu/abs/2009AGUFM.C31A0435M>
- Mote, T.L. (2007). Greenland surface melt trends 1973–2007: Evidence of a large increase in 2007. *Geophysical Research Letters* 34, L22507. doi:10.1029/2007GL031976
- NAP. (2012). Using science as evidence in public policy. In K. Prewitt, T.A. Schwandt & M.L. Straf (Eds.), Committee on the Use of Social Science Knowledge in Public Policy, National Research Council.
- NASA. (1986). Earth observing system. Data and information system. Volume 2A: Report of the EOS Data Panel. NASA Technical Memorandum 87777. Retrieved from <http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19860021622.pdf>
- NDSA. (2014). National agenda for digital stewardship: National Digital Stewardship Alliance. Retrieved from <http://www.digitalpreservation.gov/ndsa/documents/2014NationalAgenda.pdf>.
- NRC. (1999). Review of NASA's distributed active archive centers. Washington D.C.: National Research Council, The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=6396
- NRC. (2003). Government data centers: Meeting increasing demands. National Research Council, The National Academies Press, Washington D.C. Retrieved from http://www.nap.edu/catalog.php?record_id=10664

- NRC. (2009). Scientific value of Arctic sea ice imagery derived products. Washington D.C.: National Academies Press.
- NRC. (2012a). Communicating science and engineering data in the information age. Panel on Communicating National Science Foundation Science and Engineering Information to Data Users, National Research Council.
- NRC. (2012b). Seasonal-to-decadal predictions of Arctic sea ice: Challenges and strategies. Washington D.C: National Academies Press.
- NRC. (2014). Sustainable infrastructures for life science communication: Workshop summary. National Research Council. National Academies of Science.
- NSF. (2011). Data management plan requirements, grant proposal guide. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp - dmp
- NSF. (2013). Earthcube: Developing a community-driven data and knowledge environment for the geosciences. Retrieved from <http://www.nsf.gov/pubs/2013/nsf13529/nsf13529.pdf>
- Nutley, S. M., Walter, I., & Davies, H.T. (2007). *Using evidence: How research can inform public services*. The Policy Press.
- OSTP. (2013). Increasing public access to the results of scientific research. Office of Science and Technology Policy. Retrieved from <https://petitions.whitehouse.gov/response/increasing-public-access-results-scientific-research>
- Palmer, C.L. (2004). Thematic research collections. In S. Schreibman, R. Siemens & J. Unsworth (Eds.), *A companion to digital humanities* (Vol. 21, pp. 348-365). Oxford: Blackwell.
- Palmer, C.L., Weber, N.M., & Cragin, M.H. (2011). The analytic potential of scientific data: Understanding re-use value. In Proceedings of the American Society of Information Science and Technology, ASIST 2011, New Orleans.
- Palmius, J. (2005). Defining the ‘information’ part of ‘information system’. Paper presented at the Proceedings of Information Systems Research in Scandinavia, IRIS 2005.
- Parkinson, C.L. (1989). On the value of long-term satellite passive microwave data sets for sea ice/climate studies. *GeoJournal*, 18(1), 9-20. Retrieved from <http://www.jstor.org/stable/41144360>
- Parkinson, C. (1992). Spatial patterns of increases and decreases in the length of the sea-ice season in the north polar region, 1979–1986. *Journal of Geophysical Research: Oceans* (1978–2012), 97, 14377-14388. doi:10.1029/92JC01367

- Parkinson, C.L., Cavalieri, D.J., Gloersen, P., Zwally, H.J., & Comiso, J.C. (1999). Arctic sea ice extents, areas, and trends, 1978–1996. *Journal of Geophysical Research: Oceans (1978–2012)*, 104(C9), 20837-20856.
doi:10.1029/1999JC900082
- Parkinson, C.L. (2006). Earth's cryosphere: Current state and recent changes. *Annual Review of Environmental Resources*, 31, 33-60.
doi:10.1146/annurev.energy.31.041105.095552
- Parsons, M.A., & Duerr, R. (2005). Designating user communities for scientific data: challenges and solutions. *Data Science Journal*, 4, 31-38. Retrieved from http://sdl.syr.edu/course/week9/9_DataAggregation_ParsonsDuerr.pdf
- Patel, D.J. (2012). *Data jujitsu: The art of turning data into data product*. Radar, O'Reilly. Retrieved from <http://radar.oreilly.com/2012/07/data-jujitsu.html>
- Penev, L., Mietchen, D., Chavan, V., Hagedorn, G., Remsen, D., Smith, V., & Shotton, D. (2011). Pensoft data publishing policies and guidelines for biodiversity data. Retrieved from http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf
- Renear, A.H., Dolan, M., Trainor, K., & Cragin, M.H. (2010). Towards a cross-disciplinary notion of data level in data curation. *Proceedings of the American Society for Information Science and Technology* 46(1), 1-8.
doi:10.1002/meet.2009.14504603103
- Renfrow, S., Mahoney, A.R., Meier, W.N., Scambos, T.A., Serreze, M.C., & Stroeve, J. (2008). Turning the tide: Serving science to a more-than-media audience. *Eos Transactions of the American Geophysical Union*, 89(53). Retrieved from <http://abstractsearch.agu.org/meetings/2008/FM/ED41A-0556.html>
- Ribes, D., & Finholt, T. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, 10(5), 375–398. Retrieved from <http://aisel.aisnet.org/jais/vol10/iss5/5>
- Royal Society. (1985). The public understanding of science. Retrieved from <http://royalsociety.org/policy/publications/1985/public-understanding-science>
- Schuurman, N., & Balka, E. (2009). alt. metadata. health: Ontological context for data use and integration. *Computer Supported Cooperative Work (CSCW)*, 18(1), 83-108.
- SSHRC. (2008). CCL/SSHRC Knowledge Mobilization Symposium (2008). Harrison Hot Spring, BC.
- SSHRC. (2009). SSHRC's knowledge mobilization strategy. Social Sciences and Humanities Research Council of Canada. Retrieved from https://www.ucalgary.ca/research/files/research/sshrc_knowledge_mobilization.pdf

- 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 from <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 from http://nsidc.org/pubs/documents/special/nsidc_special_report_10.pdf
- Suchman, L.A., & Trigg, R.H. (1991). Understanding practice: Video as a medium for reflection and design. In J. Greenbaum & M. Kyng (Eds.), *Design at Work: Cooperative Design of Computer Systems* (pp. 65-89). Hillsdale, NJ: Lawrence Erlbaum.
- Tedesco, M., Serreze, M., & Fettweis, X. (2008). Diagnosing the extreme surface melt event over southwestern Greenland in 2007. *The Cryosphere*, 2, 159–166. doi:10.5194/tc-2-159-2008
- Wallis, J. (2014). Data producers courting data reusers: Two cases from modeling communities. *International Journal of Digital Curation*, 9(1), 98-109. doi:10.2218/ijdc.v9i1.304
- Watts, A. (2008a). Arctic ice extent discrepancy: NSIDC vs cryosphere today. Retrieved from <http://wattsupwiththat.com/2008/08/15/arctic-ice-extent-discrepancy-nsidc-versus-cryosphere-today/>
- Watts, A. (2008b). NSIDC's Dr. Walt Meier answers 10 questions. Retrieved from <http://wattsupwiththat.com/2008/09/21/nsidc-s-dr-walt-meier-answers-10-questions/>
- Weiss, C.H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426-431. doi:10.2307/3109916