LUMA-GIS Thesis nr 31

Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure

Sarah S. Tressel

2014
Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Sarah S. Tressel (2014). Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure
Master degree thesis nr 31, 30/ credits in Master in Geographical Information Sciences
Department of Physical Geography and Ecosystems Science, Lund University

Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure

---

Sarah S. Tressel
Master thesis, 30 credits, in Geographical Information Sciences

Margareta Johansson
Thesis Advisor

Roger Groth
Technical Advisor

Abstract

In recent years, the Arctic has been one of the most dynamic environments on Earth. Climate indicators have measured changes in the Arctic occurring at nearly twice the rate of the rest the world during the last 100 years (IPCC, 2007). This has resulted in increased scientific inquiry into the Arctic, and beyond to the Antarctic, as both polar regions pose similar questions for scientists. Satellites have been launched, airborne missions are under way, and field expeditions have been undertaken to collect scientific data that may be used to study these areas of recent change.

On the receiving end of these data, a challenge remains as how to best manage, archive and distribute the scientific observations so that they may be easily studied, analyzed and modeled. Given the geospatial aspect of the data, working from a shared representation of geographic features in these areas facilitates the interoperability of the data. Such shared standards can be defined as Spatial Data Infrastructure (SDI). SDI provides a base upon which the data can be structured to allow for widespread use and understanding of the information. In the Arctic, shared geographic data standards, or Arctic SDI, has yet to be defined. Because of the increase of scientific inquiry in this area, defining an Arctic SDI would be beneficial.

Additionally, developing a centralized location from which to distribute the scientific observations from these areas would facilitate the research efforts underway. Such a distribution center would likely take the form of a scientific data portal or Earth browser, which would utilize the geospatial definitions identified by the SDI. This interface could service science data from the Arctic as well as science data from the Antarctic because of the unique geospatial methods for mapping at the poles. Shared development initiatives towards a data portal between the Arctic and Antarctic data management communities would result in more unified and succinct polar science. Thus, defining an Arctic SDI and sharing portal development initiatives with the Antarctic community would be of great benefit to those seeking a better understanding of the changing Arctic.

Table of Contents

1. Introduction

The Arctic is defined geographically as the area north of the Arctic Circle. It comprises the Arctic Ocean and intersects with land from eight different countries. Although it consists of varied environmental and political spaces, it represents its own unique ecosystem which has become increasingly prominent, especially in light of climate change studies. The Intergovernmental Panel on Climate Change reported that temperatures in the Arctic have been rising twice as fast as global average temperatures (IPCC, 2007), Arctic sea ice at its summer minimum is reducing by 11.5 percent each decade, in reference to the 1979 to 2000 average (Arctic Sea Ice News and Analysis, NSIDC), and the Greenland Ice Sheet set records in 2012 in reference to the satellite era for the percentage of the surface area experiencing seasonal melt, the duration of the melt period, the reflectivity of the ice surface, and the total ice loss due to melting and glacial calving (Tedesco *et al*, 2012).

In order to study and understand these environmental issues as a whole, scientists, policy makers and other researchers can benefit from the development and utilization of an Arctic Spatial Data Infrastructure (ASDI; see appendix 1 for a list of acronyms). The purpose of this would be to create shared geospatial, metadata and organizational standards for archiving scientific data related to the Arctic. These standards would allow for the facilitated sharing and understanding of the data in addition to greater interoperability across platforms and disciplines. While much work has been done towards creating an Arctic Spatial Data Infrastructure, such standards have not yet been implemented.

An important feature enabling facilitated data sharing would be the creation of a web portal from which researchers can search, access, overlay and visualize the data. This would allow for an integrated center point facilitating comprehensive analysis of the Arctic. The development of such a portal may be best undertaken by means of a joint effort between the Arctic and the Antarctic data management communities. This is because of the unique geospatial methods needed for mapping areas of high latitudes. For the North and South Polar Regions, the map projection, the geographic coordinate definition of the polar point of origin, the unique grid layout, and the prevalence of the International Date Line require special attention within a portal interface. Since these special mapping cases apply the same to either polar region, a collaborative effort would save duplication of effort and would facilitate bi-polar scientific analysis.

In the pages that follow, a discussion will be presented on the background of infrastructure studies and

SDI development efforts for the Arctic environment. The methods for implementing infrastructure in the Arctic will be considered in relation to current trends and best practices in data management and cyberinfrastructure. The aim of this study is to take these factors into consideration in order to recommend a way forward for polar science data dissemination. Relevant publications, feedback from researchers who use the data, and workshop documents resulting from discussions about science data management will be used as the information upon which to base these recommendations. The needs, goals and trends of the science community as a whole will be considered in order to propose a way forward. In the end, the unique aspects of mapping at the polar regions point the way to implementing a method of science data dissemination via a scientific data portal specific to both the north and south polar regions.

2. Background

2.1 Defining infrastructure

Infrastructure can be defined as the building blocks, systems and relationships in place to facilitate and enable a comprehensive goal. It may refer to a railway system, a telephone network or the World Wide Web. The infrastructure of a system represents the technology, structures and standards that work together to serve its overall purpose. It compiles the integral network of components into a systematic whole. Information infrastructure, or cyberinfrastructure, is a type of infrastructure mediated through computer technology. Typically, it refers to systems and services for distributing information on the internet (Bowker *et al*, 2010).

The nature of infrastructure is that it is the underlayment of a system. Its various components serve as a base from which to launch the outward activities of the specific application. This can mean that it is largely invisible to the function for which it was intended to serve, residing unnoticed in the background even though it is integral to the operation at hand. This precisely, is the role of infrastructure. In fact, well-envisioned implementations might be described as ones that ‘submerge as infrastructure’ (Star & Ruhleder, 1996).

The information scientists, Star and Ruhleder (1996) define infrastructure as having the following qualities:

- Embeddedness – infrastructure as underlayment
- Transparency – is not obvious to the general function of the application
- Reach or scope – it reaches beyond just local usage
- Learned as part of membership – newcomers learn about it through becoming familiar with its user community and their customs
- Links with conventions of practice – it is affected by (and affects) practical realities
- Embodiment of standards – it often utilizes and borrows from other infrastructures or other widely adopted tools and technologies
- Built on an installed base – it is often built upon existing infrastructure
- Becomes visible upon breakdown – it becomes obvious only once it stops working and causes a disruption in the service it normally provides

It is precisely because of these qualities that it is important when creating new systems to challenge the tendency to overlook infrastructure. Questioning the inner workings already established for a particular function will result in the development of an infrastructure better able to meet the evolving needs and goals of a system. In this study, the information infrastructure necessary for disseminating scientific data, specifically Earth science data for the polar regions, will be discussed in detail.

2.2. Spatial Data Infrastructure

Spatial Data Infrastructure refers to internet-based systems and tools in place for the distribution and use of geographic data (Bernard and Ostländer, 2007). Also inherent to the infrastructure are the standards, policies, services and relationships in place for the distribution of data. Geographic, or spatial data, is information related to space and time. A Spatial Data Infrastructure (SDI) is an initiative to seamlessly integrate a collection of spatial data (Yang *et al*, 2010a). It manages and allows the use of spatial data for a defined purpose. An SDI creates the geographic platform from which heterogeneous and distributed information can be organized for use under the specific purpose or goal.

In certain communities, SDI is more commonly referred to as Geospatial Cyberinfrastructure. Edwards *et al*. (2007) suggest that cyberinfrastructure projects tend to be more dynamic, agile and short-term focused, while infrastructure studies are based more in academia and have a broader, longer-term perspective. However for most general purposes, these two terms refer to the same basic idea. Because

SDI is the term more common to the Arctic community, this is the term that will predominate throughout this study.

The technical components of an SDI are the spatial data and the geographic information services that use or manipulate the data, while the organizational components of an SDI are the standards, policies and relationships enabling interoperability and allowing for the seamless distribution of the data (Bernard & Ostländer, 2007). In the context of SDI for Earth science data, the spatial data components are often called framework data sets. Examples of these are such as imagery layers and boundary features. In addition to this, spatially-enabled scientific data layers, such as a grid of temperature data, may be overlain upon and integrated to work in conjunction with the framework data.

Figure 1 shows a diagram showing the components and relationships thereof within an SDI. The box on the right represents the geographic and scientific data sets, in the middle are the organizational components of the SDI, and on the left is the end-use for the SDI.

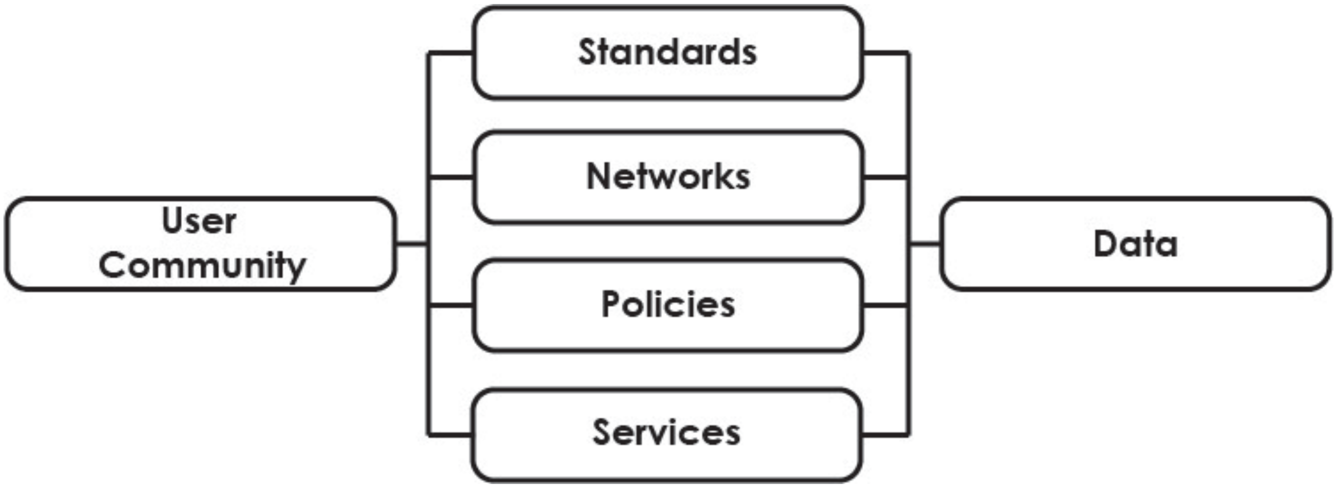


Figure 1: The various components of Spatial Data Infrastructure.

In recent years, a number of nations, states and regions across the world have developed their own SDIs. In doing so, they have created ‘a framework of standards, policies, data, procedures and technology to support the effective coordination and dissemination of spatial information across many sectors and levels of government and society’ (Sorensen *et al*, 2004). For example, the INSPIRE initiative for Europe-wide spatial data infrastructure was launched in 2007, which insures compatibility of spatial information across each of the participating countries (INSPIRE directive, 2014).

The Arctic area has yet to deploy a Spatial Data Infrastructure specific to the Arctic region. However, a group of representatives from the multi-national Arctic community have come together to develop an Arctic SDI, or ASDI, based upon international SDI standards (Sorensen *et al*, 2004). This effort is being undertaken alongside of various infrastructure and data management initiatives targeting the Earth science community and both of the polar regions (the Arctic and Antarctica). As such, the work towards an ASDI overlaps with many simultaneous efforts driving work forward regarding the adoption of standards, the creation of data sharing expectations, and the development of spatially-enabled technologies. While the scope of this study is not broad enough to detail the history of each branch of these initiatives separately, this study draws from publications, user feedback and workshop documents resulting from a number of these efforts. The purpose is to learn from the work and conversation across the Earth science community in order to put forth a set of recommendations for ASDI development that reflects the needs, goals and trends of the community as a whole.

2.2.1 History of the Arctic Spatial Data Infrastructure initiative

The groundwork for ASDI was laid by the GIT Barents project, which established SDI for the Barents Region of the Arctic. This represented a cooperation between the countries of Russia, Finland, Sweden and Norway (GIT Barents, 2006). The project helped to establish pathways for collaboration in regards to environmental, resource, land use and transportation management between these several countries at the Arctic boundary.

Later, the development of an Arctic-specific SDI was proposed at GeoNorth conferences in 2007 and 2009. Reports from several Arctic Council Working Groups also mentioned the need for developing an ASDI. Then, during the Senior Arctic Officials meeting in late 2009, the importance of ASDI development was agreed upon by each participating state, and the ASDI initiative was officially launched. The leadership for the initiative was assumed by the Conservation of Arctic Flora and Fauna (CAFF) project (Palmer *et al*, 2011).

In 2011, a meeting was held in Brussels to establish the next steps for the initiative. Since then, Arctic

SDI board meetings and actions undertaken at the International Polar Year (IPY) conference have furthered the initiative. Concurrent development of standards have been shouldered by related organizations towards ASDI, such as Sustaining Arctic Observing Networks (SAON), the International Arctic Science Committee (IASC), the Research Data Alliance (RDA), and the Arctic Data Coordination Network (ADCN). Collaboration with similar initiatives for Antarctic infrastructure have also been crucial to the recent development of Arctic infrastructure standards (Pulsifer, 2014).

2.2.2 Antarctic Spatial Data Infrastructure

Steps have also been undertaken to develop an Antarctic SDI. The Standing Committee on Antarctic Research (SCAR) is a committee dedicated to coordinating scientific research relating to Antarctica. The committee is interdisciplinary and international, serving as a comprehensive authority guiding the research undertakings across the continent and working to promote the conservation of the natural environment. It is organized into various groups and committees that advise on different elements of scientific research in Antarctica. There are groups on GeoSciences, Life Sciences and Physical Sciences. Also, there are committees on data management and geographic information. Lastly, there are committees on finance and Antarctic treaties. Each group is tasked with ensuring that collaboration and coordination exist across disciplines and enabling high quality, consistent and safe scientific undertakings (About SCAR, 2013).

In regards to the data and information management components to SCAR, these efforts were initiated in the 1980's following a stated need for better compatibility and accessibility of scientific data related to Antarctica. A committee to address this need was formed in 1989 under the title of the Committee on the Coordination of Antarctic Data (CCAD). In 1992, a joint SCAR and COMNAP (Council of Managers of National Antarctic Programs) initiative was formed to carry this goal further. Eventually, this turned into the Joint SCAR/COMNAP Committee on Antarctic Data Management (JCADM) which proposed and implemented standards for Antarctic data management. In 2009 after COMNAP withdrew from the joint committee, the current Standing Committee on Antarctic Data Management (SCADM) was formed. The goal to share scientific observations freely and effectively is afforded by recruiting Antarctic data centers into its system and by promoting the use of the Antarctic Master Directory (AMD), which is a part of NASA's Global Change Master Directory (GCMD), which is discussed later in this study.

In 2006, the Standing Committee on Antarctic Geographic Information (SC-AGI) was formed following a SCAR meeting in Hobart, New Zealand. This committee's goal has been to develop and manage Antarctic Spatial Data Infrastructure (AntSDI). The components of the AntSDI are spatial data sets (topography and bathymetry maps, management zones, maps of protected areas, etc.) in addition to the platforms for easily sharing these data. Another important component is the adoption of standards for spatial reference systems, datums, and other mapping standards. Also stated is a goal to maintain a regularly updated and complete map of the Antarctic area. Lastly, a crucial element of the AntSDI is the deployment of a system to allow easy access to the data with meaningful data services for scientific visualization and analysis (Finney, 2009).

These goals are yet to be fully realized. However, traction has been gained of late towards the development of AntSDI services. For example, in response to the stated need for a regularly updated and complete map of the entire Antarctic area, the Polar Geospatial Center at the University of Minnesota has taken a big step forward. Recently, the center published a web-based map of Antarctica using imagery at a resolution 900 times better than the previous leading Antarctic-wide map. This imagery also has the ability to be updated every 45 days (Maxwell, 2014).

It may be important to note however, the Polar Geospatial Center offers data and services for both poles, not just Antarctica. Centers such as this one will likely be the way forward for developing SDI framework data and GIS services applicable for both poles.

2.3 Understanding the Arctic environment

The Arctic is defined geographically as the area north of the Arctic Circle (approximately 65.5 degrees north). However, the environmental definitions (tree line, average temperature, or permafrost extent) do not heed any strict boundary. Ocean circulations bring cool Arctic water further south down the Greenland coastline and the Canadian archipelago, while the Gulf Stream warms the Norwegian coastline northwards towards Svalbard. On land, the definition of Arctic tundra is a treeless environment. However, areas near the Arctic boundary include boreal forest, while areas of high elevation south of the Arctic resemble tundra. The thickness of the permafrost layer is greater on north-facing slopes than it is on south-facing slopes. Also, the topography of the landforms determines where snow drifts or ice flows, complicating the definition yet further (Callaghan *et al*, 2005).

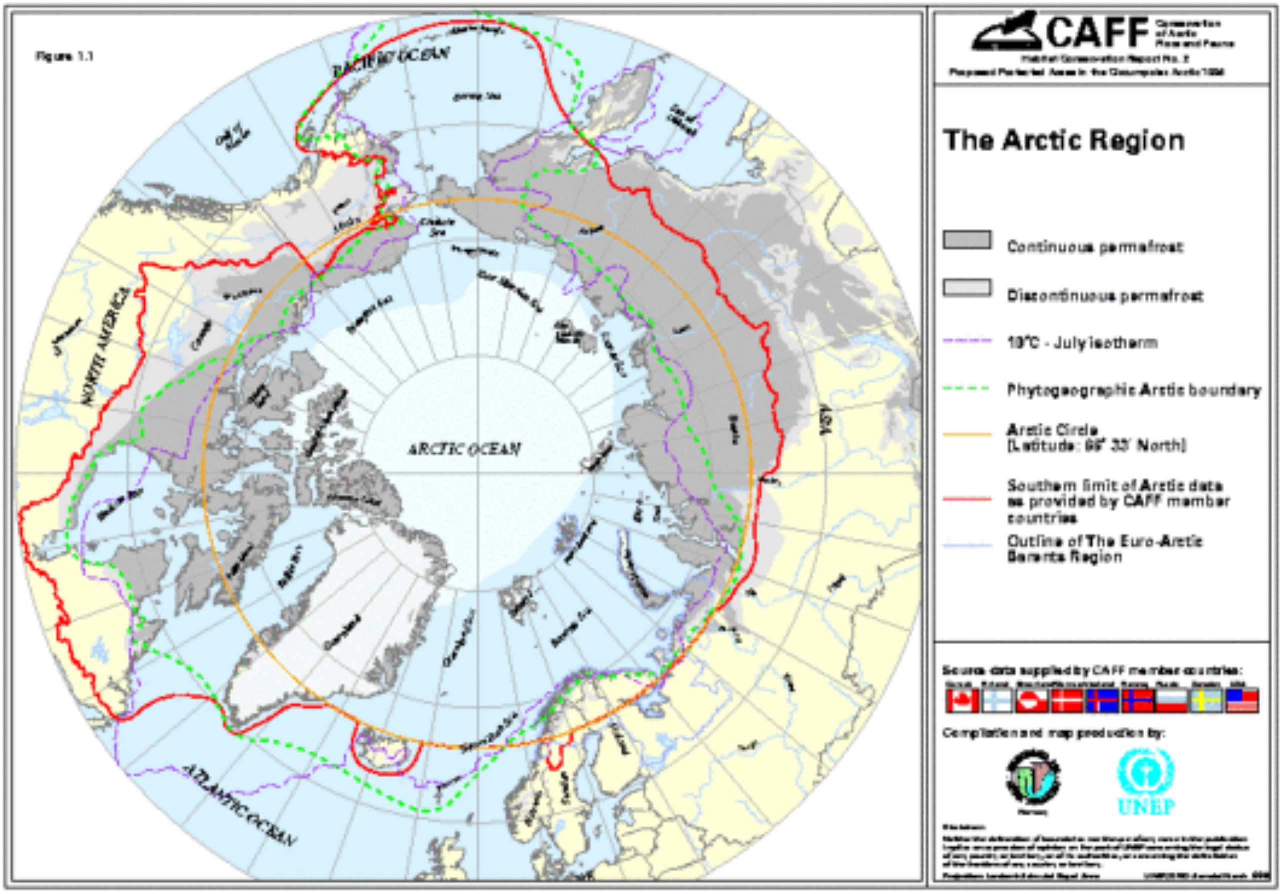


Figure 2: The image above shows the geographic boundary of the Arctic Circle (yellow), the temperature boundary of the Arctic (purple), the tree line (green), the areas of continuous permafrost (dark gray), and the boundary of Arctic data provided by the Conservation of Arctic Flora and Fauna (red) (UNEP/GRID-Arendal. *The Arctic region – definitions*, 2006).

Of primary importance is the observation that the Arctic is currently undergoing unprecedented change due to a warming climate, and these changes may only accelerate in the near future. Key indicators that signal this change are currently being observed across various scientific disciplines. Examples of these changes are as follows:

- Changes in the atmosphere
    - Rising average air temperature

    - Increase in precipitation
    - Increase in greenhouse gases
    - Increase in ultraviolet radiation

- Changes in the cryosphere
    - Diminishing Arctic sea ice
    - Lesser snow cover
    - Thawing of frozen ground
    - Melting glacial ice

- Changes in the oceans
    - Changes in ocean salinity
    - Rising sea level

- Changes in the biosphere
    - A greening of the northern environment
    - Alterations in animal migrations

(Satellite Observations of Arctic Change, NSIDC, 2014; ACIA, 2005; NASA Science, 2013)

Further, the changes in the Arctic environment have brought about an increased interest for industry and travel in the Arctic areas. Lesser sea ice coverage during the summer season creates the potential for increased shipping along the Northern Sea Route and the Northwest Passage. The decrease in summer sea ice has also encouraged activity related to oil and natural gas production (Satellite Observations of Arctic Change, NSIDC, 2014). This activity has the potential to cause yet further alterations to the fragile and already quickly changing Arctic environment.

Given the complexity of the Arctic ecosystem and the changes thus far observed, it is especially important for scientists, policy makers, and industry researchers to have access to information across scientific disciplines in order to identify new trends, make connections between indicators and respond in real time to the changes occurring. Without standards for infrastructure or a venue through which to share scientific observations, the accessibility of environmental information related to the Arctic will continue to be limited.

‘There is a clear need – particularly in light of today's climate change scenarios – to be able to access, integrate and use spatial data from disparate sources in guiding decision making. Our ability then, to make sound decisions collectively at the local, regional, and global levels, is dependent on the implementation of a sound infrastructure that provides compatibility across jurisdictions and promotes data access and use’ (Sorensen *et al*, 2004).

2.4 Informatics as its own science

The integration of heterogeneous information is a science in itself. Computer science is the study of the processing of information in a computer system. Information science, or informatics as it is often called, encompasses the definition of computer science, but can be opened up to include the technological integration of data, information, and services with the reception of the end user (Indiana University, 2013). Within the Earth science community of late, informatics and data management have been recognized as having an influential position alongside scientific research and its data. This is because offering services for organization, sharing, integration, interoperability, documentation, and accessibility of the data adds value to the process of science in its own way.

Widely-shared scientific data that are easy to integrate with other disparate data sets can pave the way for new discoveries in the field of science. Well-managed scientific data provide the reference and resources needed for future scientific inquiry and can lead towards interdisciplinary and collaborative science (Finney, 2009). Further, having well-managed, integrated data sets removes the overhead from scientists who may otherwise have to locate, reformat and reprocess data before they are ready for use. Offering data along with a suite of services simplifies the process for those who have scientific expertise, but may not excel in data management and computer science (Fox & Hendler, 2009).

Data relating to the Arctic environment can be especially complex, and thus, have much to gain from meaningful informatics systems. The sharing and transfer of information relating to the Arctic is crucial in light of its changing climate.

2.5 Flexible infrastructure and user-centered design

Technology and the total volume of data have the ability to change and grow very rapidly. The human reception to these advancements, however, may or may not be able to match the pace of the technology itself. In light of this, it is important to not forget about the human aspect of creating data infrastructure. Technology is only useful if it is applicable to the end user. Defining exactly what this means is a daunting task, especially when spearheading a multi-national effort towards defining a set of standards and services for ASDI. Mitigating this issue would include gathering input from and collaborating with representatives from the user community. For the Arctic, this community consists of polar scientists, information scientists, governmental policymakers, industry analysts, indigenous peoples, and citizen scientists. The ASDI initiative has thus far included bottom-up as well as top-down design methodologies to account for this diverse user community (Pulsifer, 2014). All have much to gain from integrating and sharing Arctic environmental data, and all have much to offer to the data, structure and organization of an ASDI.

In light of the need for user-centered design, successful infrastructure implementation will be transferable to multiple use-case scenarios, understandable for both specialists and the uninitiated, and will have the built-in flexibility to adjust in order to extended uses and applications. These ideas implicate a paradox in that the technology must be adaptable, but also be structured and standardized. ‘The competing requirements of openness and malleability, coupled with the structure and navigability, create a fascinating design challenge – even a new science. The emergence of an infrastructure... is thus an 'organic' one, evolving in response to the community evolution and adoption of infrastructure as natural’ (Star & Ruhleder, 1996). While there is no easy solution to this conundrum, a key to implementing a successful infrastructure will depend upon an invested end-user community willing to help guide the design down the path that allows for both functionality and usability.

Suggestions for how to implement this come from Bernard & Ostländer (2007) who recommend organizing feedback and acquiring metrics in order to harness stakeholders' knowledge. Also, using interactive assessment to allow for participatory development will allow for a built-in flexibility of the tools.

Abbot (2009) describes a picture of what a more integrated science network may look like: ‘Instead of well-defined data networks and factories coupled with an individually based publishing system that relies on peer review and tenure, this new research enterprise will be more unruly and less predictable, resembling an ecosystem in its approach to knowledge discovery. That is, it will include loose networks of potential services, rapid innovation at the edges, and a much closer partnership between those who create knowledge and those who use it. As with every ecosystem, emergent (and sometimes unpredictable) behavior will be a dominant factor.'

Recent conversations initiated by the National Science Foundation's Cyberinfrastructure Program mentioned the role of the citizen scientist in regards to both providing and using data. 'General public, including citizen scientists... represent an increasingly important group, as both data publishers and data consumers, and have specific challenges, in particular related to data quality, transparency of access and use of data, motivation and governance' (Altinas *et al*, 2012). Also, the suggestion was made to have the end user involved in developing software tools for data analysis, such that the software can be guided in the direction of what the scientists need rather than what the developers might conceive that they need (EarthCube executive summaries, 2013).

It should be noted that community-driven development for ASDI may differ from that for Antarctica. Antarctica benefits from a far less complex geopolitical scene. Political and territorial concerns amongst the countries intersecting with the Arctic have caused a delay in the development of ASDI. Each country also has a somewhat different approach to what SDI standards should be in place for Arctic spatial data. The differing opinions have led to various countries developing their own national approach to mapping the Arctic rather than finding a collaborative approach with their international partners (Taylor, 2014). Although collaboration between the Arctic and Antarctica are encouraged later in this study, considerations for the diverse user community for Arctic Earth science data may necessitate divergence from or extensions upon such collaborated efforts.

3. Methodology

Given the benefit that exists in entertaining end-user input at the beginning of the design process for infrastructure standards, it makes sense to learn from this for ASDI development. The Earth science community has recently put forth papers, workshops and reports recommending a way forward for the handling of data and critiquing data management problems that exist today. This study gives an overview of the common themes stated in these publications and pieces together the best practices suggested for a way forward.

4. Results

4.1. Towards discoverable and interoperable data

The ability to easily discover and access data sets from various sources is a primary goal for the Earth science community. A vast array of online scientific data sets exist, yet are archived at distributed data centers in various formats using differing standards. The overhead required for a research scientist to gather together the relevant data for her or his area of interest, and then be able to compare each data set to another can result in months of time and effort. To give an example of the volume of Arctic data sets, an experiment was performed by Li and others (2011) in which an active Web crawler was developed to crawl for data sets. Specifically, it searched for data sets enabled with WMS (Web Map Service). Over 12,000 data sets were found that relate to the Arctic. The standardization and the development of services for these existing data sets could greatly improve the workflow for scientists using the data.

To accommodate for this, the following are suggested components for a robust cyberinfrastructure.

- Data are collected and maintained effectively by distributed data centers
- Standardized metadata is created and maintained
- Digital Object Identifiers (DOI) for data set citations are created and maintained
- Data are made publicly and freely available for download
- Standard vocabularies and ontologies are appropriately associated to assist with data discovery
- Standards are applied for geospatial catalog services
- Spatial extent and resolution parameters are enabled for search and visualization services
- Projection and grid definitions are appropriately recorded
- Data visualization services are enabled in relation to portal services
- Project-on-the-fly capabilities are enabled in portal services
- Data interoperability services are enabled in portal services

(Sorensen *et al*, 2004; Li *et al*, 2011; Bernard & Ostländer, 2007; Yang *et al*, 2010a; Fox & Hendler, 2009; Altinas *et al*, 2012)

The goal is for these standards and services to work together to allow heterogeneous, distributed data sets firstly, to be discoverable, and secondly, to be interoperable with one another. Further, creating a single location from which researchers can search for and access disparate data sets for a particular geographic area, or even world-wide, would advance these efforts. This could take the form of a data portal in which data can be located and visualized in an Earth-browser map environment. Recently, several of the prominent science organizations, scientific working groups, and independent researchers have been compiling recommendations towards these goals. Several of these are noted below.

4.1.1 Metadata standards

There exists a common saying in information technology circles that one's data is only as good as the metadata behind it. In regards to data management protocol, this is also the case. The relative richness of the metadata associated with a given data set will determine how often and appropriately it appears in a listing of search results, whether or not mapping services will be able to be applied to it, or how easily data access options are enabled for it. Many of the services described further below in this study are dependent upon sufficient and standardized metadata. Thus, it remains a crucial element in data discovery and access. Various international database standards exist for notation and structure. A broadly accepted standard would be best suited for use in a scientific data repository.

In regards to metadata standards for Earth science data, the Global Change Master Directory (GCMD) provides a common format and a directory listing of data sets. The GCMD was developed by NASA and exists today as one of the largest public metadata repositories in the world. The metadata structure for GCMD takes the form of the Directory Interchange Format (DIF) standard notated in XML schema, and has been cross-mapped with other standards in order to synchronize the content.

In addition, the GCMD offers seven sets of controlled keyword vocabulary fields within their metadata format to enhance search capabilities and to provide meaningful groupings and categorization of data sets. This provides a set vocabulary or ontology “to disambiguate between different terminologies by explicitly defining a conceptualization” (Li *et al,* 2011). Some of these are implemented in a hierarchical format to allow search and categorization functionality to operate at the level of detail desired (GCMD, 2014).

For example, one of the vocabularies offered provides a controlled listing of scientific keywords, which describe the geophysical parameter of a given data set. This vocabulary is structured in a hierarchical format. One such hierarchy is as follows:

`Land Surface > Topography > Terrain Elevation`

The various levels of the hierarchy allow computer systems to use the information at the level most appropriate for the given function. If the function is simply to differentiate land data sets from ocean data sets, the top-level category is most appropriate, and most easily implemented. However, if the function is to differentiate elevation data sets from topographic relief data sets, the lowest level should be targeted. Having a defined vocabulary for geophysical parameters is especially important, as search metrics from one scientific data center have shown that data users tend most often to use these kinds of terms (snow cover, etc.) when searching for data (NSIDC, 2013).

4.1.2 Ontologies

As is described in section 4.1.1, the GCMD keyword vocabularies provide controlled listings of scientific terminologies to standardize the descriptions of the data sets within the metadata records. However, such vocabularies can also exist apart from the metadata, serving to group related terminology together to allow for better integration of data sets. These are commonly referred to as ontologies and can be defined as the ‘machine encodings of terms, concepts and relations among them’ (Fox & Hendler, 2009). Ontologies enhance with semantics-based search functionality, and enable meaningful associations between science terms noted across data registries. For example, ‘rainfall’ and ‘precipitation’ could be noted as synonyms in such an ontology (Hunt *et al*, 2009). An example of a recommended ontology system, which could be implemented for data catalogs is the Simple Knowledge Organization System (SKOS). It provides a controlled vocabulary of classifications (Altinas *et al*, 2012).

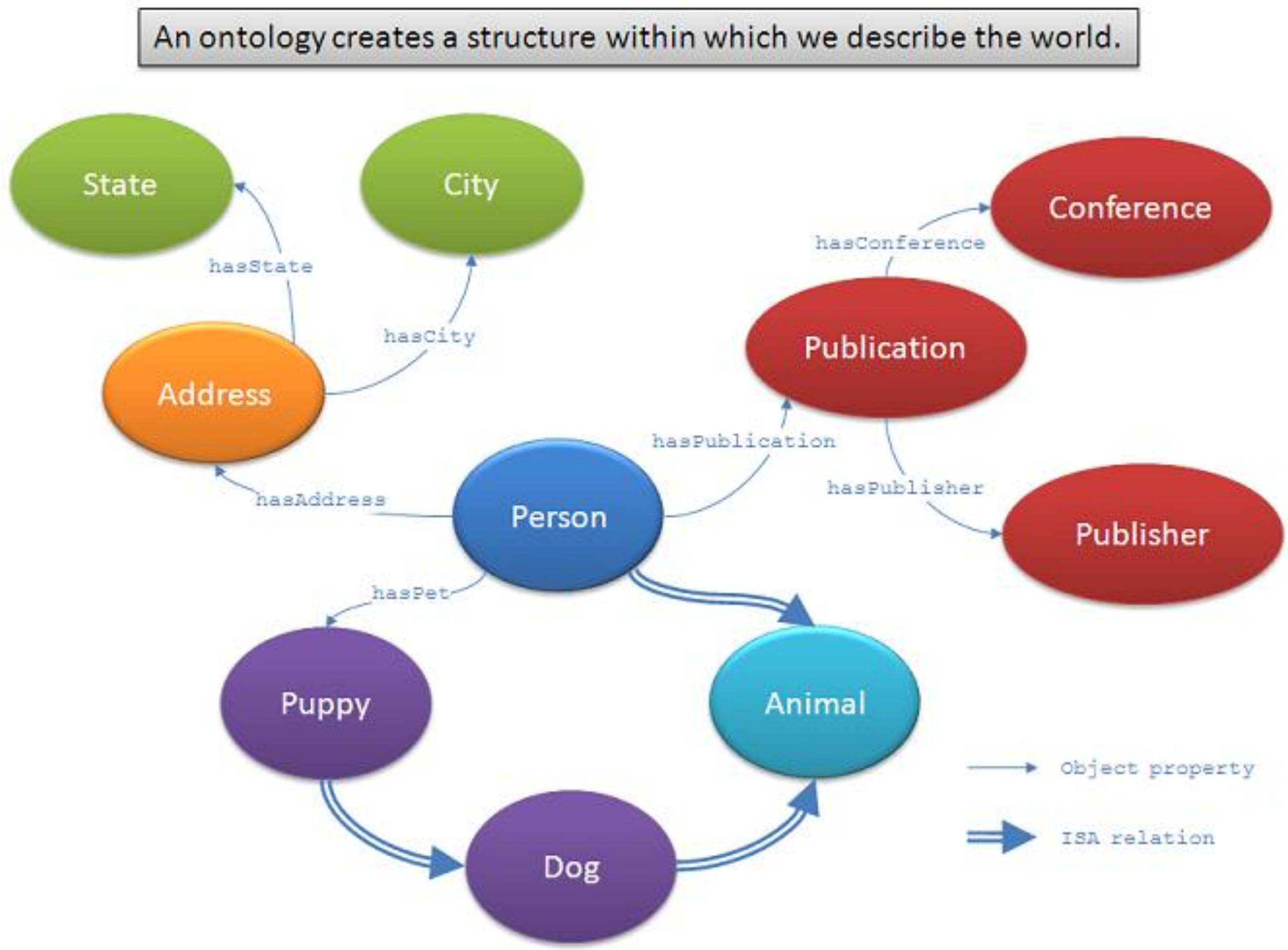


Figure 3: The image above shows an example of an ontology. An ‘ISA’ (or is a) relation indicates that the given word belongs to the class of the word that it points to, such as: a puppy is a dog. The ‘has-‘ relations above indicate that the given word has the property of ownership over the word that it points to, such as: a publication has a publisher (Data Annotation, NEMO, 2010).

As is mapped in Figure 3, an ontology can store a vocabulary of words along with the relationships between words. This adds a useful functionality for the dissemination of scientific data sets.

4.1.3 Catalog Services

Catalog services allow for a metadata listing to be queriable in a standardized way. This is important for data set search and retrieval operations within an SDI. Open Geospatial Consortium (OGC) Catalog services for the Web (CSW) is one recommended service for scientific data set listings. This has become a primary catalog service for scientific data and allows for the support of multiple users accessing information from distributed data archives (Li *et al*, 2011). An advanced capability of an SDI as described by Li and others (2011) is to chain services together, such that if one such service is not available for certain data, another is attempted. This can provide a seamless experience for the user of an SDI when perusing scientific data sets.

4.1.4 Open Geospatial Consortium standards

In general, using Open Geospatial Consortium (OGC) standards are often the recommended practice for the development of data access services. These standards represent the outcome of a consensus-based development practice that was influenced by an international collaboration of geospatial software experts, academic professionals as well as end-users (Bernard & Ostländer, 2007). A recent effort to develop a series of Arctic science data maps, virtual globes and services exemplifies why OGC open source services may be preferred over commercial software packages.

The Arctic Research Mapping Application (ARMAP) was developed to provide access to disparate scientific data for the Arctic. The groundwork performed for the project, including its database, framework and services, serve to enhance the efforts of Arctic SDI development. At one point in the project, it was determined that ESRI software products would be utilized to implement some of the desired services. A reason behind choosing commercial software over open source software was that the student developers working on the project had access to extensive training websites, which would reduce the overall effort to spin up the technology (Johnson *et al*, 2011). However, in light of open source technologies and standards that are typically adopted by the science community, the ARMAP effort could be determined less relevant and lacking in interoperability due to the use of the proprietary software. Presently, it is of increasing importance to lay the proper foundations of sustainable, buildable, and extendable infrastructure, as a great wave of data availability is forthcoming.

4.2 Data access paradigm shift

Earth science today generates a vast amount of data. Earth-observing satellites constantly monitor the surface of the planet with a variety of instruments. Airborne and in-situ platforms add to the collection of data. New data acquisition techniques have increased the resolution, and therefore the volume, of the data (Yang *et al*, 2010a). In addition, modelers create new ways to combine and make future projections for the data at hand. The current trend is leaning in the direction of further complexity and larger and larger data sets.

In order to analyze the data, many scientists have been pushed to become well-versed in computer science and may need to have high-performance software packages that can manipulate vast amounts of data effectively and efficiently. With the growing demand for extensive disk space and complex software to do science, this paradigm of scientific data analysis at the desktop level becomes less feasible. Instead, it may be more efficient for the scientists to bring their analysis to where the data are stored online, rather than bring the data down to the level of the individual scientist (Gray, 2009).

With new development in online data visualization, it is now becoming more and more possible to analyze data sets on the fly instead of downloading large amounts of data. 'This performance improvement contributes to a 21st century paradigm shift in data access – from media shipping to on the fly, online access' (Yang *et al*, 2010a). While this could provide rapid access to data and the democratization of scientific analysis, special consideration will need to be paid to the methodology behind the analytical services provided for the data. It is possible for minor calibrations of data to result in erroneous results. Therefore, any interpolation, averaging, smoothing, omissions, etc. in the processing would need to be transparent and well-documented. It also may be that some data may have complexity that is unable to be sufficiently addressed in the online analysis environment, and therefore would require handling by other software.

The trend towards greater data availability is also changing the culture of the way science is done. Scientists who once felt the need to shelter their data prior to publication are now realizing that the increasing amounts of available data are not necessarily feeding rivals ideas, but rather are creating possibilities for data manipulation and use beyond what may have been imagined prior to widespread data sharing practices. For example, below is an account from a scientist reflecting on current trends in data availability.

'I used to be worried that… people would publish before me if I put my data out there too early, but now I have access to such a wide range of data, in real or near real-time, that there is more data to interpret than any of us can reasonably keep up with. The sheer volume and variety of the data has even provided some of us with new fields of enquiry focused purely on mining the data. Of course I can see now that without the foresight that was put into standardizing and specifying many aspects of the system, like data, metadata and communication standards, none of what we now enjoy would have been possible. We might still be capturing these large volumes of data – but sharing in its use and using it efficiently like we do today just wouldn't have happened' (Finney, 2009).

Organizations that fund scientific research, such as the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) in the USA, adhere to policies that promote the open sharing of data and the publishing of data sets collected for the research they fund. Further, the Antarctic Treaty includes an article that requires the free exchange of data collected for research on the continent. The commitment to public accessibility for scientific data exemplifies the understanding that the data collected today will provide the foundation for the science of tomorrow (Tressel, Bauer & Scambos, 2013).

The International Polar Year campaign of 2007 and 2008 launched a network of polar observation initiatives which could result in a system of long-term data collection and archive. The hope is that these data will, over the decades, enable the 'establishment of a baseline against which to detect and forecast future environmental, biological and climate change' (Finney, 2009).

4.3 Data portals

A primary venue in which the shift to online scientific data analysis will take place is within the context of a data portal. Such portals are akin to Earth browsers or virtual globes, for which spatially-enabled scientific data are displayed on top of framework or base map layers. The portals that exist today have typically been launched to serve segments of the scientific community, and tend to have a scope limited to a particular discipline or a certain funding organization. The science community has expressed the goal of developing a single portal environment, which allows access to the breadth of scientific data across disciplines and data centers (Altinas *et al*, 2012). However, such a single repository for this does not yet exist.

In recent discussions, the need for interdisciplinary science is emerging as a common theme. Scientists have expressed interest in a method by which to study the Earth as a single system. There seems to be a 'growing understanding of environmental interconnectedness' alongside of advancements in technology that facilitate the interoperability of geospatial and science data (Altinas *et al*, 2012). Taking advantage of this are ‘new Renaissance-type scientists’ who are comfortable using data and analysis outside of their normal fields of expertise (Altinas *et al*, 2012). Through all this, the need for the cross-pollination of data between scientific disciplines is gaining recognition, and a single portal in which different sources of science data can be integrated would aid in this. ‘A new breed of scientist is emerging. These scientists exhibit knowledge of their particular discipline in addition to expertise in the areas of computer science and mathematics’ (Tedesco, 2013).

The portal environment takes full advantage of the framework data sets mentioned previously in this study. These layers serve as foundational data sets that provide points of reference for the spatially-enabled science data sets which are overlain upon this base. There is no clear distinction between which data sets are chosen as base layers and which ones are overlain. Scientific data portals tend to show base layers customized for the particular branch of science for which the portal was built. However, a common trend is for portals to provide base layers that are more static than dynamic as well as layers that provide a general context for the layers that may be overlain. Examples of these types of foundational data sets are imagery, coastlines, administrative boundaries, digital elevation models, bathymetry, protected zones, etc. Science data sets intended to be visualized in context of these base layers might include air temperature, sea surface temperature, atmospheric properties, snow cover, sea ice, soil moisture, etc. These data sets may be gridded satellite data or in-situ point data, current or historical, as long as there is a geographic aspect that ties the data to a certain spot on the Earth.

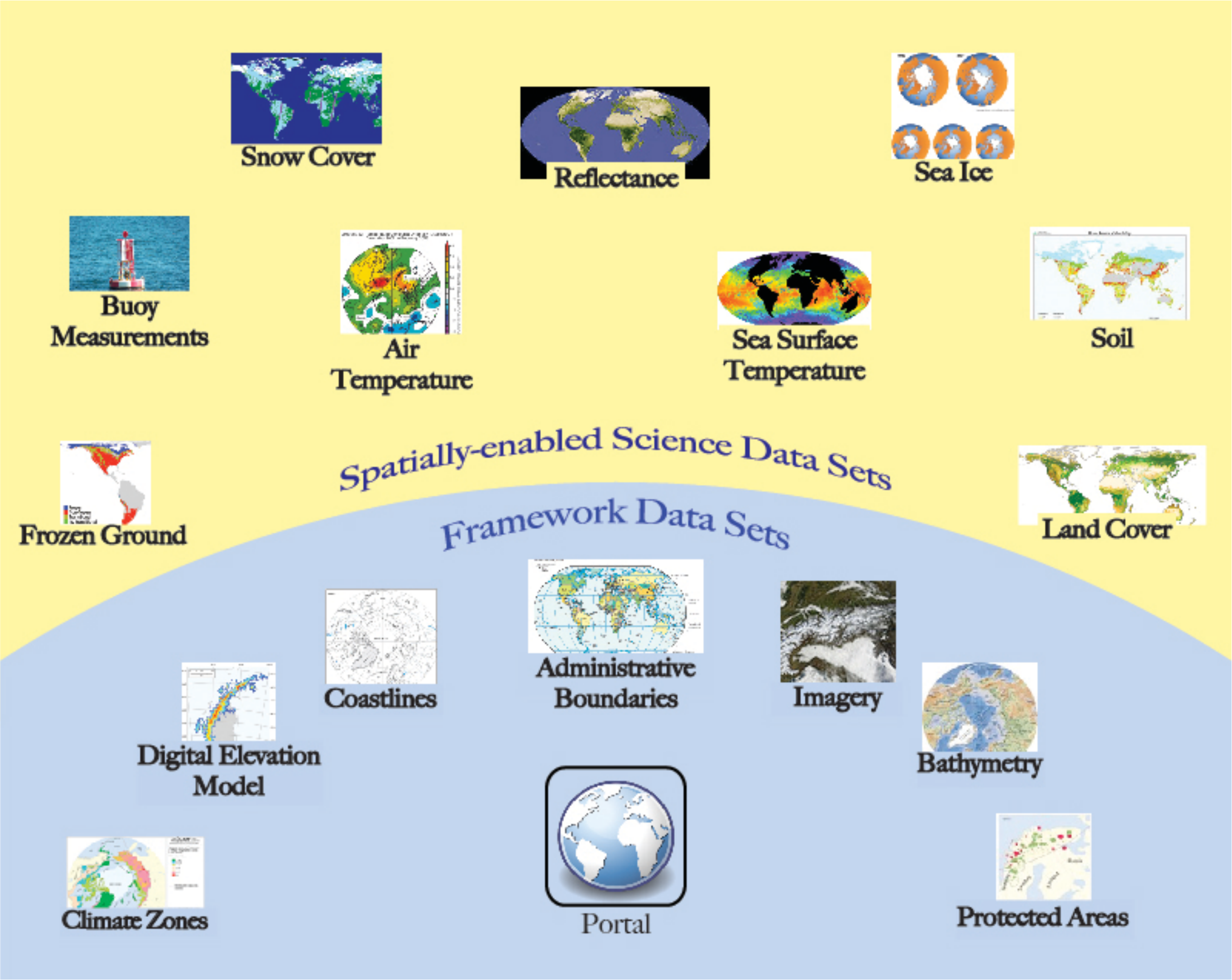


Figure 4: Examples of framework (base) map layers are shown in the lower section, while examples of science data overlays are shown above (see appendix 2 for references).

4.3.1 A single repository for metadata

The services of a data portal will be most effectively and efficiently realized with the implementation of a centralized metadata catalog (Sorensen *et al*, 2004). Keeping a single repository of the metadata means that the search and services can act more quickly and consistently within the portal environment. The metadata records are originally created and maintained by the distributed data center, which holds the associated data. To create the centralized repository, the portal will then have to harvest the metadata on a regular basis in order to have the most complete and recent data descriptions ready to be used for portal services. Depending on the frequency of the harvest, however, this could mean that the portal will not have the most up-to-date records (Sorensen *et al*, 2004).

Also, the process of harvesting metadata may take time, and it can be difficult to find an acceptable point in the day to perform the harvesting, as people from all over the world in many different time zones may be interested in accessing the data. For example, a metadata harvest scheduled to take place at midnight on the west coast of the United States, may impact European data users at the point at which they have just arrived at work.

Although the practice of having one centralized repository is recommended for portal services (Sorensen *et al*, 2004), it is important to note that this is not the only way forward. It is also possible that metadata be retrieved upon each service call in order to populate the required information fields.

4.3.2 Visualization and interoperability techniques

There are a number of visualization techniques that may range from static thumbnail images, to dynamically-generated map services, to interactive graphics and charts (Altinas *et al*, 2012). High spatial and temporal resolution data can pose visualization challenges, which may require for data to be simplified or compressed before the visual display is available. 'Visualization requires the consideration of spatial principles to solve spatial complexity issues' (Yang *et al*, 2010a).

The Open Geospatial Consortium (OCG) standards include the widely adopted Web Map Service (WMS), Web Coverage Service (WCS) and Web Feature Service (WFS). These services allow data in the form of maps, features and images to be delivered through a web service. WMS calls forth a simple map image via HTTP from a geospatially-enabled database. WCS renders multidimensional data, delivering the data together with their original attributes, allowing a rich, analyzable map. A WFS service call returns specific geographic features, allowing for the retrieval at the feature level rather than the file level (Open Geospatial Consortium, 2014). WMS is the most popular of these services, and is enabled for a large amount of scientific data. These services are recommended across many publications to do the job of enabling data visualization and geospatial interoperability in an SDI portal (Li *et al,* 2011; Altinas *et al*, 2012; Yang *et al*, 2010a; Sorensen *et al*, 2004).

As these services advance, questions that will need to be answered include how to integrate data sets at different resolutions, how to interpolate values for missing data, and how to integrate the fourth dimension of time into the portal functionality (Yang *et al*, 2010a).

4.3.3 Quality flag

Data sets that are accessible and discoverable may not, however, be useful to the scientist searching for data. It has been suggested that a quality of metadata flag or quality of service flag be inserted into the data set description in order to help users dig through the diverse data sets (Yang *et al*, 2010b). While the first challenge to be addressed is the ability to find and easily access the data, the search may not end there. The next challenge is to determine if the data are suitable for the purpose at hand and high enough quality for the desired level of analysis. This can be especially difficult if the scientist is searching for and using data outside of his or her field of expertise (Altinas *et al*, 2012).

Possible solutions to this could be a ranking system based upon when the data were last updated, assuming that more recently updated data will be more relevant and error-free than other data. Another solution could be a ranking system based upon the number of times a data set is cited in a publication, assuming that frequently cited data will be of better quality than uncited data. The assumption is that data of high quality will be more likely to be used for scientific analysis and more often cited in academic publications. Finally, there could be a peer review of the data available within the data set description (EarthCube executive summaries, 2013).

4.4 Interdisciplinary nature of GIS

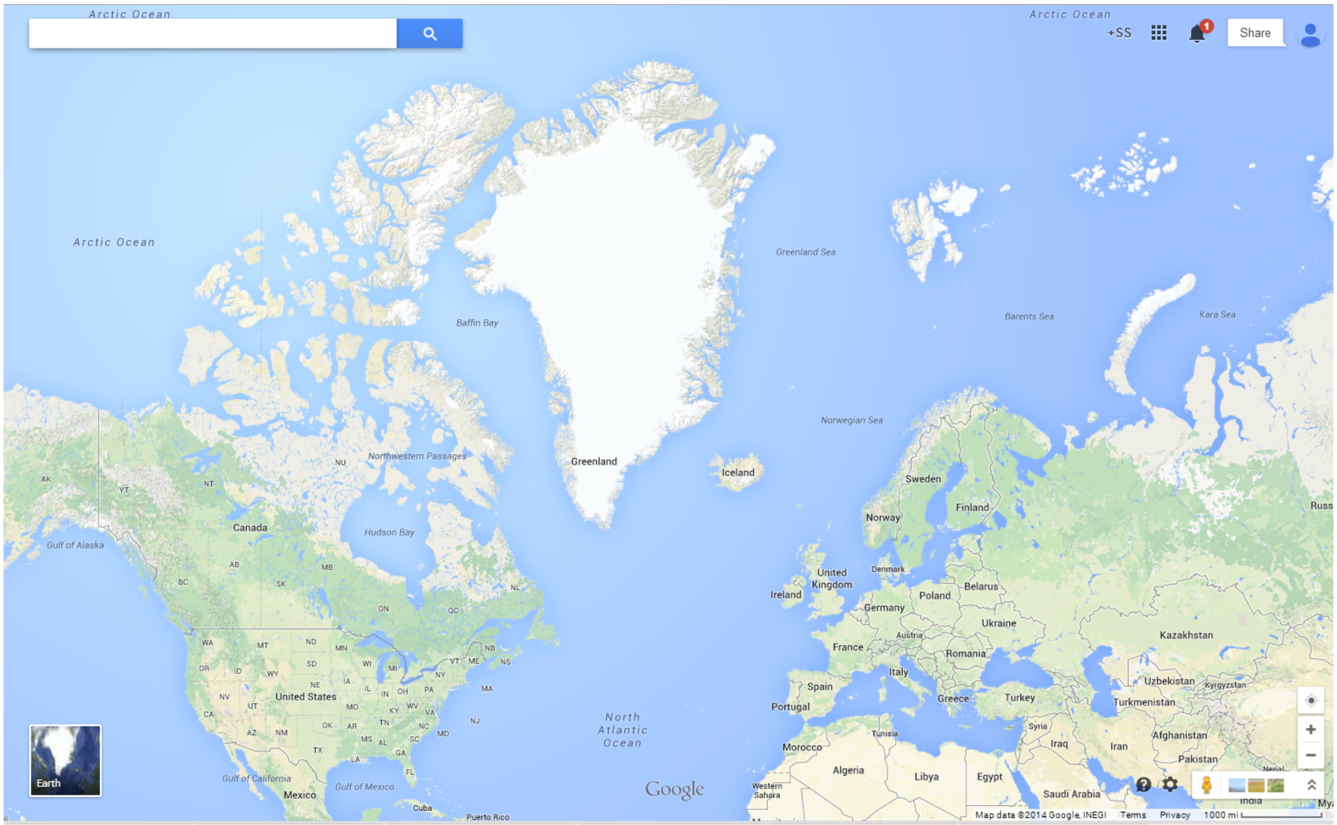
A GIS-enabled Earth browser for scientific data has many possible uses and extensions. It represents the coming together of geospatial information with other quantitative and qualitative information about the world. In a single image, scientific data sets can be overlain upon a map in order to identify trends in geographic areas. Connections can be made between disparate sources of information. Models can be generated that include boundaries, buffers and interpolation. A GIS allows the context for scientists to portray and analyze the data they gather about the Earth. However, the methods by which most Earth browsers are geared to handle scientific data, tend to fall short when mapping areas at the polar regions of the world.

4.5 Unique geospatial methods for mapping at the poles

Human population density shows preference for the mid-latitude regions of the Earth. Because of this, it is not unusual that the way we typically portray the globe on a two-dimensional surface is by centering the focus on the mid-latitude belt of the world. Historically, early cartographers placed their region of origin near the center of the map and drew the other geographical features as extending out from their conceptual axis. An outgrowth of this tendency is the dominance of a world map view similar to what is shown in Figure 5.

In Figures 5 and 6, two screen shots are presented: one from Google Maps (Figure 5) and the other from Google Earth (Figure 6). Both center on Greenland. Google Maps displays its imagery based upon a variant of the Mercator projection (Google Maps projection, 2014). The Mercator projection is a cylindrical projection tangent to the equator. As is true for any method for representing the Earth in two dimensions, the Mercator projection results in some areas of distortion. In general, creating a two-dimensional map requires a reinterpretation of an ellipsoid onto a plane. The result is a map of the Earth with a degree of error introduced in conformality, distance, direction, scale, and/or area (Dana, 1999).

In the case shown in the Google Maps image in Figure 5, the geographic features near to the equator appear true to their actual form, while features near the poles, such as the northern part of Greenland, appear exaggerated in size. A common trend in scientific websites, or websites in general, is to utilize Google Maps Application Programming Interfaces (APIs) to implement mapping services within the site. Thus, the image similar to Figure 5 can be seen in many scientific mapping applications. This works for mid-latitude geographic features, but is not suitable for features at the poles. A much more accurate representation of the polar regions appears in the Google Earth image (Figure 6). However this image does not show all areas of the Earth in one view, and thus requires a much more complex approach for online mapping applications.



Figures 5: This Google Maps image shows distortions in the high Arctic (Google Maps, 2014).

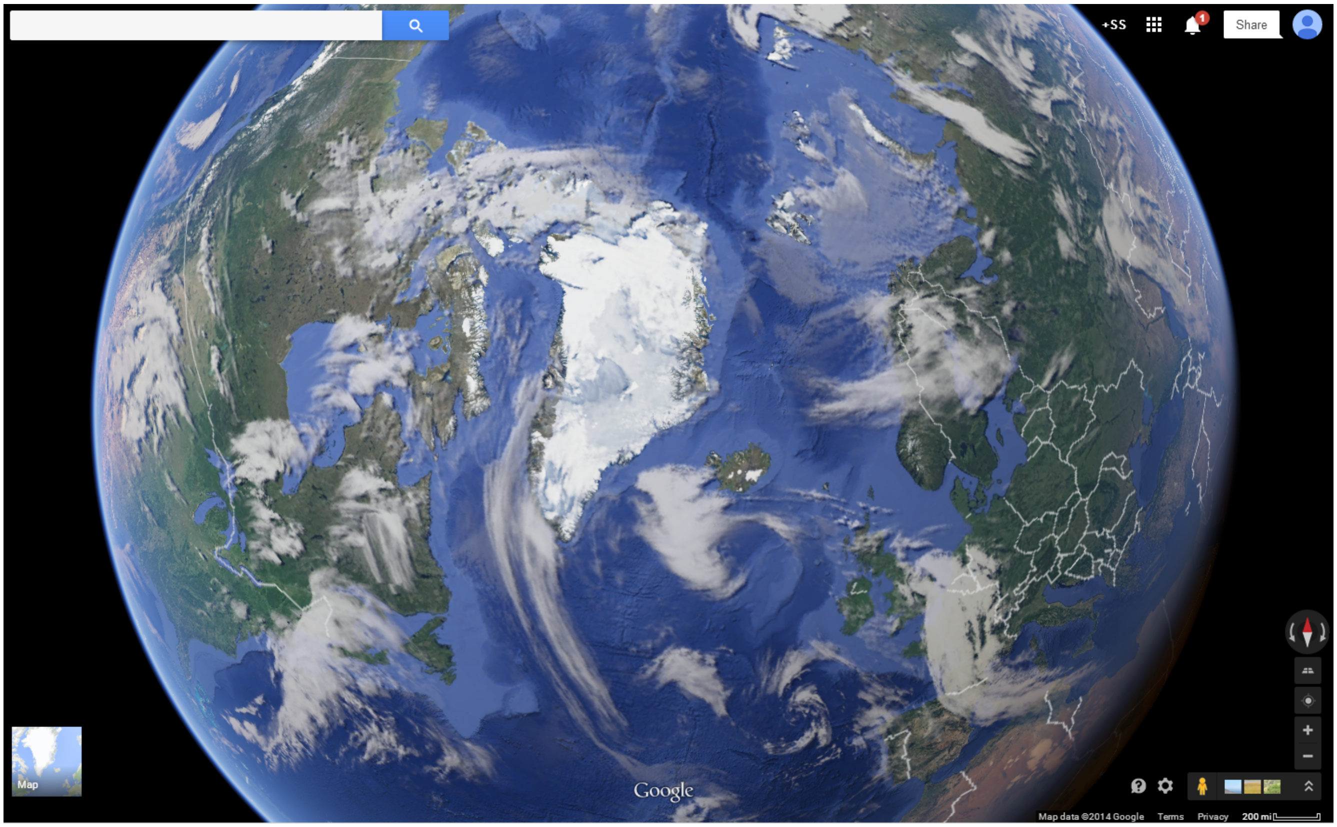


Figure 6: This Google Earth snapshot shows Greenland with more natural proportions when compared to the Google Maps image in Figure 5 (Google Earth, 2014).

4.5.1 Polar projections

In order to create a reference system that represents the polar areas of the Earth as a contiguous area with little distortion, a reconfiguration of the map is required. For this purpose, polar map projections are used. Polar projections result in maps with the North or South Pole taking the focal point at the center. The lines of latitude are shown as concentric circles. The outermost circle, representing the edge of the map, is typically the equator. Thus, the map shows only one hemisphere, either north or south. The area of the map with the greatest distortion is the area at the edge of the map. In the polar areas, the accuracy of the features are preserved (Van Sickle, 2004).

There exist a couple of polar projections that are commonly used with polar Earth science data. The first is the Stereographic projection. This projection is an azimuthal, which projects a sphere on to a plane. Pictured in Figure 7 is an example of a Northern Hemisphere Polar Stereographic projection. In this implementation, the projection plane is secant at 70 degrees of latitude. This means that the areas near to 70 degrees will be most accurate. Some degree of distortion will be present at the pole as well as at the outer areas of the map. This particular projection definition is popular for Arctic sea ice data, for example, since the sea ice edge can be examined in the zone most accurately portrayed on the map (Polar Stereographic data, 2014).

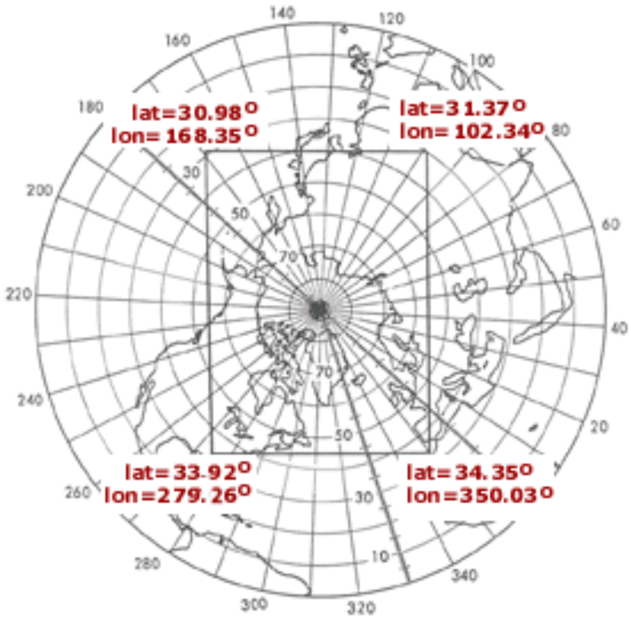


Figure 7: The map above shows a rectangular grid defined by the National Snow and Ice Data Center on top of the circular representation of the Polar Stereographic projection (Polar Stereographic data, NSIDC, 2014).

Another example is the Azimuthal Equal Area projection. The difference between this and the Polar Stereographic example above is that areas will be presented accurately across the entire map. This polar projection is common for scientific data for which area comparisons and calculations are important. The example shown in Figure 8 is the Southern Hemisphere EASE-Grid projection, which is used for various gridded snow cover, ice cover and soil moisture data sets.

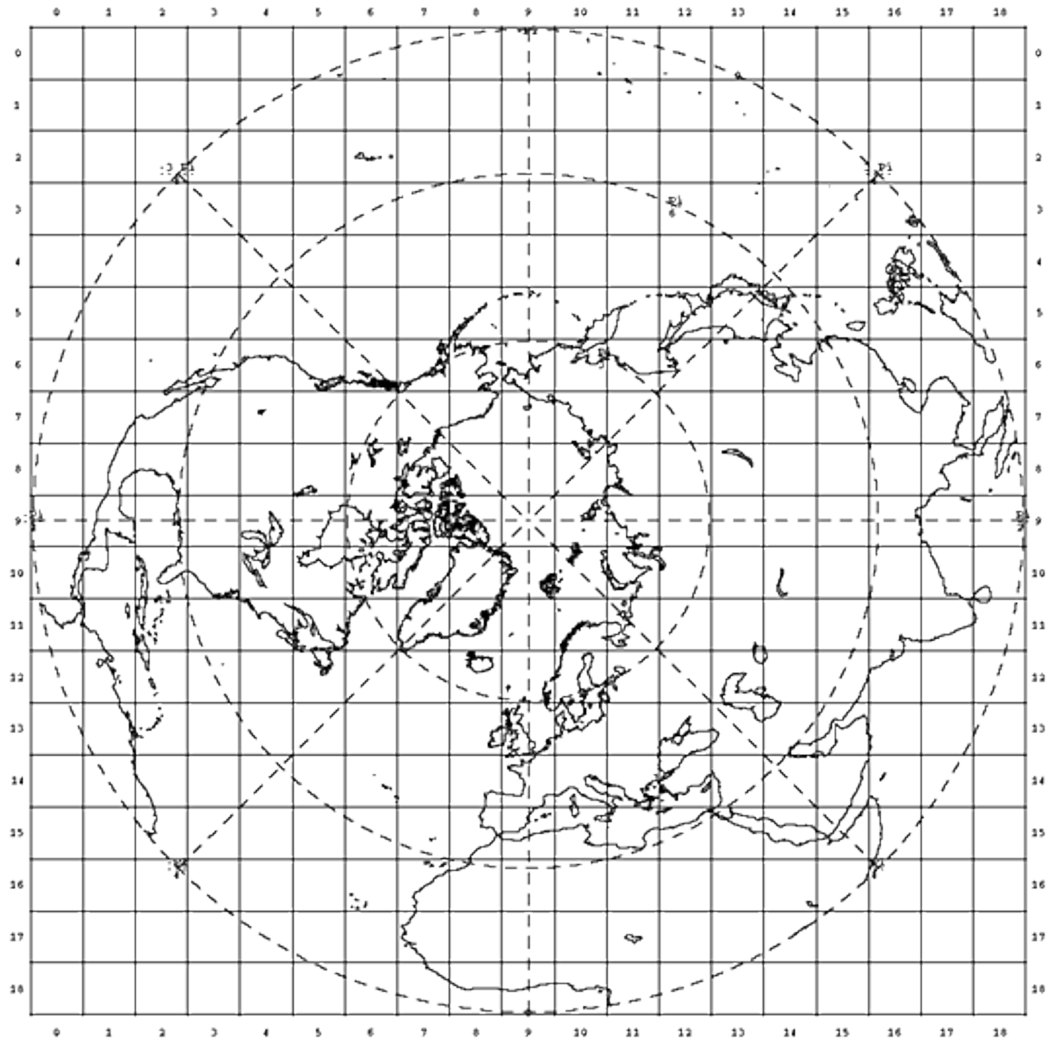


Figure 8: The map above shows the original EASE-Grid atop the Azimuthal Equal Area projection (Original EASE-Grid Format Description, NSIDC, 2014).

As is shown above, the cylindrical projection of the world map is quite different from the two azimuthal projections of the Northern and Southern Hemispheres shown thereafter. This demonstrates the need to rotate the perspective when taking a closer look at the polar regions. It also shows that conventions such as Google Maps API cannot properly address the mapping of the Arctic or the Antarctic. In fact, the projection used by Google Maps is cut off at 85 degrees of latitude (Google Maps projection, 2014). Also, tools and services built for a portal assuming a Mercator-like view of the Earth, may not always be appropriately built for use in the polar areas. Due to these reasons, Arctic and Antarctic data present a special use case for scientific data portal development.

4.5.2 Scientific data portal interfaces

Existing scientific data portals typically show a large map view of the Earth within which a user can search for and learn about the data. A few examples of this are shown below.

NASA’s Worldview interface

This data portal displays near-real-time satellite imagery of the Earth. Images are available for visualization in the Worldview interface as soon as three hours after they are collected. Because of the rapid response time of this tool, activities such as forecasting and emergency response are possible through the imagery provided. The interface excels at data visualization in that full-resolution images of the various data layers are provided. Worldview also offers the ability to browse imagery using the fourth dimension of time. The timeline at the bottom of the map in Figure 9 allows one to choose the date of the imagery portrayed (Worldview, NASA, 2014).

In Figure 9, Moderate-Resolution Imaging Spectroradiometer (MODIS) Terra/Aqua true color satellite imagery is displayed. On top of this, MODIS Terra/Aqua sea ice coverage data are overlain. The sea ice data are represented in pink at the top of the image. As is shown, the color indicating sea ice coverage is strewn across the top of the map in such a way that the geographic extent of the sea ice cannot be ascertained from the image. The sea ice edge, the area of coverage, and the sea ice shape and formation cannot be meaningfully displayed in this map because of the global projection utilized.

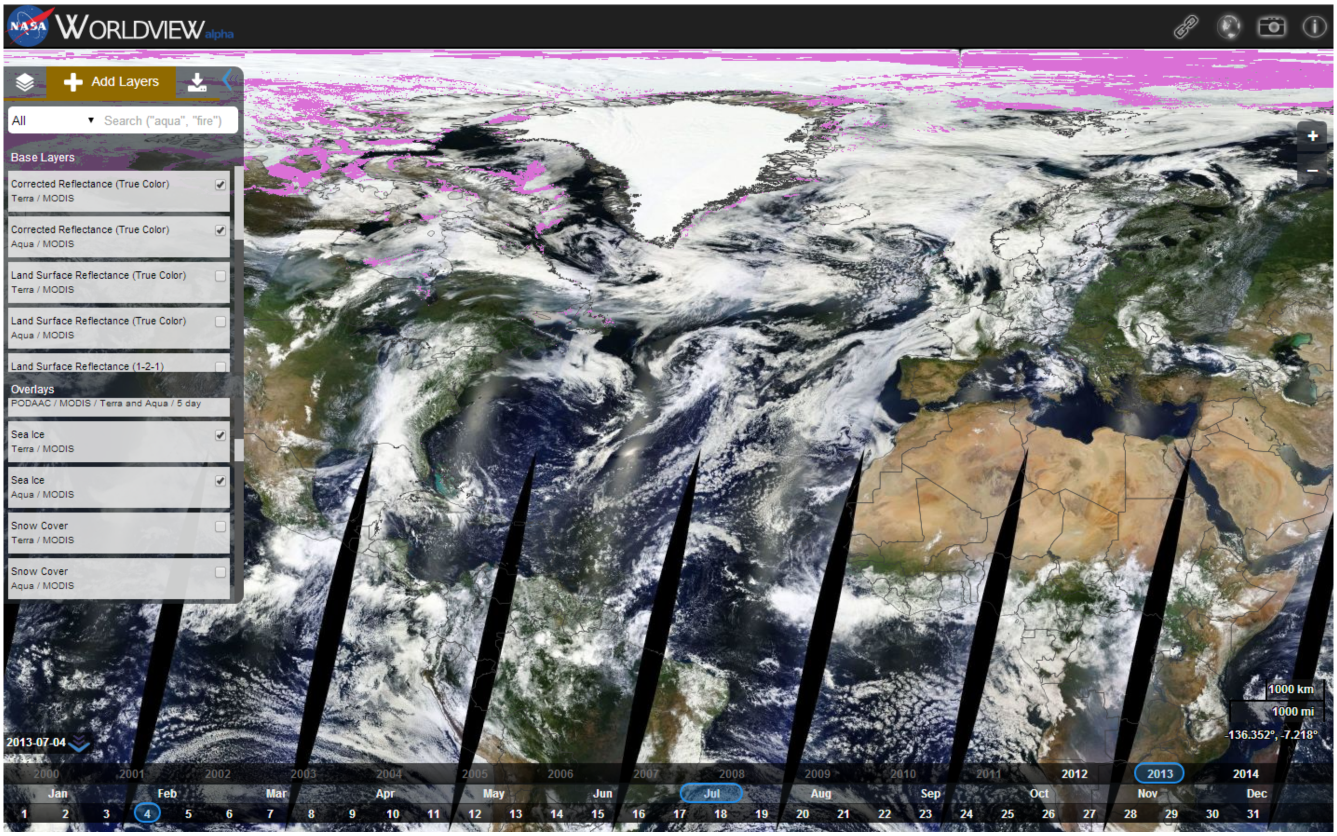


Figure 9: The Worldview interface above displays MODIS true color imagery and MODIS sea ice coverage data (Worldview, NASA, 2014).

Global Data Explorer

The next screen shot is of the Global Data Explorer portal, which was created by the Land Processes Distributed Active Archive Center (LP DAAC). Displayed is the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Digital Elevation Model. Areas of very high elevation are shown in red, while areas of low elevation appear white. For Greenland and Antarctica, the ice surface elevation is displayed. The thick ice sheets result in areas of very high elevation in both Greenland and Antarctica.

Looking more closely at Antarctica, it is obvious that the shape and scale of the continent are not able to be meaningfully displayed. Also, there are areas of missing information. Scientists wishing to visualize the elevation of the ice sheet domes on the Antarctic high plateau, would either have trouble locating the dome areas, or would see that the areas are not included in the data displayed in the portal.

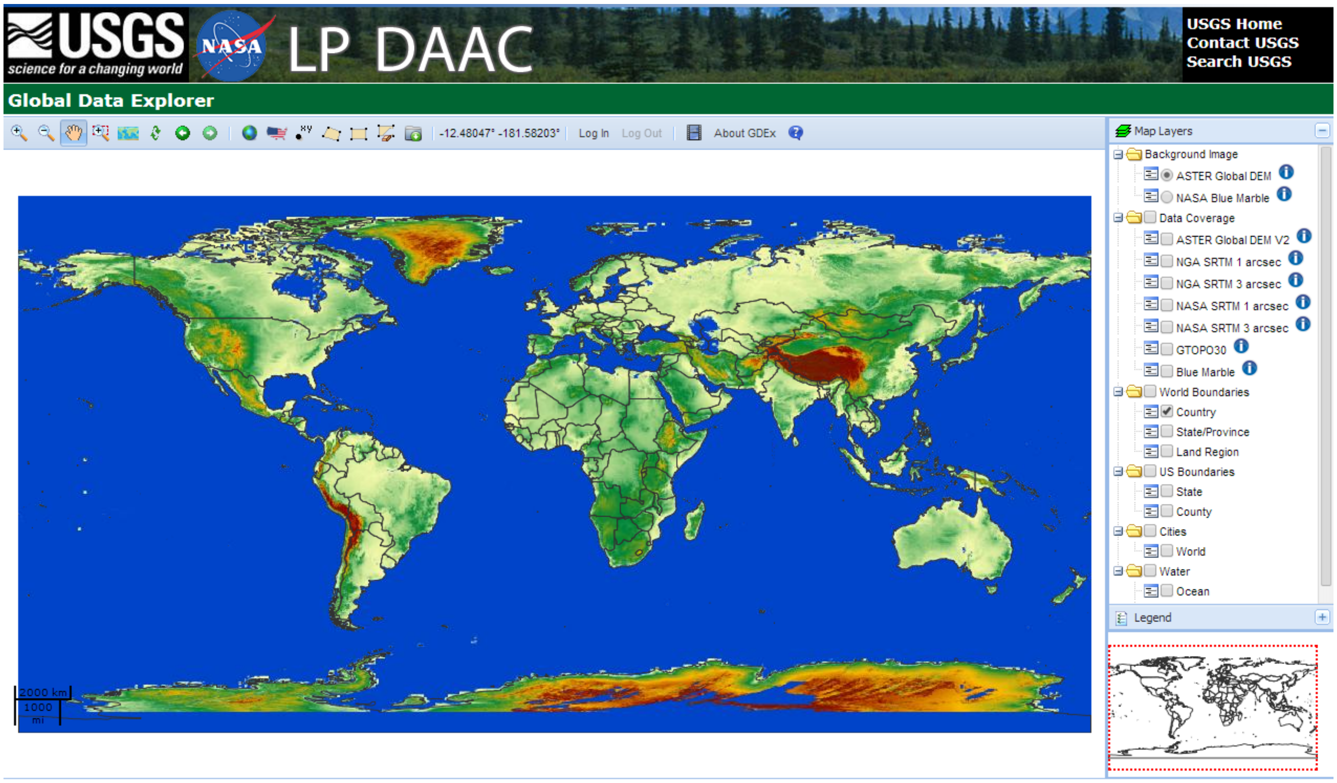


Figure 10: The image above shows the ASTER Digital Elevation Model of the world (Credit: Global Data Explorer, LP DAAC, 2014).

4.5.3 Bounding boxes for a polar web map

One common way to search for data sets in a scientific data portal is to do so geographically. The researcher will draw a bounding box on the map around a particular area of interest and search for data found at that geographic location. The bounding box is often defined by corner points, such as the upper-left and the lower-right coordinate pairs of the box. Typically, the user will click on the map to set the upper-left corner of the box, then drag the mouse downwards and to the right, clicking again to mark the lower-right corner. When overlain on a map of the world such as is shown in Figure 11, the lines of the bounding box are assumed to follow lines of latitude and longitude. The rectangular shape of the box follows what we assume to be the layout also for the grid lines on the map.

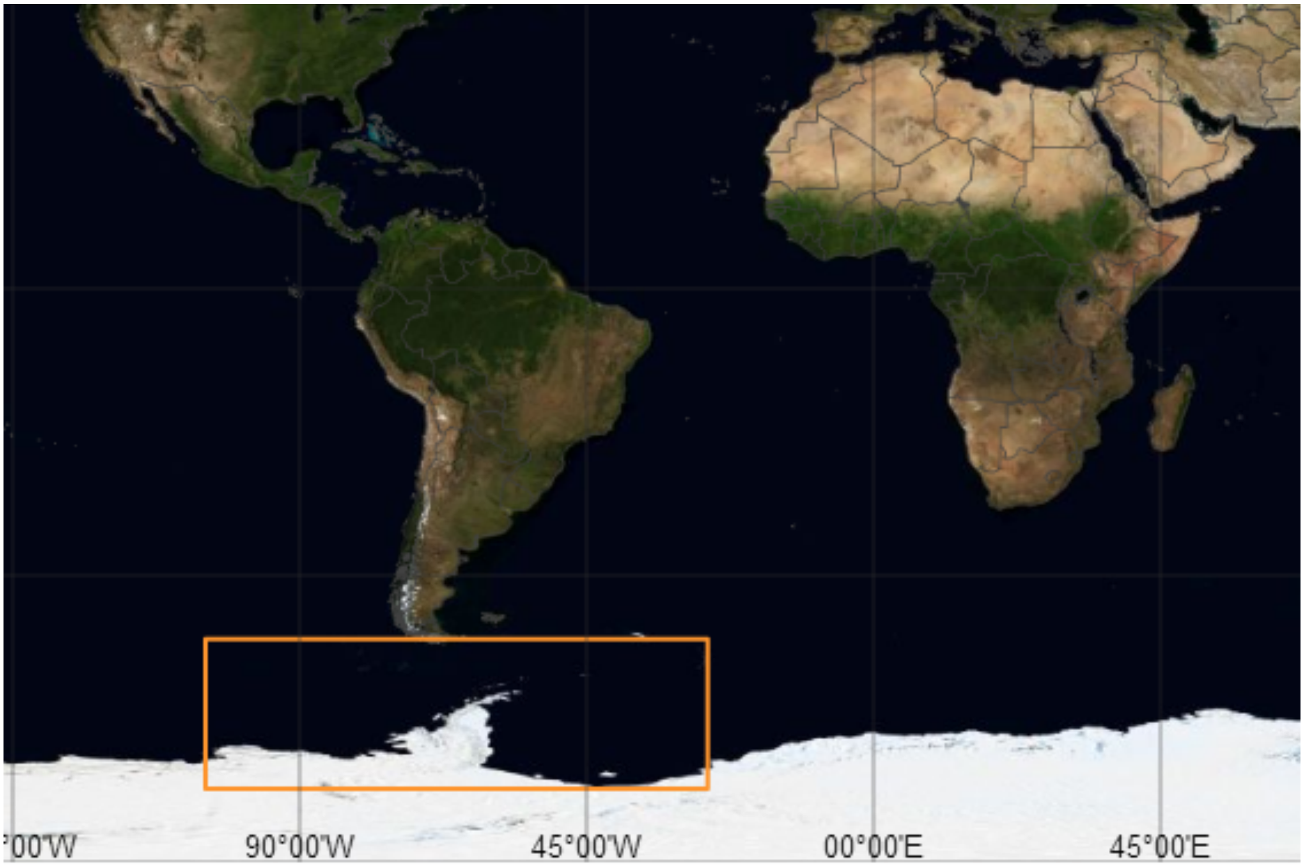


Figure 11: A common way to search for data sets in a particular geographic area is to use a bounding box (here drawn over the Antarctic Peninsula) in a global view of the Earth (Credit: Reverb, NASA, 2014).

In Figure 11, drawing a bounding box in the polar regions can prove difficult in the global view of the Earth, as the Arctic and Antarctic regions are strung out along the upper and lower borders of the map, and do not retain much of their actual form. To account for this, some scientific data portals offer polar views of the map, which give an alternative map centered at either the North or South Pole.

As was demonstrated in the polar projections in the section above, an azimuthal polar map view will result in lines of latitude radiating in concentric circles out from the pole at the center of the map. The lines of longitude appear like rays radiating outward from the pole. A bounding box drawn to follow lines of latitude and longitude will appear more like a piece of pie or a segment of a circle rather than like a box. These shapes would likely not be expected by the researcher selecting the study area.

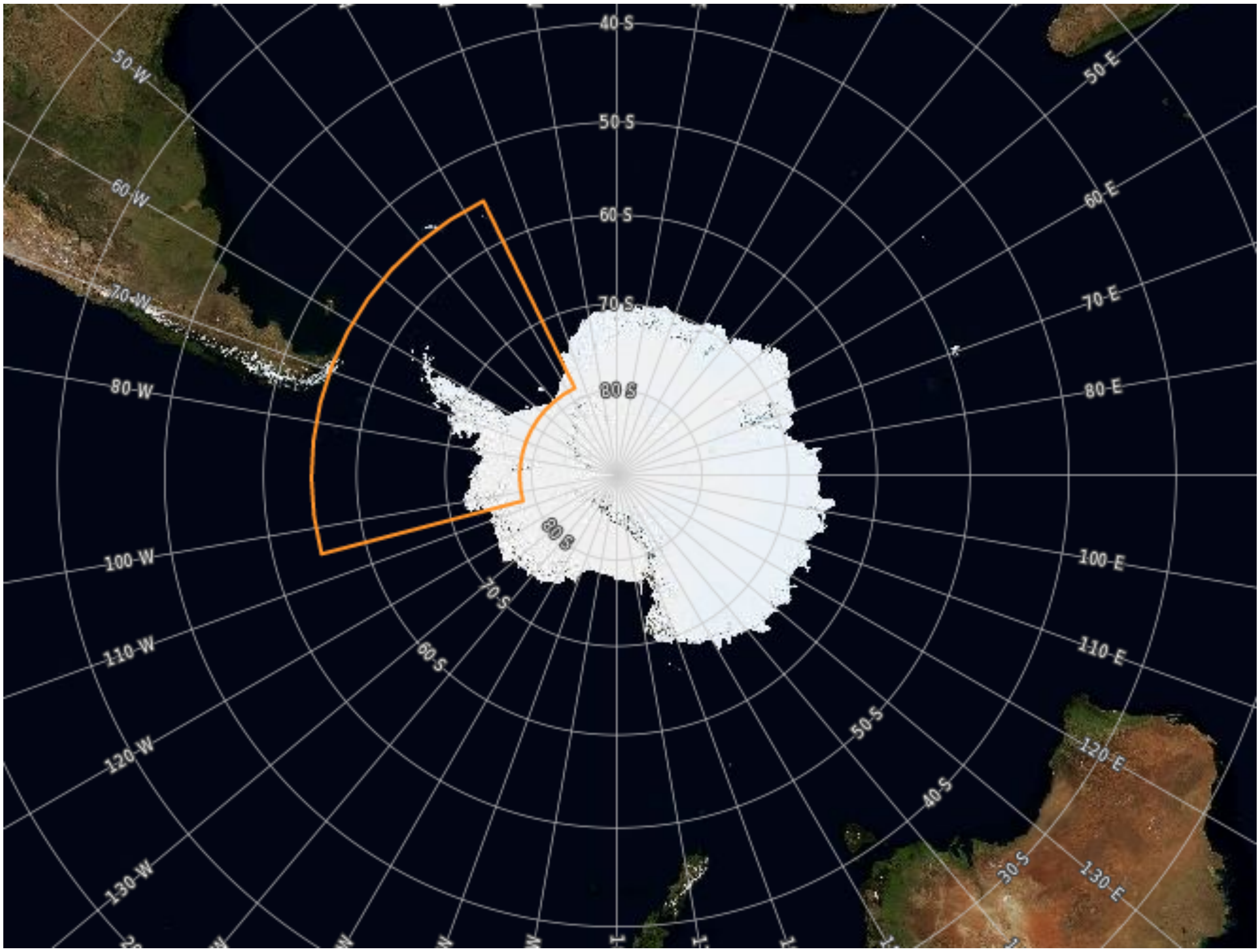


Figure 12: A pie-shaped bounding box around the Antarctic Peninsula results in the polar view of the map. This bounding box is defined by the same corner points as in the rectangular-shaped example in the previous image (Reverb, NASA, 2014).

Instead, the best practice may be to allow for a rectangular bounding box to be drawn on the polar view of the data portal. Then, the coordinates may be stored in reference to a polar grid selected as the standard grid definition for this purpose. Another option would be to enable project-on-the-fly capabilities for the bounding box and the spatial area definitions for the data set stored in the metadata.

4.5.4 Crossing the International Date Line

The International Date Line lies in the Pacific Ocean, far from any major population centers. In many Earth science disciplines, this area does not tend to be a focal point for data. However, in the polar regions, the international date line presents itself right in the midst of areas significant to scientific inquiry. For example, the Bering Strait region, where sea ice, whale migration, ocean circulation, etc. are studied, has the International Date Line running along the center of the study area. Data that include a time element will need special processing in order to account for the crossing of the date line. Also portal tools, such as visualization displays, will need to be programmed to be able to cross the date line without mishandling the data.

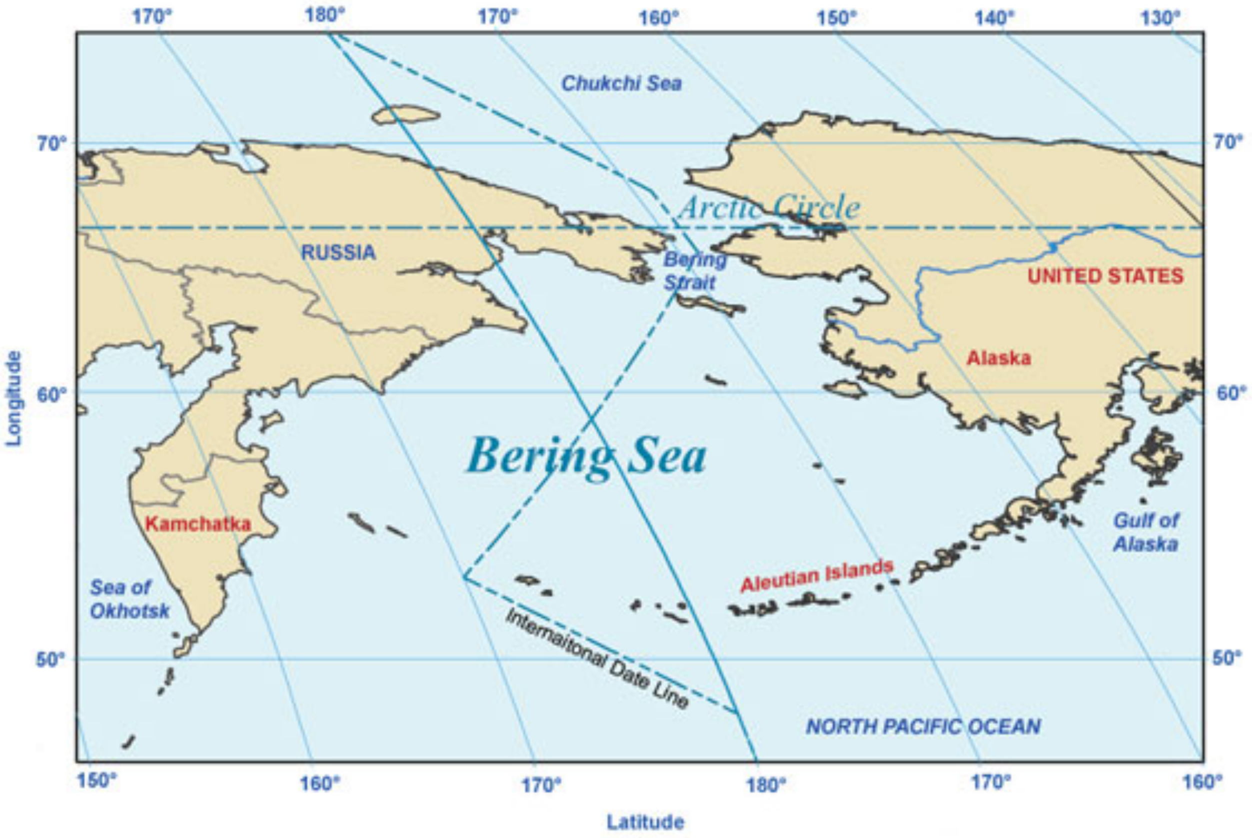


Figure 13: The boundary of the International Date Line traverses areas of relevance to Arctic science (Teacher at sea, NOAA, 2013).

4.5.5 Mapping close to the North or South Pole

The geographic North Pole is defined as 90 degrees north, while the geographic South Pole is defined as 90 degrees south. In terms of longitude, each pole is defined as simultaneously every degree of longitude, -180 to 180. This can make mapping at the poles challenging, especially when using latitude and longitude coordinates in a Geographic Coordinate System, or when using a global projection. For example, the mathematical model for the Mercator projection results in projecting the poles at infinity. Thus, maps based on this projection are cut off before they reach 90 degrees of latitude (Google Maps projection, 2013).

Figure 14 shows a rendering of the NASA Operation IceBridge areal flight lines over northern Greenland and the Arctic Ocean. The flight line highlighted in yellow passes directly over the North Pole. This flight line represents a path of science data that includes position, time, elevation, and ice properties. The location of the data over the pole requires that the data are projected in a way that allows for a spatial definition of 90 degrees of latitude. Also, the crossing of the International Date Line, which extends to the pole, requires that the time element of the data is handled properly.

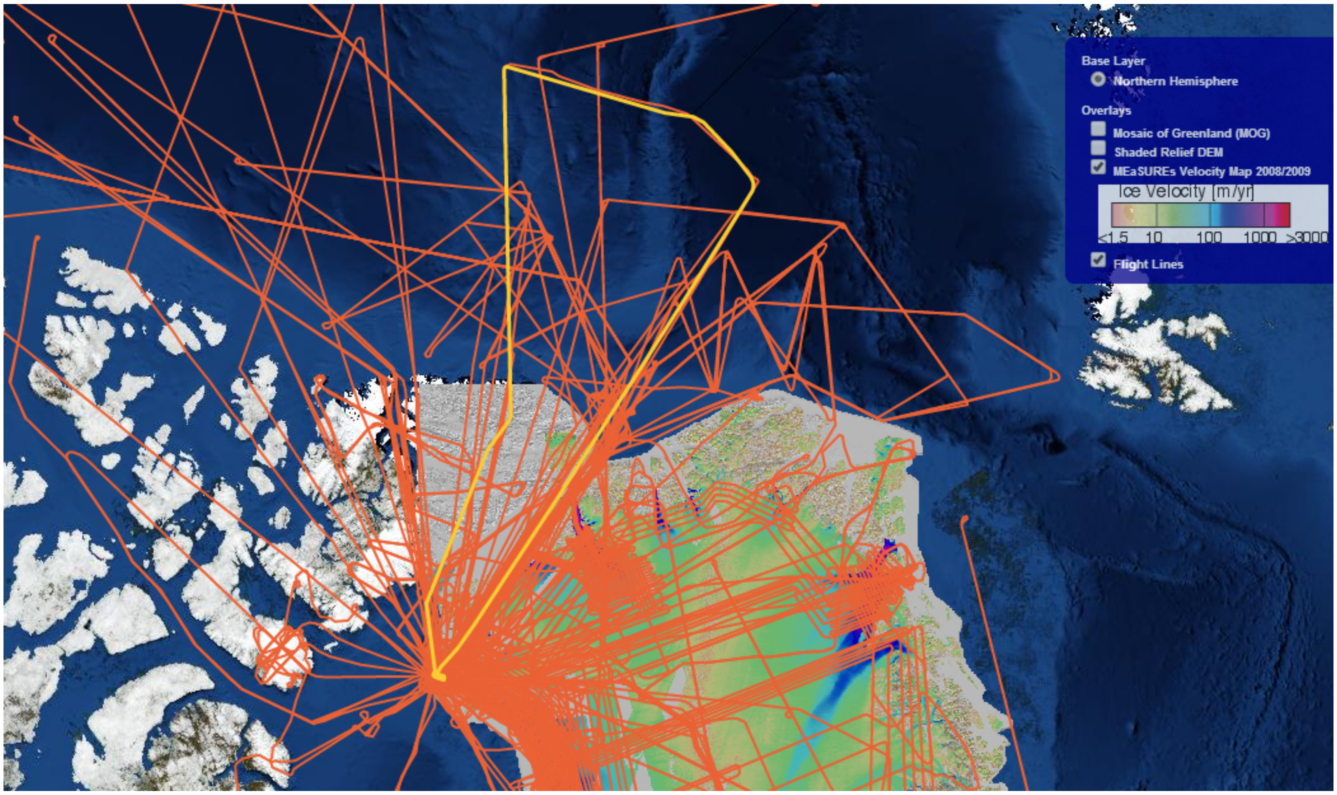


Figure 14: Flight line paths cross over the Arctic where aircraft with special instrumentation flew to collect data for NASA Operation IceBridge (Credit: NASA IceBridge Data Portal, NSIDC, 2014).

4.6 GIS as the bridge between the two poles

The discussion above exemplifies the degree to which portal development to date shows a preference for global-view, mid-latitude mapping services and tools. Taking the special circumstances for mapping at the poles into consideration requires additional development and work before data access is optimized for polar environments. Because of this, it makes sense for Arctic and Antarctic SDI development and portal initiatives to join forces in order to benefit from the other's progress. The uniqueness of the polar areas requires a special approach to mapping services that does not apply to most other aspects of Earth science data. Spatial data layers such as imagery and digital elevation models could be created for both poles with much the same approach, and would not require doubling the effort. Organizations such as the Polar Geospatial Center, which focus on science at the poles, could develop framework data sets and GIS services that target these unique areas. Polar science also leans in the direction of collaboration between the two poles since glaciology, sea ice studies, snow studies, and climatology lends itself to seeking cross-over reference from one pole to the other (Pulsifer, 2014). Because mapping techniques and spatial data at the poles are so unique, it makes sense for Antarctic and Arctic SDI efforts to build a common portal – or common standards and services for a portal.

5. Conclusion

To summarize the key points of this thesis, developing an ASDI will prove to be a difficult task because of the international collaboration required between countries with differing priorities. However, defining spatial data standards for the Arctic is increasingly necessary. Climate change, especially, has resulted in elevated interest in the Arctic area from scientists, policymakers, industry analysts and individuals alike. Providing accurate, timely and complete scientific information about the Arctic environment is of great importance. The scientific indicators involved do not pay attention to geopolitical boundaries, and these should not ultimately inhibit the development of ASDI.

The political, cultural and geographic diversity in the Arctic suggests the need for flexible or extendable infrastructure standards. Few regions can purport to have the wide range of uses for SDI as the Arctic. Furthermore, infrastructure studies indicate that building flexibility into infrastructure can create a more solid core. Involvement from the end-user early in the development process can make a difference in which infrastructure standards are adopted by the user community and which are not. Therefore, gaining input from science data end-users regarding SDI development may be a beneficial step.

Currently, the science community has been discussing the aims and goals of science data services in order to meet the needs for the future. Themes from these discussions have arisen to suggest the best practices for metadata, data management, web services, open standards services, etc. These themes can be used to inform the ASDI development in order to meet the needs and expectations of the end-user community.

In addition to this, a trend towards bigger data, greater collaboration across scientific disciplines and a more diversified pool of data users has come about. This trend implicates the need for online access to interdisciplinary data and services. In general, scientific data portals are best fit to fill this need, and should be conceptualized hand-in-hand with ASDI development.

Science data portals to date do not tend to handle polar data very well. Projections, visualization, bounding boxes, and special considerations for mapping around the International Date Line and the North and South Poles cause science data in the polar regions to require an entirely different approach than data in the mid-latitude range. Because of this, it makes sense for ASDI development to partner with Antarctic SDI development in regards to a scientific data portal.

In conclusion, the geospatial methods for mapping polar science data bring the north and south together and take precedence over their inherent differences when it comes to SDI portal development. Each could gain from working with the other towards the common goal of accessible and interoperable polar data.

References

Abbott, M. R. (2009) A new path for science? In: Hey, T., Tansley, S. & Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft, Redmond, WA. pp. 111-116. Available from: http://research.microsoft.com/en-us/collaboration/fourthparadigm/ [22 February 2014].

ACIA (2005) Arctic Climate Impact Assessment. Cambridge University Press, 1042 p. Available from: http://www.acia.uaf.edu [25 March 2014].

Altinas, I., Arctur, D., Brotzge, J., Couch, A., Dibner, P., Domenico, B., Hooper, R., Lehnert, P., Nativi, S., Plale, B., Richard, S., Stocks, K., Valentine, D. & Zaslavsky, I. (2012) *Cross-domain interoperability. EarthCube roadmap*. EarthCube Report version 1.1. Available from: http://workspace.earthcube.org/sites/default/files/files/document-repository/ [21 March 2014].

Anonymous (2009) The way ahead for polar science. Nature. 457 (7233), 1057.

Barcelona Field Studies Centre (2014) *Google Maps Projection* [Online] Available from: http://googlecompass.com/GoogleMapsProjection.htm [28 March 2014].

Bernard, L. & Ostländer, N. (2007) Assessing climate change vulnerability in the arctic using geographic information services in spatial data infrastructures. *Climate Change,* 87, 263-281.

Bowker, G., Baker, K., Millerand, F. & Ribes, D. (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger et al. (eds), *International Handbook of Internet Research*. Available from: DOI 10.1007/978-1-4020-9789-8_5.

Callaghan, T., Björn. L. O., Chapin III, S., Chernov, Y., Christensen, T. R., Huntley, B., Ims, R., Johansson, M., Riedlinger, D. J., Jonasson, S., Matveyeva, N., Occhel, W., Panikov, N. & Shaver, G. (2005) Arctic tundra and polar desert ecosystems. In: Arris, L. (ed.), *Arctic Climate Impact Assessment*, University of Alaska Fairbanks, ACIA Secretariat and Cooperative Institute for Arctic Research, pp. 244-335. Available from: http://www.acia.uaf.edu/ [21 March 2014].

Dana, Peter H. (1999), *Map Projection Overview* [Online] Available from: http://www.colorado.edu/geography/gcraft/notes/mapproj/mapproj_f.html [28 March 2014].

EarthCube (2013) *EarthCube end-user workshop: Executive summaries.* Available from: http://workspace.earthcube.org/sites/default/files/files/document-repository [21 March 2014].

Edwards, P., Jackson, S., Bowker, G. & Knobel, C. (2007) Understanding infrastructure: Dynamics, tensions, and design. In: *History & theory of infrastructure: Lessons for new scientific cyberinfrastructures*, Report from a workshop.

Finney, K. (2009) *Data and information management strategy (DIMS)*. International Council for Science. SCAR Report number: 34. Available from: http://www.scar.org/publications/reports/ [21 March 2014].

Fox, P. & Hendler, J. (2009) Semantic eScience: Encoding meaning in next-generation digitally enhanced science. In: Hey, T., Tansley, S. & Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive*

Scientific Discovery. Microsoft, Redmond, WA. pp. 111-116. Available from: http://research.microsoft.com/en-us/collaboration/fourthparadigm/ [22 February 2014].

GIT Barents: Graphic Information Technology within the Barents Region (2006) *About the project*. [Online] Available from: http://www.gitbarents.com/AboutProjects.aspx [22 March 2014].

Gray, J. (2009) Jim Gray on eScience: A transformed scientific method. In: Hey, T., Tansley, S. & Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft, Redmond, WA. pp. 111-116. Available from: http://research.microsoft.com/en-us/collaboration/fourthparadigm/ [22 February 2014].

Hunt, J. R., Baldocchi, D. & Van Ingen, C. (2009) Redefining ecological science using data. In: Hey, T., Tansley, S. & Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft, Redmond, WA. pp. 111-116. Available from: http://research.microsoft.com/en-us/collaboration/fourthparadigm/ [22 February 2014].

Indiana University (2013) *Informatics Defined*, School of Informatics and Computing [Online] Available from: http://soic.iupui.edu/about/what-is-informatics/ [22 March 2014].

Infrastructure for Spatial Information in the European Community (n.d.) *INSPIRE directive*. [Online] Available from: http://inspire.ec.europa.eu/ [22 March 2014].

IPCC (2007) *Climate change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, Pachuari, R. K. and Reisinger, A. (eds)]. IPCC, Geneva, Switzerland.

Johnson, G. W., Gaylord, A. G., Franco, J. C., Cody, R. P., Brady, J. J., Manley W., Dover, M., Garcia-Lavigne, D., Score, R. & Tweedie, C. E. (2011) Development of the Arctic Research Mapping Application (ARMAP): Interoperability challenge and solutions. *Computers & Geosciences*, 37, 1735-1742.

Li, W., Yang, C., Nebert, D., Raskin, R., Houser, P., Wu, H. & Li, Z. (2011) Semantic-based web service discovery and chaining for building an Arctic spatial data infrastructure. *Computers & Geosciences,* 37, 1752-1762.

Maxwell, R. (2014) Mapping the Earth's polar regions. *GIS Lounge*, GIS Data, 2 January [Online] Available from: http://www.gislounge.com/mapping-earths-polar-regions/ [22 March 2014].

National Aeronautics and Space Administration (NASA) (2014) *Global Change Master Directory* [Online] Available from:http://gcmd.nasa.gov/index.html [28 March 2014].

National Aeronautics and Space Administration (NASA) *Earth Observing System Data and Information System (EOSDIS)* [Online] Available from:http://reverb.echo.nasa.gov/reverb/ [28 March 2014].

National Aeronautics and Space Administration (NASA) *Worldview* [Online] Available from: https://earthdata.nasa.gov/labs/worldview/ [28 March 2014].

NASA Science (2013) *How is the global earth system changing?* [Online] Available from: http://science.nasa.gov/ [22 March 2014].

National Oceanic and Atmospheric Administration: NOAA (2013) Teacher at sea [Online] Available from: http://teacheratsea.wordpress.com/tag/international-date-line/ [30 March 2014].

National Snow and Ice Data Center (2014) *Arctic Sea Ice News and Analysis* [Online] Available from: http://nsidc.org/arcticseaicenews [22 March 2014].

National Snow and Ice Data Center (2014) *Satellite Observations of Arctic Change* [Online] Available from: http://nsidc.org/soac [22 March 2014].

National Snow and Ice Data Center (2014) *Polar Stereographic Data* [Online] Available from: http://nsidc.org/data/polar_stereo/ps_grids.html [28 March 2014].

National Snow and Ice Data Center (2014) *Original EASE-Grid Format Description* [Online] Available from: http://nsidc.org/data/ease/ease_grid.html[28 March 2014].

National Snow and Ice Data Center (2013) *NASA Operation IceBridge Data Portal* [Online] Available from: http://nsidc.org/icebridge/portal [28 March 2014].

National Snow and Ice Data Center (2014) *Metrics reports giving statistics on the NSIDC Data Set Search Tool at:* http://nsidc.org/data/search.

Neural Electric Magnetic Ontologies (NEMO) (2014) *Data Annotation* [Online] Available from: http://nemo.nic.uoregon.edu/wiki/Data_Annotation [28 March 2014].

Open Geospatial Consortium (2014) *OGC Standards*[Online] Available from: http://www.opengeospatial.org/standards [15 June 2014].

Palmer, O., Skedsmo, M., Gudmundsson, M., Finnbogadottir, E. L., Barry, T., Ursin, H., Thaulow, I., Obinyakov, V., Nebert, D., O'Brien, D. & Hartmann, J. P. W. (2011) The Arctic Spatial Data Infrastructure. *Arctic SDI,* Project Plan v1. Available from: http://arctic-sdi.org/wp-content/uploads/2011/12/ [23 March 2014].

Polar Geospatial Center. (n.d.) *Our history.* [Online] Available from: http://www.agic.umn.edu/about/ [21 March 2014].

Pulsifer, P. Research Scientist. (Personal communication, 24 March 2014).

Scientific Committee on Antarctic Research (2013) *About SCAR* [Online] Available from: http://www.scar.org/about/ [22 March 2014].

Sorensen, M., Rich, S., Hoover, S. & Tweedie, C. (2004) *Arctic spatial data infrastructure (SDI) for scientific research: Draft White Paper.* Circumarctic Portal Framework. Report v2.

Star, S. L. & Ruhleder, K. (1996) Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research,* 7 (1), 111-134.

Taylor, D. R. F. (2014) Email sent to Sarah Tressel, 25th March.

Tedesco, M., Fettweis, X., Mote, T., Wahr, J., Alexander, P., Box, J. & Wouter, B. (2012) Evidence and analysis of 2012 Greenland records from spaceborne observations, a regional climate model and reanalysis. *The Cryosphere Discussion*s, 6, 4939-4976.

Tedesco, M. (2013) *The polar cyberinfrastructure program* [Presentation] PTC Meeting. Available from: http://polarpower.org/PTC [22 March 2014].

Tressel, S., Bauer, R. & Scambos, T. (2013) The Antarctic Glaciological Data Center provides support for NSF Division of Polar Programs Principal Investigators to share their snow, ice and related data. In: *Polar snow and firn: physical and chemical processes in the snowpack*. C13B-0678 Poster. American Geophysical Union Fall Meeting. 9-13 December.

UNEP/GRID-Arendal (2006) *The Arctic region-definitions*, graphic [Online] Available from: http://www.grida.no/graphicslib/detail/the-arctic-region-definitions_beb4 [22 March 2014].

United States Geological Survey (2014) *Global Data Explorer* [Online] Available from: http://gdex.cr.usgs.gov/gdex/ [28 March 2014].

Van Sickle, J. (2004) Basic GIS coordinates. USA, CRC Press.

Yang, C., Li, W., Xie, J. & Zhou, B. (2008) Distributed geospatial information processing: sharing distributed geospatial resources to support Digital Earth. *International Journal of Digital Earth*, 1 (3), 259-278.

Yang, C., Raskin, R., Goodchild, M. & Gahegan, M. (2010a) Geospatial Cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems,* 34, 264-277.

Yang, C., Wu, H., Huang, Q., Li, Z. & Li, J. (2010b) Using spatial principles to optimize distributed computing for enabling the physical science discoveries. *Proceedings of the National Academy of Sciences*, Early Edition. Available from: DOI: 10.1073/pnas.0909315108. [22 February 2014].

Yang, C., Nebert, D. & Taylor D. R. F. (2011) Establishing a sustainable and cross-boundary geospatial cyberinfrastructure to enable polar research. *Computers & Geosciences*, 37, 1721-1726.

Appendix 1:

Table of Acronyms

Acronym	Description
ACIA	Arctic Climate Impact Assessment
ADCN	Arctic Data Coordination Network
AMD	Antarctic Master Directory
AntSDI	Antarctic Spatial Data Infrastructure
API	Application Programming Interface
ARMAP	Arctic Research Mapping Application
ASDI	Arctic Spatial Data Infrastructure
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
CAFF	Conservation of Arctic Flora and Fauna
CCAD	Committee on the Coordination of Antarctic Data
COMNAP	Council of Managers of National Antarctic Programs
CSW	Catalog services for the Web
DIF	Directory Interchange Format
DOI	Digital Object Identifiers
EASE-Grid	Equal-Area Scalable Earth Grid
GCMD	Global Change Master Directory
GIS	Geographic Information Systems
GIT Barents	Geographic Information Technology within the Barents Region
IASC	International Arctic Science Committee
INSPIRE	Infrastructure for Spatial Information in the European Community
IPCC	Intergovernmental Panel on Climate Change
IPY	International Polar Year
JCADM	Joint SCAR/COMNAP Committee on Antarctic Data Management
LP DAAC	Land Processes Distributed Active Archive Center
MODIS	Moderate-Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NEMO	Neural Electric Magnetic Ontologies
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
NSIDC	National Snow and Ice Data Center
OGC	Open Geospatial Consortium
RDA	Research Data Alliance
SAON	Sustaining Arctic Observing Networks
SCADM	Standing Committee on Antarctic Data Management
SC-AGI	Standing Committee on Antarctic Geographic Information
SCAR	Standing Committee on Antarctic Research
SDI	Spatial Data Infrastructure
SKOS	Simple Knowledge Organization System
UNEP	United National Environment Programme
USA	United States of America
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service

Appendix 2:

References for Figure 4

Images for Figure 4 [Online] Available from:
http://pixabay.com/en/world-globe-global-earth-97864/
http://www.grida.no/graphicslib/region/geoarctic
http://nsidc.org/data/measures/
http://nsidc.org/data/nsidc-0516
http://pixabay.com/en/buoy-buoy-marker-ocean-sea-72044/
http://visibleearth.nasa.gov/view.php?id=54229
http://upload.wikimedia.org/wikipedia/commons/c/cf/Worldmap_LandAndPolitical.jpg
http://nsidc.org/images/arcticseaicenews/20090406_Figure4.png
https://farm6.staticflickr.com/5086/5367930090_d03698c365_s.jpg
https://upload.wikimedia.org/wikipedia/commons/thumb/a/a9/Globalndvi_tmo_200711_lrg.jpg/300px-Globalndvi_tmo_200711_lrg.jpg
http://upload.wikimedia.org/wikipedia/commons/thumb/a/ad/Globcov.png/320px-Globcov.png
http://upload.wikimedia.org/wikipedia/commons/e/e0/Water_erosion_map.jpg
[26 March 2014].

Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science (LUMA-GIS)

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland. (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research. (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing. (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia. (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010)
8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010)
9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010)
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011)
11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011)
12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).
13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).
14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011)
15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011)
16. *David Kanyari*: Nairobi City Journey Planner An online and a Mobile Application (2011)

17 *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012)
18 *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012)
19 *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012)
20 *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013)
21 *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013)
22 *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013)
23 *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013)
24 *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013)
25 *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013)
26 *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014)
27 *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014)
28 *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014)
29 *Brynja Guðmundsdóttir* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014)
30 *Oleksandr Nekrasov* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014)
31 *Sarah Tressel* Recommendations for a polar Earth science portal
in the context of Arctic Spatial Data Infrastructure (2014)