

OGC Arctic Spatial Data Pilot

Phase 2 Report

version 1.2 - 2017-07-19

Table of Contents

1. Summary	4
1.1. Preface	4
1.2. Scope	4
1.3. Report Structure	5
1.4. Foreword	5
2. Abbreviated terms	6
3. SDI Concepts	7
4. Use Case/Scenarios	8
4.1. New Shipping Routes in the Arctic	8
4.2. Search & Rescue in the Hudson Strait	8
4.3. Modeling, Forecasting	9
4.4. 3D Data Visualization & Temporal Patterns	10
4.5. Food Security in the Arctic	11
4.6. Arctic SDI	12
4.7. Complex Data Analysis	13
4.8. Sea Ice Age Evolution	14
5. Arctic SDI	16
6. Lessons Learned	19
6.1. Architecture	19
6.2. Data	21
6.2.1. Missing Metadata for OGC Web Services Content	21
6.2.2. Data Formats	21
6.2.3. Styling of Vector Data	21
6.2.4. Temporal Characteristics	22
6.2.5. Vendor Specific Solutions	23
6.2.6. Open Data and Data Access	23
6.2.7. Shared Semantics & Quality Information	24
6.2.8. Aggregation and Data Fusion	24
6.3. Communication	25
6.3.1. Discovery & Rapid Exploration	25
6.3.2. Broken Links and Portals	25
6.3.3. Web service availability of registered dataset	26
6.3.4. CORS - Cross-origin resource sharing	27
6.3.5. HTTPS	27
6.3.6. Denied, Degraded, Intermittent, or Limited bandwidth	27
6.3.7. Projection Issues	27
6.3.8. Raw Data vs. Maps	28
6.3.9. Need for Catalogs	28
6.3.10. GetCapabilities Issues	29

6.4. Experiences made setting up the ArcticSDP Catalog	29
6.4.1. Catalog Introduction	29
6.4.2. Catalog harvesting	30
6.4.3. Implementation for harvesting metadata from OGC catalog server	30
6.4.4. Implementation for harvesting metadata from OGC Web server	30
6.4.5. Metadata indexing/ranking	31
6.5. Lessons Learned: Pilot Itself	31
6.5.1. A nice demo doesn't (automatically) make a good story	31
6.5.2. Think of the demonstration video early in the project	32
6.5.3. Align the scenario with the project requirements and the available data sources	32
7. Conclusions	33
Appendix A: Complete list of datasets analyzed for the project	35
Appendix B: Revision History	42

OGC ARCTIC SPATIAL DATA PILOT: PHASE 2 REPORT

OGC® document: 17-068

Editors: Ingo Simonis, Frédéric Houbie

COPYRIGHT

Copyright © 2017 Open Geospatial Consortium. To obtain additional rights of use, visit <http://www.opengeospatial.org/>

WARNING

This document is not an OGC Standard. This document is an OGC Public Engineering Report created as a deliverable in an OGC Interoperability Initiative and is not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, any OGC Engineering Report should not be referenced as required or mandatory technology in procurements. However, the discussions in this document could very well lead to the definition of an OGC Standard.

LICENSE AGREEMENT

Permission is hereby granted by the Open Geospatial Consortium, ("Licensor"), free of charge and subject to the terms set forth below, to any person obtaining a copy of this Intellectual Property and any associated documentation, to deal in the Intellectual Property without restriction (except as set forth below), including without limitation the rights to implement, use, copy, modify, merge, publish, distribute, and/or sublicense copies of the Intellectual Property, and to permit persons to whom the Intellectual Property is furnished to do so, provided that all copyright notices on the intellectual property are retained intact and that each person to whom the Intellectual Property is furnished agrees to the terms of this Agreement.

If you modify the Intellectual Property, all copies of the modified Intellectual Property must include, in addition to the above copyright notice, a notice that the Intellectual Property includes modifications that have not been approved or adopted by LICENSOR.

THIS LICENSE IS A COPYRIGHT LICENSE ONLY, AND DOES NOT CONVEY ANY RIGHTS UNDER ANY PATENTS THAT MAY BE IN FORCE ANYWHERE IN THE WORLD. THE INTELLECTUAL PROPERTY IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE DO NOT WARRANT THAT THE FUNCTIONS CONTAINED IN THE INTELLECTUAL PROPERTY WILL MEET YOUR REQUIREMENTS OR THAT THE OPERATION OF THE INTELLECTUAL PROPERTY WILL BE UNINTERRUPTED OR ERROR FREE. ANY USE OF THE INTELLECTUAL PROPERTY SHALL BE MADE ENTIRELY AT THE USER'S OWN RISK. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR ANY CONTRIBUTOR OF INTELLECTUAL PROPERTY RIGHTS TO THE INTELLECTUAL PROPERTY BE LIABLE FOR ANY CLAIM, OR ANY DIRECT, SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM ANY ALLEGED INFRINGEMENT OR ANY LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR UNDER ANY OTHER LEGAL THEORY, ARISING OUT OF OR IN CONNECTION WITH THE IMPLEMENTATION, USE, COMMERCIALIZATION OR PERFORMANCE OF THIS INTELLECTUAL PROPERTY.

This license is effective until terminated. You may terminate it at any time by destroying the Intellectual Property together with all copies in any form. The license will also terminate if you fail to comply with any term or condition of this Agreement. Except as provided in the following sentence, no such termination of this license shall require the termination of any third party end-user sublicense to the Intellectual Property which is in force as of the date of notice of such termination. In addition, should the Intellectual Property, or the operation of the Intellectual Property, infringe, or in LICENSOR's sole opinion be likely to infringe, any patent, copyright, trademark or other right of a third

party, you agree that LICENSOR, in its sole discretion, may terminate this license without any compensation or liability to you, your licensees or any other party. You agree upon termination of any kind to destroy or cause to be destroyed the Intellectual Property together with all copies in any form, whether held by you or by any third party.

Except as contained in this notice, the name of LICENSOR or of any other holder of a copyright in all or part of the Intellectual Property shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Intellectual Property without prior written authorization of LICENSOR or such copyright holder. LICENSOR is and shall at all times be the sole entity that may authorize you or any third party to use certification marks, trademarks or other special designations to indicate compliance with any LICENSOR standards or specifications.

This Agreement is governed by the laws of the Commonwealth of Massachusetts. The application to this Agreement of the United Nations Convention on Contracts for the International Sale of Goods is hereby expressly excluded. In the event any provision of this Agreement shall be deemed unenforceable, void or invalid, such provision shall be modified so as to make it valid and enforceable, and as so modified the entire Agreement shall remain in full force and effect. No decision, action or inaction by LICENSOR shall be construed to be a waiver of any rights or remedies available to it.

None of the Intellectual Property or underlying information or technology may be downloaded or otherwise exported or reexported in violation of U.S. export laws and regulations. In addition, you are responsible for complying with any local laws in your jurisdiction which may impact your right to import, export or use the Intellectual Property, and you represent that you have complied with any regulations or registration procedures required by applicable law to make this license enforceable.

Chapter 1. Summary

1.1. Preface

The OGC Arctic Spatial Data Pilot (Arctic SDP), sponsored by US Geological Survey (USGS) and Natural Resources Canada (NRCan), was initiated to demonstrate the diversity, richness and value of providing geospatial data using International Standards in support of Spatial Data Infrastructures. It demonstrated how standards and interoperability arrangements help stakeholders to gain new perspectives into social, economic, and environmental issues by providing an online network of resources that improves the sharing, use and integration of information tied to geographic locations in North America, the Arctic, and around the world.

Arctic scenarios were developed with assistance from stakeholders including the Inuit Circumpolar Council (Alaska) and Arctic Council's Conservation of Arctic Flora and Fauna working group. Data intensive scenarios covering sea ice evolution, caribou migration analysis, effects of new shipping routes in the Arctic, food security, and landslide susceptibility mapping using spatial data infrastructure components were tested and implemented.

Further, pan-Arctic science, monitoring, and societal, economic, and environmental decision support are improved with increased data sharing. In a reciprocal process, the Arctic SDP helps to generate a better understanding of how the national spatial data infrastructures can be developed and applied to support Arctic priorities. By implementing consistent means to share geographic data among all users, costs for collecting and using data can be significantly reduced while decision-making is enhanced.

1.2. Scope

This OGC Engineering Report summarizes all experiences made during the implementation phase, provides guidelines for future service setup and data handling, and identifies future work items and potential approaches. Topics detailed in this document are:

- use cases, their implementation, and experiences during their implementation
- implemented architecture, services, and data
- ideas for a future improved architecture of an SDI for the Arctic
- an integration concept for data that is available at non-OGC compliant interfaces (e.g. at proprietary portals)
- recommendations on the handling of metadata and semantics

- general recommendations for the setup and operation of a successful SDI in the Arctic
- reference architecture discussion for an SDI for the Arctic
- ideas for future activities

This Engineering Report complements Arctic Spatial Data Pilot phase 1 report, which described the results of an OGC concept study executed during the first months of the pilot.

1.3. Report Structure

This report starts with a brief **Introduction to Spatial Data Infrastructures** that has been taken from the GSDI cookbook to help the inexperienced reader to get familiar with concepts and ideas behind this important approach.

This phase of the pilot has been developed around a set of **Scenarios** that are described in the following chapter. Each scenario explored specific aspects of the Arctic and helped to generate a more complete picture of the value of spatial data infrastructures for the arctic. Chapter **Arctic SDI** briefly introduces the Arctic SDI is an important driver for that pilot. Experiences made while implementing the scenarios are the base for the **Lessons Learned** that are discussed in the following next chapter.

Chapter **Conclusions** discusses the overall situation of spatial data infrastructures for the arctic. It provides some guidelines that need to be implemented by national and international SDI initiatives in order to further improve the user's experiences with SDIs for the Arctic.

Annex A lists all data sets that have been registered at a temporary OGC Catalog Service during the pilot.

1.4. Foreword

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

Chapter 2. Abbreviated terms

- INSPIRE Infrastructure for Spatial Information in Europe.
- IPR Intellectual Property Rights
- SDI Spatial Data Infrastructure

Chapter 3. SDI Concepts

The following paragraphs have been taken from GSDI's SDI Cookbook [1: From the SDI Cookbook, GSDI.] They are repeated here to help the inexperienced reader to understand the concept and relevance of Spatial Data Infrastructures.

The term "Spatial Data Infrastructure" (SDI) is often used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. The SDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general.

The word infrastructure is used to promote the concept of a reliable, supporting environment, analogous to a road or telecommunications network, that, in this case, facilitates the access to geographically-related information using a minimum set of standard practices, protocols, and specifications. Like roads and wires, an SDI facilitates the conveyance of virtually unlimited packages of geographic information.

An SDI must be more than a single data set or database; an SDI hosts **geographic data** and attributes, sufficient documentation (**metadata**), a means to discover, visualize, and evaluate the data (catalogues and Web mapping), and some method to provide access to the geographic data. Beyond this are additional services or software to support applications of the data. To make an SDI functional, it must also include the organizational agreements needed to coordinate and administer it on a local, regional, national, and or trans-national scale.

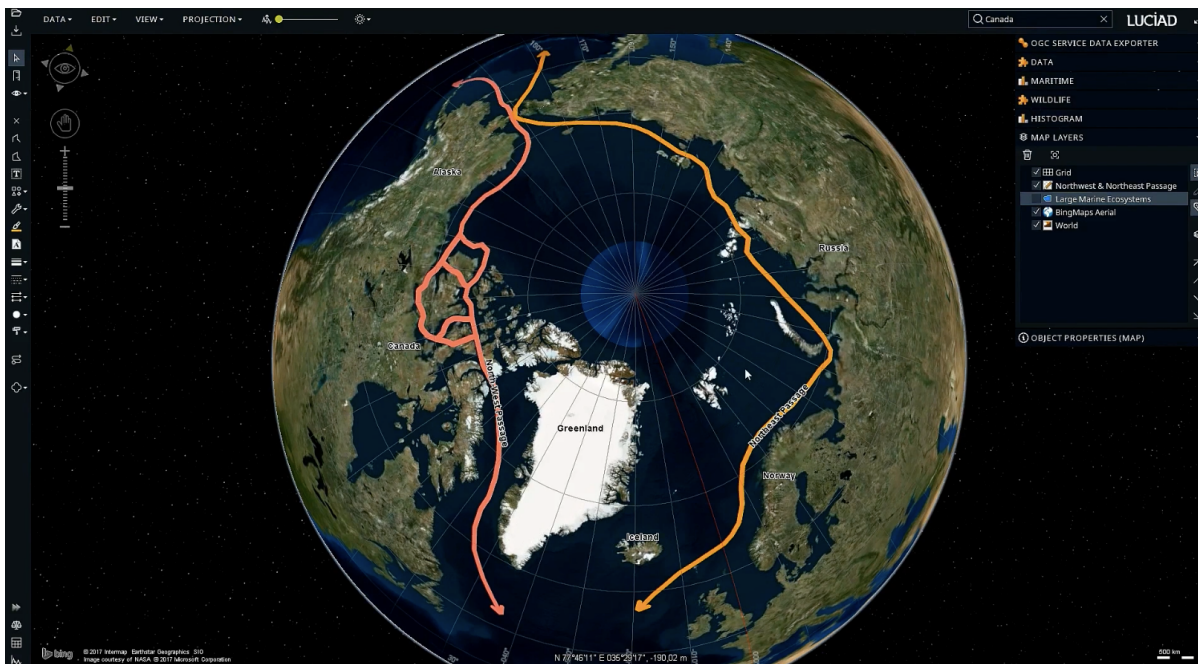
Although the core SDI concept includes within its scope neither base data collection activities or myriad applications built upon it, the infrastructure provides the ideal environment to connect applications to data – influencing both data collection and construction of applications through minimal appropriate standards and policies.

Chapter 4. Use Case/Scenarios

The OGC Arctic Spatial Data Pilot used a scenario-based approach to demonstrate the value and richness of standards-based Spatial Data Infrastructures for the Arctic. Each scenario is briefly outlined in the following sub-clauses. Further on, each scenario is documented as video material, which is available online on the project [website](http://www.opengeospatial.org/pub/ArcticSDP/index.html) [http://www.opengeospatial.org/pub/ArcticSDP/index.html].

4.1. New Shipping Routes in the Arctic

New Shipping Routes in the Arctic: Reduced Sea Ice Bears Both Great Potential and Risk scenario: The Arctic encompasses a number of shipping routes, grouped into a Northwest Passage and a Northeast Passage. Each passage crosses a number of Large Marine Ecosystems (LMEs), potentially impacting their large amount of wildlife species by disturbances and implications from shipping activity.

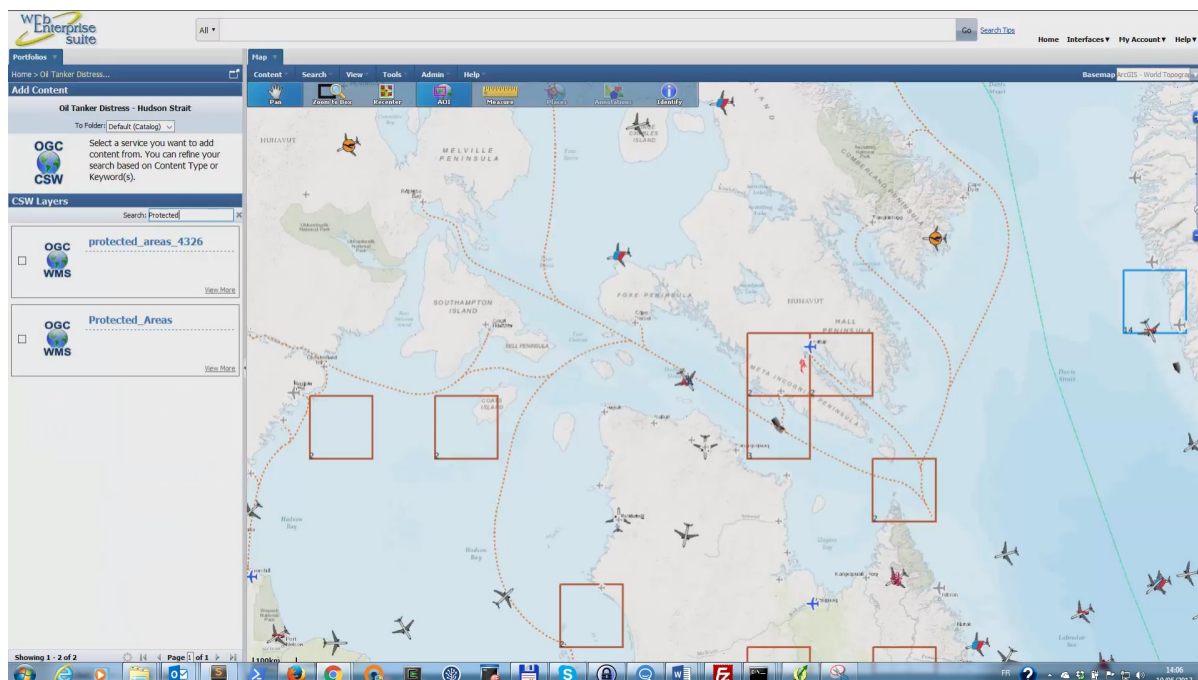


The [video](http://www.opengeospatial.org/pub/ArcticSDP/blog-shippingRoutes.html) [http://www.opengeospatial.org/pub/ArcticSDP/blog-shippingRoutes.html] illustrates how standardized data infrastructures allow the rapid integration of data from various sources, independent of their physical location. For the "New Shipping Routes" scenario, data from the US, Canada, Finland, Sweden, Norway, and Russia has been integrated easily thanks to common standards. On the fly re-projection allows users to visualize all data on a single virtual globe without unintended distortions or skews.

4.2. Search & Rescue in the Hudson Strait

Search & Rescue in the Hudson Strait: Real Time Data Integration and Offline Situation Handling in Canada's North: The "Search & Rescue in the Hudson Strait" scenario shows

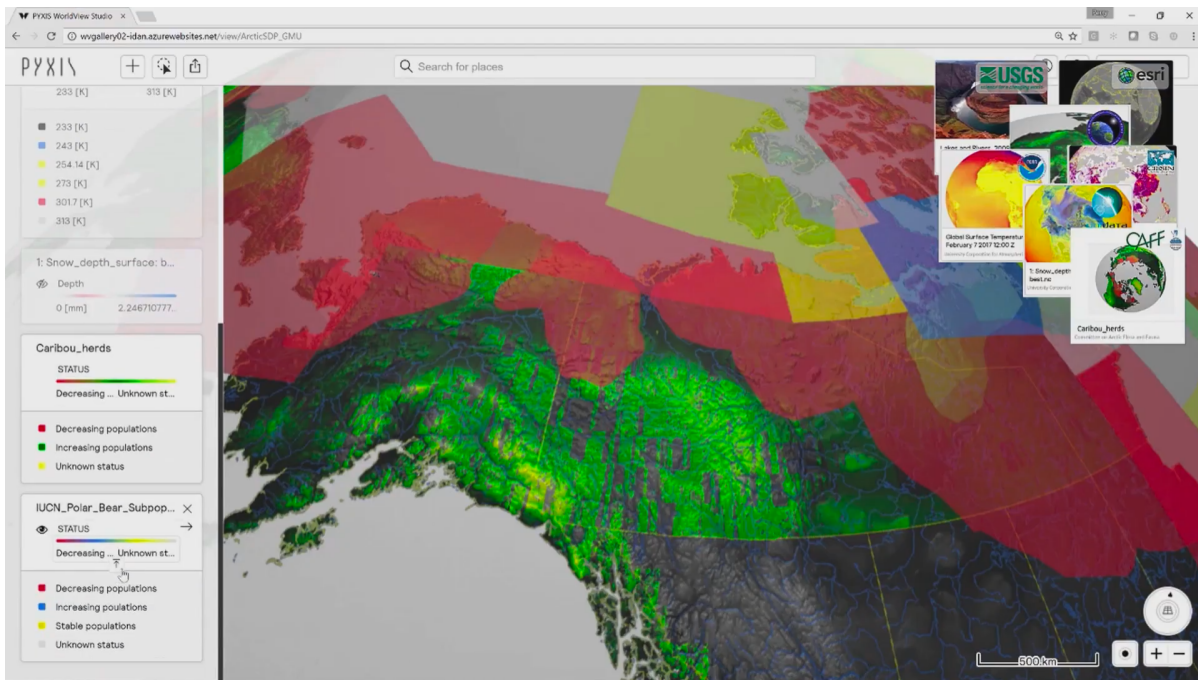
how data that is accessible via OGC standards could be used to support Search and Rescue in an emergency situation. The Canadian Coast Guard (CCG) receives a distress message from an oil tanker in the Hudson Strait. The oil tanker reports there has been an explosion on board and there are a few minor casualties, which require medical attention.



The video [<http://www.opengeospatial.org/pub/ArcticSDP/blog-searchRescue.html>] illustrates how various real-time data is integrated on the fly with archived data to generate a single operational picture. The problem of intermittent Internet connectivity is mitigated by the offline data handling of GeoPackages. Data is subsequently synchronized with the master data set when connectivity is restored.

4.3. Modeling, Forecasting

Modeling, Forecasting: Analysis of Scientific Data to Project Thawing of Permafrost: The video [<http://www.opengeospatial.org/pub/ArcticSDP/blog-modelingForecasting.html>] demonstrates the value of a Discrete Global Grid System (DGGS), a data-agnostic information grid that allows users to acquire and integrate data from various sources at any location on a virtual globe and at any resolution. The data sources, which can be real-time or archived, include any geospatial information source, such as earth observation data, social and economic data and/or statistics, as well as conventional GIS feature and coverage data. In particular, the overlay of socio-economic data with earth observation data is a powerful feature that helps in understanding and optimizing many processes on earth.



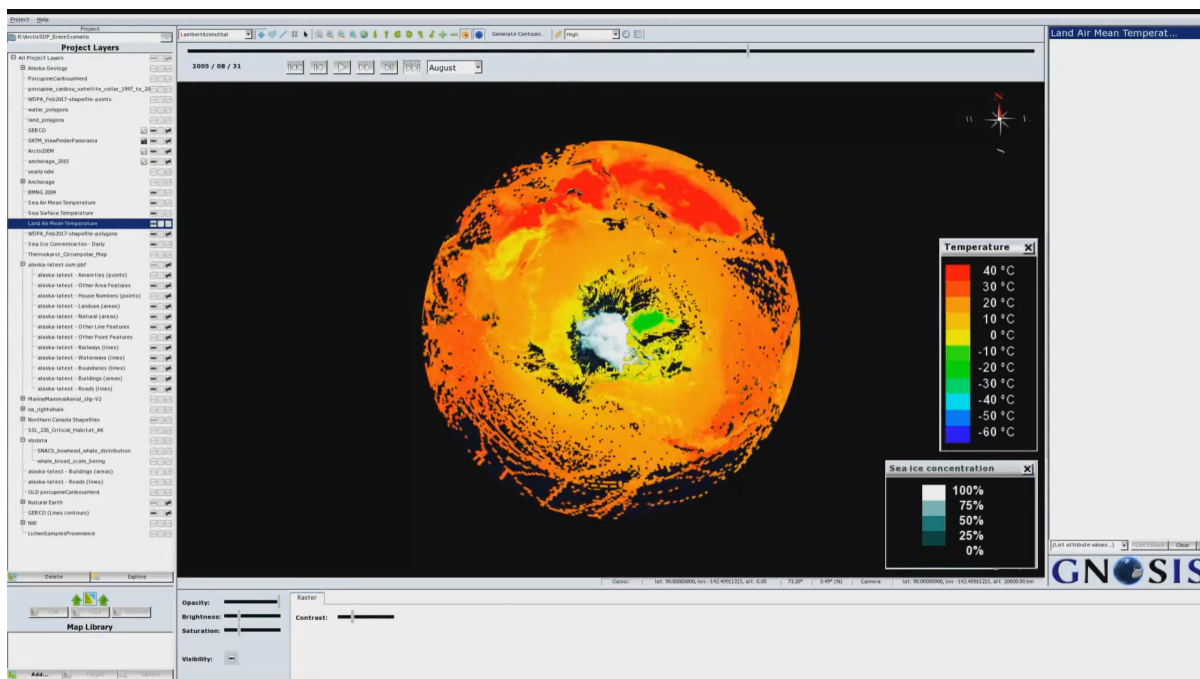
The DGGS allows users to calculate new data values as expressions, to select and refine data values to answer questions such as "Where is it?" and to aggregate data values to answer questions such as "What is here?"

All data is accessed through OGC Web Feature Services or OGC Web Coverage Services. Both services provide actual data, preserving all attributes of the original data sources. This allows rich calculations and analysis directly in the client application. The DGGS fuses all data sets on the fly, independent of the resolution, location, or original map projection.

In the video, the change in permafrost distribution based on the assumed temperature change is forecasted for the coming decades. Starting with the actual distribution of permafrost, the future extent is calculated and the visualized using time sliders. Eventually, the potential consequences of climate change are emphasized by overlaying with socio-economic and wildlife data, showing that critical infrastructure and migration paths are at risk.

4.4. 3D Data Visualization & Temporal Patterns

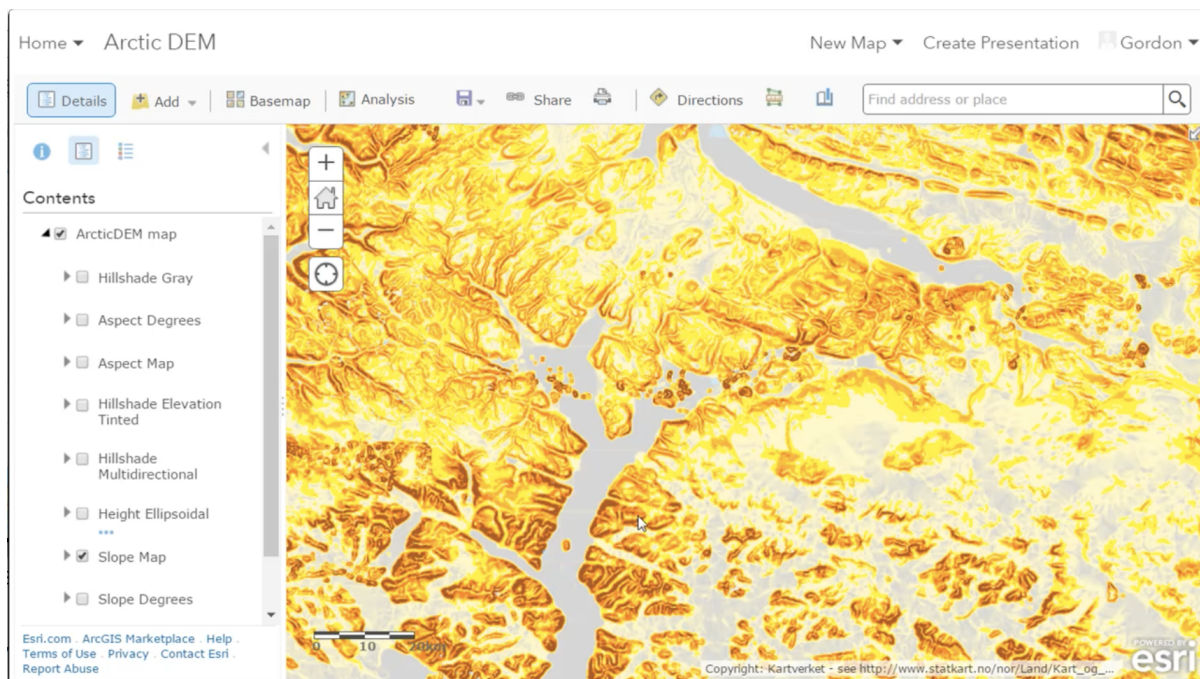
3D Data Visualization & Temporal Patterns: New Ways of Data Exploration and Visualization: The [video](http://www.opengeospatial.org/pub/ArcticSDP/blog-3dVis.html) [http://www.opengeospatial.org/pub/ArcticSDP/blog-3dVis.html] shows the Porcupine Caribou Herd's migration patterns, which have been overlaid with topographic and climatic information in a 3D environment.



Modern spatial data infrastructures are not restricted to two-dimensional representations anymore. As demonstrated in this video, current standards support full 3D and even 3D+time visualizations and analysis of spatial data.

4.5. Food Security in the Arctic

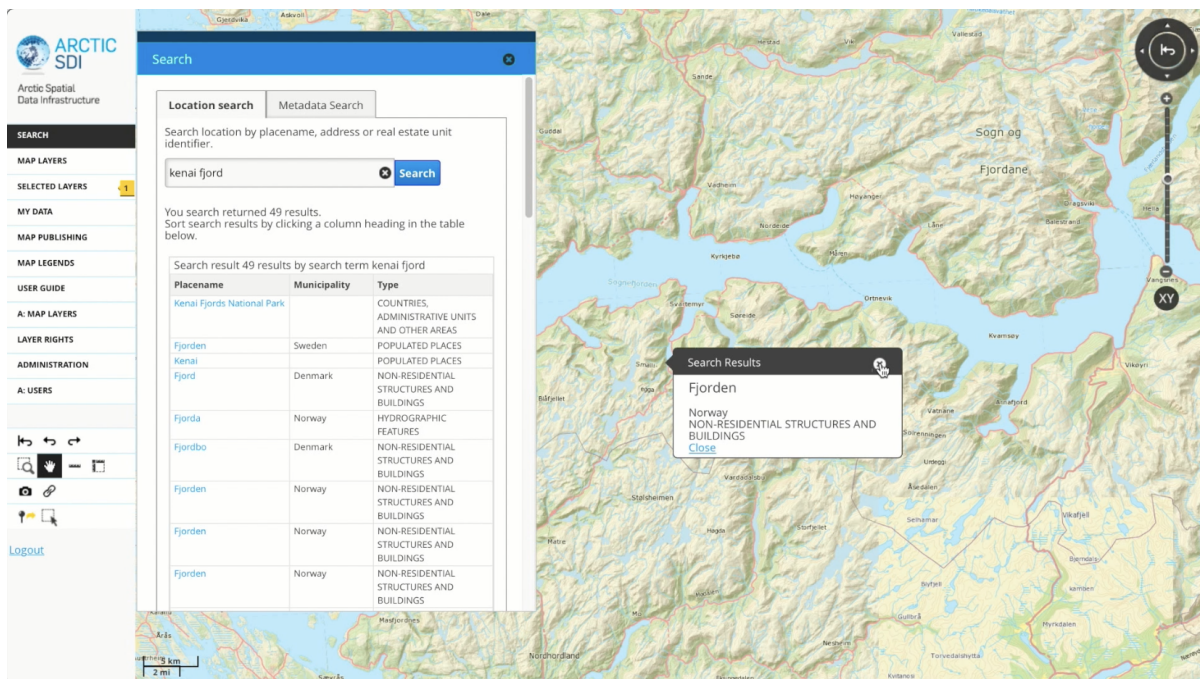
Food Security in the Arctic: Food Security Policy Demonstration Policy Workbench: “Food security is a significant issue in the Arctic and the principal challenges to food security across the Arctic are: high cost often coupled with economic vulnerability and decreasing consumption of country foods. Exacerbating these challenges are major issues linked to: contaminants and climate change.” - Food Security across the Arctic, Background paper of the Steering Committee of the Circumpolar Inuit Health Strategy Inuit Circumpolar Council – Canada, May 2012.



Governments and NGOs are continually assessing and monitoring the situation to ensure a sufficient food supply for Arctic residents. Food security policy depends on the sharing and interoperability of spatial data. For this pilot, a demonstration policy workbench was developed which allows policy workers to monitor issues, assess situations, analyse data and review progress. The workbench allows easy addition of issues and simple accumulation of interoperable data sources - all within a web browser, as shown in the [video](http://www.opengeospatial.org/pub/ArcticSDP/blog-foodSecurity.html) [http://www.opengeospatial.org/pub/ArcticSDP/blog-foodSecurity.html].

4.6. Arctic SDI

Arctic SDI: Functionality & Sustainability: Demonstration of the Arctic SDI Geoportal: The [video](http://www.opengeospatial.org/pub/ArcticSDP/blog-arcticSDI.html) [http://www.opengeospatial.org/pub/ArcticSDP/blog-arcticSDI.html] shows a demonstration of the Arctic SDI Geoportal an open source platform that adheres to Open Data Principles leveraging distributed SDIs based on standards. The National Mapping Agencies of the eight Arctic Council Member countries are working in cooperation to facilitate access and use of Arctic data to and from local, national, regional and global stakeholders.



The video illustrates both the importance and capabilities of authoritative data services provided by National Mapping Agencies of the eight countries adjacent to the Arctic. It illustrates how common standards allow the integration of data independent of national borders, local projection systems, or other national particularities.

The Arctic SDI Geoportal allows the rapid generation of customized maps that can be integrated into any web site. Thanks to the underlying standards and infrastructure, these apps auto-update themselves when new data becomes available.

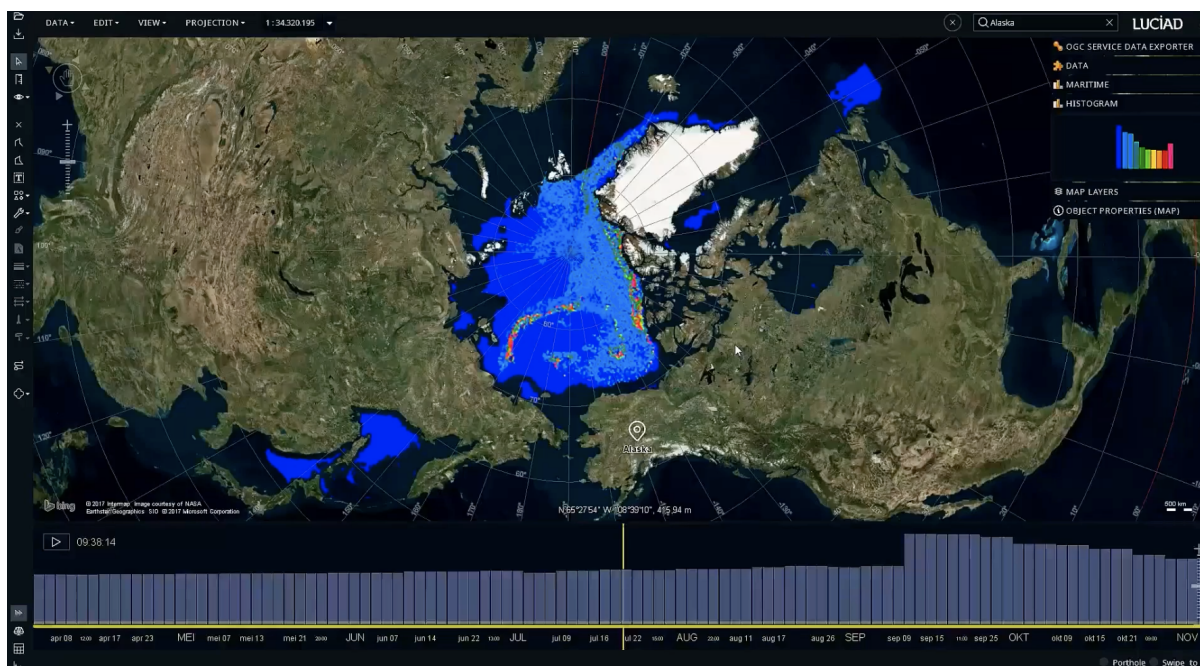
4.7. Complex Data Analysis

Complex Data Analysis: Modeling Land Susceptibility to Failure due to Permafrost Loss: In the [video](http://www.opengeospatial.org/pub/ArcticSDP/blog-landslides.html) [http://www.opengeospatial.org/pub/ArcticSDP/blog-landslides.html], a landslide susceptibility study, as previously implemented by National Resources Canada, is executed using state of the art data infrastructures on the Web. Standards of interoperability make it extremely simple to discover, access, and use data from arbitrary sources, and a study that previously took days is now executed within minutes. The video demonstrates the full five-step process, just this time executed within a Discrete Global Grid Client accessing data directly via the Web.



4.8. Sea Ice Age Evolution

Sea Ice Age Evolution: Beaufort Gyre: The Evolution of Sea Ice Age and Extent North of Alaska: Arctic Sea Ice Age measurements show that the sea ice is becoming younger. Since the 1980s, the amount of multiyear ice has declined dramatically.



The video illustrates the evolution of sea ice in the entire Arctic over a period of twenty years. Using a swipe controller, the Beaufort Gyre area just north of Alaska is analyzed in closer detail, highlighting changes in the sea ice extent and age by comparing 1995 and 2015.

This case study highlights how data available at standardized Web service interfaces can be

accessed dynamically from a client application. Maps served as image files can be overlayed with actual data stemming from data services such as OGC Web Feature Service (WFS) and OGC Web Coverage Service (WCS). Thanks to the standardized Web service interfaces and data encoding models, data can be integrated and analyzed within minutes. Elements such as swipe controllers allow new ways of data exploration and visualization without additional complexity.

Chapter 5. Arctic SDI

The Arctic SDI is a voluntary, multilateral cooperation between the National Mapping Agencies in the Arctic:

Earth Sciences Sector of the Department of Natural Resources Canada	Norwegian Mapping Authority
Danish Geodata Agency	Federal Service for State Registration, Cadastre and Mapping of the Russian Federation
National Land Survey of Finland	Swedish Mapping, Cadastral and Land Registration Authority
National Land Survey of Iceland	United States Geological Survey

*"The **Arctic Spatial Data Infrastructure initiative** [<https://arctic-sdi.org>] brings together geospatial experts and scientists in a voluntary cooperation between the eight national mapping agencies of the Arctic countries in direct support of the priorities of the Arctic Council and other important stakeholders. The purpose of the Arctic Spatial Data Infrastructure is to support the Arctic Council and other users and stakeholders in meeting their goals and objectives by using reliable and interoperable geospatial data of the Arctic, accessible via the Arctic SDI Geoportal."* (**Arctic Spatial Data Infrastructure 2015–2017 BIENNIAL REPORT, June 2017** [<https://arctic-sdi.org/wp-content/uploads/2017/06/20170621-Arctic-SDI-Biennial-Report-small.pdf>])

The OGC Arctic Spatial Data pilot was motivated to support the Arctic SDI and its strategic objectives as described in the Arctic Spatial Data Infrastructure (Arctic SDI) Strategic Plan. The plan *"contains a high-level overview of the background, organization and philosophy of the Arctic SDI, and it provides a mid-range vision identifying the primary strategic objectives of the Arctic SDI over the five year span from 2015 to 2020"*. (**Arctic Spatial Data Infrastructure Strategic Plan, 2015-2020** [https://arctic-sdi.org/wp-content/uploads/2014/08/20151119-Arctic-SDI-Strategic-Plan-2015-2020_FINAL.pdf])

Objective	Objective Description	Primary Arctic SDI Working Group
Objective 1	User and Stakeholder Needs and Requirements	Strategy Working Group
Objective 2	Reference Datasets	Technical Working Group
Objective 3	Thematic Datasets	Technical Working Group
Objective 4	Data and Technical Interoperability	Technical Working Group
Objective 5	Spatial Operational Policies	Operational Policies Working Group
Objective 6	Communications	Communication Working Group

Figure 1. Arctic SDI Strategic Plan 2015-2020 objectives including the primary Arctic SDI working group responsible for implementing the actions described in the Arctic SDI Strategic Plan Implementation Plan.

The pilot addressed the **strategic objectives** at various levels of detail with initial focus on objectives 1, 3, and 4. It is one of the important **lessons learned** that these objectives need to be combined with a revised communication paradigm in order to leverage the full potential of a Spatial Data Infrastructure for the Arctic. The full set of lessons learned and conclusions related to the Arctic SDI strategic goals are available in sections **Lessons Learned** and **Conclusions**. The most important elements include:

- If data is well documented and provided at a standardized interface, then integration into various client components or data processing workflows is extremely efficient.
- There are some specific requirements that differentiate an SDI in the Arctic from other SDI, in particular the limited telecommunication resources in the North, as well as the difficulty to integrate Indigenous knowledge with scientific or sensor based data.
- The amount of available data is impressive, discovery and access of data is still difficult. Data not being delivered via standard services, the lack of consolidated catalogs, the lack of any data or service annotation mechanism combined with missing entry links into aggregating catalogs, and almost no relevant inventory of available services or data produce high entry hurdles for any type of scenario.
- Many high value data sets are archived and not available via Web services.
- Despite the fact that all ArcticSDP participants have been experts in geospatial data and services, access to, and use of, fundamental layers such as base maps wasn't clear.
- The value of portals such as the Arctic SDI Geoportal that hide a lot of the complexity from the user and provide relevant data sets and base maps right away cannot be overestimated. Otherwise, the different spatial projections used in the Arctic, paired with the missing support for high latitudes of frequently used tools such as Google, Bing, or Open Street Maps cause major challenges that cannot be handled by the average data user.

- Metadata is still a major challenge. It is often missing or incorrect, which leads to long data discovery processes as individual data sets often need to be checked for applicability in a long and expensive process.
- Access to data and the lack of single sign-on to the various portals combined with often unclear, or very restrictive data usability regulations, prevent the use of many data sets. Most data portals required some sort of registration before access to data was granted.
- Quality of data and quality of service aspects need to be addressed in future. This includes links from one data set to another, or one service to another or to other data.
- To efficiently use research time and development funds and simplify the sustainability of pilot results, a repository of real world use cases containing point of contact data, would allow pilots such as ArcticSDP to build on previous work and address more real issues.

Chapter 6. Lessons Learned

The **Lessons Learned** chapter describes experiences made during the implementation of the various scenarios. In parts, these experiences describe general aspects of data handling in SDIs, while other parts indicate rather low level technical details that require further attention in order to improve the various users' experiences with SDIs.

It has to be stated that many data sets and data services are currently available for the Arctic. Catalogs provided by e.g. the **Arctic SDI Geoportal** [<https://geoportal.arctic-sdi.org/>], **Polar Hub** [<http://polar.geodacenter.org/polarhub2/>], or the **Canadian Polar Data Catalog** [<https://www.polardata.ca>] provide hundreds of entries with services serving thousands of layers for different types of data (though ease of use to access the data needs to be improved). Thus, there is a solid basis of Web services already available that future efforts can build upon. Nevertheless, many data sets are still not directly available, need to be extracted from *pdf* documents, or need to be scraped from HTML Websites. There are certainly reasons to protect data against inappropriate usage, but many data sets are locked up for other reasons that certainly not align with Open Data and FAIR principles. **Wikipedia** [https://en.wikipedia.org/wiki/Open_data] describes the Open Data principle as the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. FAIR addresses one of the grand challenges of data-intensive science, which is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows; according to **FORCE11** [<https://www.force11.org/group/fairgroup/fairprinciples>]. FAIR is a set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable.

The pilot used a scenario-driven approach to experiment with the current landscape of data and services for the Arctic. The actual effort to integrate the data, once made available behind OGC Web services, made up only a small fraction of the full efforts to discover, access, and integrate the data. This shows that the underlying principles of data advertisement, discovery, access, and conversion need to be further developed, whereas standardized access interfaces and data models work well already.

In the following, we concentrate on elements that need to be improved in order to make the full circle of data search, discovery, access, integration, processing, and presentation a smoother user experience.

6.1. Architecture

In general, the SDI architecture with its service composition, transport protocols and information models is mature and allows rapid setup of new or extended data and data processing services. Just, the underlying principles of service oriented architecture with

XML remote procedure calls and publish-find-bind interaction pattern has room for improvements in real world implementations. The following issues have been identified:

- **Shared vocabularies:** Any data owner needs to register the data/data access service at a catalog in order to allow clients to find the data/service. Without well established vocabularies, data providers and data consumers often speak different languages. If a data provider uses keywords different from what a client expects and uses for discovery, data provider and data consumer don't come together. An important part of the solution is shared vocabularies that can be used by both sides. Currently, there are a number of efforts under way to setup online vocabularies. The challenge is to find the ones that provide the best service to the Arctic community. Here, governance patterns and policies are important in addition to the actual content. It is recommended that the Arctic SDI community address this issue in close detail.
- **Never enough and too much data:** With the growing number of data and data processing services, discovery of relevant services becomes more and more a challenge. Here, new mechanisms need to be found that help consumers to discover relevant services/data sets more easily. The traditional approach with orchestrated catalogs seems to be too slow to catch up with the ever changing Internet environment. Indexed services and data sets that are not available at runtime are an important point of frustration for the user. From an architectural point of view, new discovery mechanisms need to be explored, taking into account community-based and automatic annotation mechanisms. Community-based approaches make use of user feedback that is aggregated and made available in some meaningful way. It allows others to benefit from experiences made by colleagues. Automated annotation systems could be based on linked data paradigms that help to describe links between publications, processing steps, and input data.
- **Rapid exploration:** Often, data providers provide access to their data as data services, but fail in providing rapid overview mechanisms that help clients during the decision process if a given resource is useful in the client's context. The importance of guidance and tools to help data providers to provide complete and up-to-date metadata would be very helpful.
- **Missing Digital Rights Management:** The lack of any efficient digital rights management across data providers and consumers has led to a high number of data portals that all follow their own rules and policies. Often, data consumers need to register at portals in order to get access to data. Here, portals use different registration procedures and technologies that prevent interoperability between the used components. Once access is granted, the path from the portal entry point to the actual data varies from portal to portal.
- **Changing Access Paradigms:** The service oriented paradigm with rich Web services that basically offer interfaces for remote procedure calls over HTTP(s) is more and more replaced by RESTful services that are entirely built on the limited set of HTTP(s) methods (usually HTTP GET and POST). This leads to a mixed access path situation

that complicates client-server communication. Adding linked data principles to RESTful services that allow traversing data sets by following heterogeneous link concepts adds additional complexity. Here, guidebooks are required that help in particular the data provider side to consolidate the new approaches.

- **Changing Data Encodings:** Whereas XML was the preferred data encoding model for many services developed in the late 90ies/early 2000s, data is now more and more serialized using JSON. Without discussing the pros and cons of the two approaches (that almost certainly will be complemented with additional approaches in the near future anyway), it needs to be stated that the growing number of serialization formats adds complexity to the data exchange that is not fully covered by technology yet. Hopefully, off the shelf software implementations will handle these differences in a way transparent to the user.

6.2. Data

6.2.1. Missing Metadata for OGC Web Services Content

The use of OGC Web services is well-adopted by the geospatial industry and community. However, it was observed that served data often lacks proper metadata, which makes it difficult to interpret the services' offerings. For instance, many WMS layers use default or empty titles, abstract, keywords, etc. This makes it difficult for catalogs to help clients with their data search. Also, often only the service provider is mentioned in the metadata. The original data provider is missing, which causes problems for proper citations.

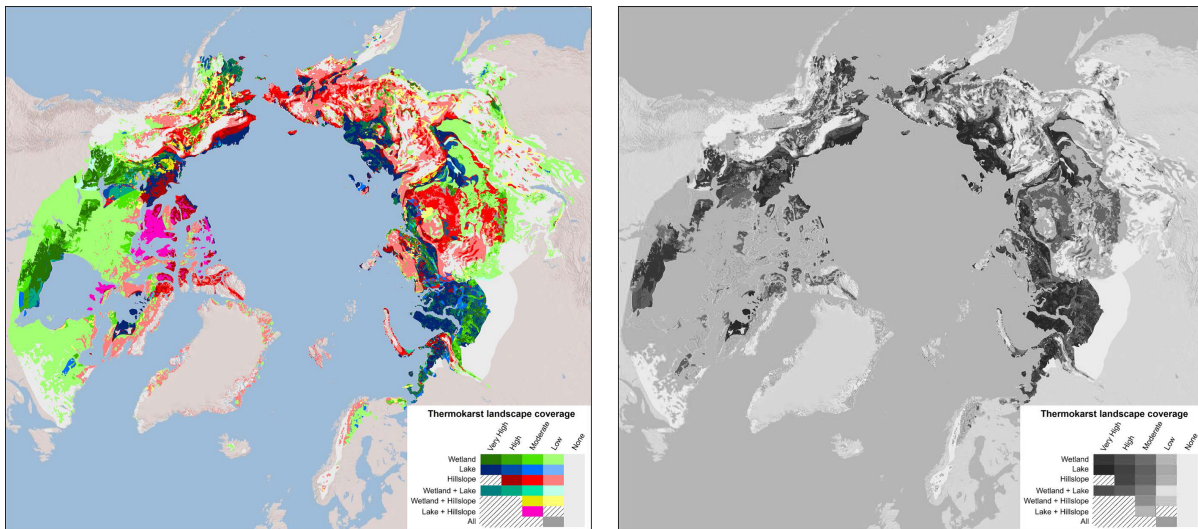
6.2.2. Data Formats

Data integration can be time consuming in case of proprietary/custom formats. This situation is often observed at portals that feature a more FTP-like data access rather than a Web service with rich query interface. As an example, the NSIDC Website offers sea ice age data for the Arctic region, covering the impressive time span of 1984 until now. This data set is stored using a custom binary format. Although it is a simple format, additional development time is required to integrate this data into applications. Additionally, the temporal dimension is not modeled in the data itself; instead, the file name is used to indicate the time instant (year and week). Though this is in principle a mechanism that is easily understood by humans, it prevents automated processing and requires humans to manually control the data integration process. Using an open, interoperable standard with support for temporal dimensions (e.g., NetCDF, OGC WCS) avoids custom development tasks related to the integration of these data.

6.2.3. Styling of Vector Data

Vector data is often made accessible using a vector format that doesn't contain any styling information - such as a **CSV file** [https://en.wikipedia.org/wiki/Comma-separated_values] or an

ESRI Shapefile. While an application can read such a file relatively easy, it will typically be displayed with a default (application-specific) style. However, having a meaningful style greatly helps to interpret the data. An example is the **Thermokarst data served by ORNL** [https://daac.ornl.gov/SOILS/guides/Thermokarst_Circumpolar_Map.html] illustrated below, which consist of a set of polygons indicating the amount of Thermokarst landscape coverage per landscape type. A visualization of the data only makes sense with proper styling instructions - such as shown in the map below left. The same map in black and white becomes pretty much useless, in particular as a two-dimensional color coding pattern has been applied (different colors and different levels of color saturation).



The OGC Symbology Encoding standard is one example standard that can solve this problem: being a stand-alone styling definition language, it is ideally suited to be created by a vector data provider and shared with data consumers - possibly through a registry / discovery service such as an OGC CSW.

6.2.4. Temporal Characteristics

To analyze the evolution of some characteristics, the temporal dimension is mandatory. It is usually referenced as the 4th dimension after x,y and z. The temporal dimension represents snapshots of the data at different points in time. Management of time in data has some impacts, the major one being the size of the dataset. For a "picture" of an area at one moment occupying a size of X, a monthly update of the data will occupy 12 times the size every year. Beyond the acquisition and storage challenge, the distribution of spatio-temporal datasets is not always easy. As explained above, some standards data formats like NetCDF and Grib are suited for multidimensional data. For raster data, most of the time, the data is organized following a specific directory or filename structure to represent the temporal dimension. The main reason is that often multiple acquisitions are not merged to be stored in a single file container. As far as distribution of temporal data is concerned, OGC standard completely fulfill the requirements for all of its Web services (WMS, WMTS, WFS, WCS).

6.2.5. Vendor Specific Solutions

A number of data sets have been discovered that provide RESTful service interfaces that are based on open standards and which are not OGC standards. While that is commendable, we understand that it is generally simple for the data provider to also provide standardized OGC Web service interface support for e.g. WMS or WFS. With so many of these services deployed across the globe that fail to enable OGC service interfaces, it is a lost opportunity for the data provider to increase exchange of their information.

6.2.6. Open Data and Data Access

We've come across a number of data sets which would have been interesting to work with (e.g. various caribous herds, belugas), but after a lot of time and communication we only managed to get access to a single caribous herd data set, late in the project. The trend of open data should be encouraged because it maximizes usability. The increased availability makes it easier for scientists and decision makers, who may be rightfully intimidated by the tedious process to obtain the data, to quickly correlate multiple data sets. If the data is sensitive, open access to a limited or out-of-date subset could be considered. This would allow stakeholders to quickly visualize or otherwise analyze the dataset to determine fitness for use or purpose, which could then lead to negotiation of terms of use based on information provided in a full and complete metadata record following international standard models would greatly help, as terms or usage are defined there.

"Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control." (Wikipedia) [https://en.wikipedia.org/wiki/Open_data]. Open Data needs to be combined with direct access to the data. Experiences have shown that it can be time consuming (order of weeks) to talk to the right parties, identify the correct / desired data set, and get the actual data delivered / made available. Then if the data is delivered in a non-standard format there is additional work to ready the data for use. This is in strong contrast with the experiences made getting the resulting data into an OGC service or client used within the Pilot: This is typically in the order of seconds / minutes (in case it is a standardized format). Lots of data found or referenced in sources used in the pilot are only available in reports or in data files that need to be downloaded and further processed. In addition, it is very difficult to find pan arctic data. There is data available for individual countries in the Arctic, like Canada and the US, but no one has aggregated this data for pan-Arctic use. Here, pan-Arctic efforts such as the [Arctic SDI Geoportal](https://geoportal.arctic-sdi.org) [<https://geoportal.arctic-sdi.org>] that are built for browsing, visualizing, analyzing, and sharing distributed geographic information for full Arctic region play an important role. Key is that these efforts adhere to Open Data principles leveraging distributed spatial data infrastructures and making extensive use of services based on OGC standards. Ideally, international metadata standards are supported. As an example, the [Arctic SDI Geoportal](https://geoportal.arctic-sdi.org) [<https://geoportal.arctic-sdi.org>] uses ISO standard 19115 Topic Categories and search to provide access to thematic and reference data layers.

6.2.7. Shared Semantics & Quality Information

Today, geospatial data is collected, processed, and used in domains as diverse as hydrology, disaster mitigation, spatial planning, statistics, public health, geology, civil protection, agriculture, nature conservation, and many others. The challenges regarding the lack of quality and quality information, often combined with heterogeneous data production processes, policies, and applied semantics to technical terms and processes are common to a large number of policies and activities, and are experienced across the various levels of public authority.

It is easier to reuse spatial data when information about their quality and fitness-for-use is available, and when technical and legal barriers for integrating these into the user systems are removed. The first condition, quality, requires that rich and meaningful metadata be used, while ‘fitness for use’ requires the involvement of technical arrangements that ensure interoperability.

Semantic issues in spatial data sharing and service interoperability have been recognized in the literature for a long time. Bishr [<http://www.tandfonline.com/doi/abs/10.1080/136588198241806>] summarized interoperability issues in 1998 under the terms semantic heterogeneity, schematic heterogeneity, and syntactic heterogeneity. Though the latter two have been addressed pretty successfully with GML [<http://www.opengeospatial.org/standards/gml>] and OGC Web service interface standards [<http://www.opengeospatial.org/standards>], semantic heterogeneity still causes several problems. These include

- discovery of data sets and services based on keywords,
- rigid metadata structures,
- missing semantics on technical terms,
- missing matching capabilities for equivalent or related terms or symbols

It appears that the Semantic Web is still in its infancy, in particular if extended to data. Nevertheless, a key concept of the Semantic Web is the usage of URI as identifiers for objects, predicates, and subjects. If, at least, URIs would be used for keywords, discovery and usage of data for the Arctic would already be largely improved.

6.2.8. Aggregation and Data Fusion

Collaboration between organizations (e.g. national mapping agencies) should be encouraged to build aggregated data sets. Tremendous value is created when the best data sources are unified together in a single data set which can cover a large geospatial extent while still offering the best data resolution. This can result in one global source which can benefit from all authoritative updates and be the go-to source for a given data type, making it easy to find the best quality data. Data fusion steps help to efficiently integrate a large amount of small files

First: One of the Arctic data sources consisted of elevation data for Alaska (Anchorage), represented by 2800+ ERDAS Imagine elevation files each representing a very small region. A data fusion step to combine everything in one logical dataset followed by the use of an OGC Web Service helps to ease the integration of the data in applications. In this particular example, a WCS and WMS were used to respectively access the raw elevation data and a rendered version of the elevation data; in case of WMS, GetFeatureInfo was enabled allowing users to access an individual elevation data sample.

Second example: The ArcticDEM was made available as 1441 GeoTIFF files, each covering a small area. Again, a data fusion step to combine everything in one logical data set and corresponding Web service helps to ease the integration of the data in applications. Loading 1441 files separately is just inefficient.

Third example: Sea ice age from **NSIDC** [<http://nsidc.org/data/docs/measures/nsidc-0532/>]. The fact that the data is available in a standardized format (OGC NetCDF) is very convenient. However, the data is spread among thousands of files in a time period between 1970s and 2012. Although access through FTP and HTTPS is offered, some difficulties were encountered to download the data in bulk.

6.3. Communication

6.3.1. Discovery & Rapid Exploration

Data Discovery is becoming an ever greater challenge as the volume of geospatial data, and the variety of sources, increase. SDIs may attempt to aggregate metadata to provide discovery services cross diverse data collections, but these will never be complete, as new data sources are constantly coming online. What is needed, in addition to centralized catalogs, is an infrastructure that can locate data wherever it is on the Internet. This can include, in addition to formal repositories, institutional or investigator Websites, publication-related repositories, self-advertising sensor Webs, and the like. This argues for development of technologies that can crawl the Web, locate and interpret data and then permit intelligent queries against the found content. The US National Science Foundation Project Polar Deep Insights (<http://polar.usc.edu/html/polar-deep-insights/>) is pursuing such a technology, using open-source big data tools.

Web mapping services are great for quick visual exploration, and can quickly highlight unanticipated questions needing further analysis. However, bulk data download is still very valuable and should continue to be an alternative provided, as it is ideal to perform offline analysis or heavy processing and optimizations (e.g. tiling and generalization).

6.3.2. Broken Links and Portals

Searching for data, one often encounters discontinued or moved pages, and the link to the

new data location is rarely to the exact data set but rather back to yet another portal where the search must start over. Often after many steps, you end up in a dead-end. By keeping alive Web locations by fully redirecting to the dataset (service end-point or download link), or at least by using unique and stable Digital Object Identifiers that could be used in searches, it could be a lot easier to find data. Although OGC catalogs serve a purpose, they are limited to the entries they contain, while search engines like Google remain the easiest way to find a data set over the entire Internet. When the top result for a particular search leads nowhere, it's a hindrance to locating useful data sets.

Geospatial portals do not always provide basic filtering capabilities for finding relevant data sets, including: nature of the data (e.g. imagery, coverage, vector), geospatial extent, data resolution, temporal extent, service type or whether the data set is available for bulk download, in addition to the obvious topic & keywords.

These should be standard in any geospatial portal interface and would avoid painful scrolling through pages of hundreds of irrelevant results. Some portals will also return publications results with no associated GIS data available, or which will require extra steps such as sifting through the publication itself to locate the actual related data set, which makes it even worse.

6.3.3. Web service availability of registered dataset

The metadata of more than 100 thousand datasets (e.g., sea ice, sea ice age, NDVI) were registered in the GMU catalog. Each metadata record contains the endpoint for accessing the real dataset from archiving catalogs. Nevertheless, users occasionally experience difficulties in discovering datasets that are available through OGC Web services and ready for directly overlaying/visualizing in desktop/Web based clients.

The main issue is with the resource's availability. The dataset access endpoints, harvested from the distributed catalogs, serve as the endpoint for downloading raw data (in native formats such as HDF or GeoTiff). Since there often doesn't exist a Web service (e.g., OGC WMS or OGC WMTS) provided by either original data catalog or the third-party Web service provider to explore the data, users need to manually visualize the data. This requires full downloading of the raw datasets and often additional local pre-processing steps.

It is recommended that future catalogs feature an indicator system that illustrates data accessibility and - ideally - its fitness for use. The following ranking of data accessibility is proposed as a first step. The indexed data accessibility indicator will facilitate users in distinguishing and searching dataset which are directly available through a Web API.

ID	Data Accessibility
1	Dataset is directly available through OGC Web services (e.g., WMS, WCS, WMTS) and ready for client-based visualization

ID	Data Accessibility
2	Dataset is directly available through standards from non-OGC standards bodies like ISO, W3C and IETF (e.g, OPeNDAP, FTP, HTTP)
3	Dataset is available through open or community standard Web APIs (e.g, RESTful API, Shapefile)
4	Dataset is not available for directly access (e.g., only available for order with credentials requirement)

6.3.4. CORS - Cross-origin resource sharing

"Cross-origin resource sharing (CORS) is a mechanism that allows restricted resources (e.g. fonts) on a web page to be requested from another domain outside the domain from which the first resource was served." (Wikipedia) [https://en.wikipedia.org/wiki/Cross-origin_resource_sharing]

We experienced a number of WMS that had CORS disabled; thus preventing any view of the map image in a Web browser located on a different domain, unless the Web applications access other servers through a proxy.

In general, Web service administrators should be informed if their services prevent cross-origin resource sharing, as this is a serious issue in spatial data infrastructures, where clients often load an number of elements in cross-domain approaches.

6.3.5. HTTPS

HTTPS should be supported for all Web services.

6.3.6. Denied, Degraded, Intermittent, or Limited bandwidth

Not all systems reside on the World Wide Web or are served through powerful connections. Many are hosted on Denied, Degraded, Intermittent, or Limited bandwidth (DDIL) communications environments. This pilot has demonstrated that **OGC GeoPackage** [<http://www.geopackage.org>] can serve as a powerful solution for those situations, as proper loading and synchronization capabilities are available.

6.3.7. Projection Issues

The primary aim of a logical choice is to select a projection in which **the extreme distortions are smaller than would occur in any other projection used to map the same area** [http://www.iho.int/mtg_docs/rhc/ArHC/ArHC3/ARHC3-3.2.7_Suitable_projections_for_the_Arctic.pdf]. Unfortunately, many tools and Web service instances that are serving data for the whole planet use, despite the large distortions in Arctic areas and angles, Mercator and other non-appropriate projections, whereas Arctic

Region is best served using Azimuthal Polar Equidistant projection and the Azimuthal Polar Stereographic projection.

Most clients overcame the requirements for polar projection by eliminating the need to display data in a projection and instead used a globe. This may indicate a positive trend not only for polar analysis but also over the Earth generally as these technologies catch up to the performance of Web Mercator products. However, we did run into the problem that many common data sources and APIs assume Web Mercator and, due to difficulties capturing data at the poles, lack polar coverage (ie hole at the pole).

In other situations, clients and servers did support some Arctic projections, but not all suitable ones, which caused labels being displayed upside down in some situations. It is recommended that clients and servers shall support at least the EPSG codes 3571 to 3576, 3413, 32661, and 104306.

Clients which perform their own runtime reprojection, often including those dealing with 3D coordinates (e.g. Ecere's GNOSIS system), have a preference for unprojected lat/lon coordinates. Therefore support for WGS84/EPSSG:4326 should be considered important.

6.3.8. Raw Data vs. Maps

There continues to be a trend of serving map products as RGB encoded final map views only instead of the valuable data values that underlie the data, i.e. WMS/WMTS instead of WFS or WCS. This prohibits secondary analysis between data sources. Having raw raster data access in a client enables capabilities beyond visualization: applications can query values, apply analytics, perform transformations, etc - in contrast with having map access. The most widespread OGC service for raster data is however WMS, which provides rendered maps; the protocol does provide an optional capability to get the actual data value for a given location (GetFeatureInfo). The primary OGC Web service candidate for raw raster data access is WCS: preferably, a raster data source is offered both through a WMS and WCS, so that applications can choose what to load depending on the desired functionality.

6.3.9. Need for Catalogs

A catalog can greatly enhance the search for data, of course, if and only if metadata are complete and correct. Without extensive and accurate metadata, a catalog is often useless. If projects decide not to register their data and service offerings at open catalogs, then a local catalog shall be set up that harvests all local data sources and services. This catalog then needs to be made available and known to the public. Often enough, catalogs are barely discoverable because they are not indexed by search engines if not properly described on an HTML page. Known catalogs compliant with international standards can be easily integrated into a federation system that allows forwarding search queries to all federation members.

Federations can be implemented using a periodic harvesting approach, where information from other catalogues is stored in a master catalog, or by a dynamic federated search where the master catalog dispatches the query to other catalog instances. This process can be done in a smart way. For example, if the master catalog is queried for metadata on a spatial extent that includes only some regions, only catalogues serving data for those regions can be queried.

6.3.10. GetCapabilities Issues

Scalability of OGC's GetCapabilities request

The GetCapabilities request is defined in the OGC OWS Common specification. The purpose of this operation is to publish the capabilities of the service. The GetCapabilities request is used throughout OGC Web service specifications as a first entry point to discover the offered data sources. By definition, the response lists all offered data in a single XML document - which can get quite big in cases with lots of data. An example encountered in the Arctic Pilot is the WMS at <http://spatial-dev.ala.org.au/geoserver/wms?service=WMS&request=GetCapabilities>, which returns a capabilities XML document of 15MB with the response time about 6 seconds.

Extended capabilities according to capability extension schemas sometimes lack the extension schema

Clients that apply schema-based parsing need valid schemas to read XML. They are well-known and often cached in clients for standards like OGC WMS, WFS, or WCS; although when extensions are used, it is necessary to have access to the extension schema.

6.4. Experiences made setting up the ArcticSDP Catalog

The aforementioned aspects have been experienced across many scenarios. Most of them are reflected in the following overview of issues that were collected by the Pilot's catalog provider George Mason University (GMU). We kept this summary to allow a consolidated view from a dedicated perspective, even though some elements might be repetitive.

6.4.1. Catalog Introduction

The geospatial catalog hosted by GMU serves as the catalog portal for ArcticSDP project. The roles of GMU catalog:

- In the front end, the GMU geospatial catalog serves as the standard geospatial data discovery portal for data clients. The geospatial dataset from distributed catalogs is served through heterogeneous discovery protocols (e.g., OPeNDAP, vendor specific

RESTful API). The standard interface (CSW 2.0.2) as well as OGC Essentials API provided by GMU catalog improve geospatial dataset discovery through standardizing the API for spatial/temporal search. This frees data clients from dealing with heterogeneous API during data discovery.

- In the back end, the GMU geospatial catalog serves as the harvester periodically harvesting arctic geospatial data from distributed catalogs through heterogeneous discovery/harvesting interfaces (e.g., OGC CSW, WMS GetCapabilities, WCS GetCapabilities, WFS GetCapabilities, OpenSearch OSDD, etc). This improves the visibility of datasets archived across distributed catalogs.

6.4.2. Catalog harvesting

The GMU catalog harvests metadata from distributed geospatial service repositories and geospatial catalogs. The harvesting implementations, interoperability issues and solutions are summarized as follow.

6.4.3. Implementation for harvesting metadata from OGC catalog server

The easiest harvesting occurred for harvesting other OGC catalog servers, as the [OGC CSW API](http://www.opengeospatial.org/standards/cat) [http://www.opengeospatial.org/standards/cat] provides direct support for harvesting. If the OGC CSW API and corresponding metadata information models (e.g., ISO19115, Dublin Core) are implemented, then the least configuration and implementation (e.g., metadata conversion logic) is required for the harvester implementation.

6.4.4. Implementation for harvesting metadata from OGC Web server

Metadata is harvested from distributed geospatial servers (e.g., WMS, WCS, WFS) through standard GetCapabilities API/document. The GetCapabilities API/document serves as the base interface/information model on which the capabilities document of other OGC services (e.g., WMS, WCS, WFS) is designed. However, the implementation of GetCapabilities (e.g., GML version support, spatial/temporal encoding) may vary across different servers. The following customization is implemented to make the metadata conversion logic more robust:

- Support for multiple-version GML encoding
- Logic to process spatial footprint encoding in harvested metadata and the logic to verify the correctness of spatial encoding
- Logic to process temporal encoding in harvested metadata and the logic to verify the correctness of temporal encoding (e.g., start date \Leftarrow end date)
- Implementation for harvesting metadata from the catalog supporting non-OGC Web APIs In this case, either the harvesting interface or the metadata information model is

heterogeneous across distributed catalogs. The following development is need to implement the metadata harvester:

- Implement Web service client based on the Web API (e.g, OpenSearch) supported distributed catalogs. Especially, spatial/temporal request handling is the key logic to implement for harvesting metadata based on spatial and temporal criteria.
- Implement metadata conversion logic to convert native metadata (e.g., ISO19115 profile, ATOM) to the supported metadata model (e.g., ISO19115 or Dublin Core) in the GMU catalog.
- The OGC CWS Transaction API is always open in the GMU catalog for data clients or third-party service provider to register Web services. Any newly created Web services will be incrementally registered in the catalog.

6.4.5. Metadata indexing/ranking

In the GMU catalog, spatial /temporal indexing was implemented and applied on each metadata. This improves the searching capabilities on spatial footprint and temporal extent of dataset. To improve the search accuracy, semantic indexing and contextual search are proposed for the future cataloging implementation. As an example, the semantic indexing model would need to capture the relationship (e.g., hierarchy) between keywords and concepts. Thus, if a keyword is used by the user, the server could offer additional hits that are registered under the same top level concept. The relationship/similarity among keywords can be leveraged to heuristically adjust the searching scope based on the searching keywords from clients. The [NASA GCMD keywords](https://wiki.earthdata.nasa.gov/display/CMR/GCMD+Keyword+Access) [https://wiki.earthdata.nasa.gov/display/CMR/GCMD+Keyword+Access] system was recommended as the reference for the future implementation.

The contextual searching benefits the searching accuracy by analyzing the searching context information such as searching history or searching preference from clients. The contextual searching will return more accurate searching results based on clients' search preferences.

6.5. Lessons Learned: Pilot Itself

6.5.1. A nice demo doesn't (automatically) make a good story

Component providers are technically skilled to integrate the data into a (client/server) component in an interoperable, standardized and - in case of a client component - visually appealing way. There is however a difference between a demo that looks good and a demo that tells a meaningful story. It is therefore important to reflect about the ultimate goal(s) of the scenario. In other words, what are the questions to be addressed and answered? In case of the Arctic SDP, one question is to showcase the value of rich data environments, i.e. to highlight why people should make their data available in a standardized and interoperable way: what are the benefits of this? What is the advantage of an SDI? Related to this, it can

be meaningful to show how such a (sustainable) infrastructure can be used to answer geospatial questions. In general, such questions clearly need input from domain experts, which were sought out in the pilot process by bringing in additional experts.

6.5.2. Think of the demonstration video early in the project

Think of the message that has to be conveyed in the video and the destination audience. In general, it is advisable to have a reasonably short (4 to 8 minutes) video that gives an overview of the complete project. This should not be too technical: it should be accessible by people that want to know more about the project but don't have all the technical and/or geospatial expertise to understand everything in detail. The summary video is ideally complemented with a set of additional videos that describe individual aspects of the big picture. Even these videos shall not be too technical. The integration of technical aspects and domain elements are very difficult to be explained in short videos.

In general, when planning for video material as output of technical exercises, there is a high risk that the abilities of technical experts to express the overall value of technical details in a simple language should are overrated.

6.5.3. Align the scenario with the project requirements and the available data sources

This sounds logical, but it can take a lot of discussions / iterations to get it right. From a component provider's perspective, an ideal scenario is one that defines a set of clear actions for readily available data sources. In practice, this will often not be the case: - a common research goal is to investigate new possibilities and capabilities when using new / existing open, interoperable standards; hence, a clear set of scenario actions will often be lacking - definitely in the early stages of the project. This is often only clarified at a later stage, when the exploration of data & capabilities lead to new insights applicable to the scenario - parts of the required / desired data sources will often be lacking; it should not be a problem if some data is lacking, as long as the general picture is clear (and preferably shows the benefit of having the data readily available)

Chapter 7. Conclusions

In general, the Arctic SDI community has an impressive amount of data and services at their disposal, but discovery and access issues prevent users from making efficient use of that data. In future, those elements need to be explored in more detail and matched with tools that simplify the integration and use of data while data rights and usage policies are matched and acknowledged. Improved communication and best practices are further key elements to better leverage available resources.

Current efforts focus on technology rather than communication and education. The long-term value to stakeholders, which include data owners and users of “create once and use many times” data, cannot be overstated. To achieve this new type of data communication, all users, i.e. data providers, services operators, or catalog providers need to implement the best practices mentioned in this report, need to provide appropriate descriptions for their product and categorize their information, and need to establish links between data, information and services. This new paradigm of data communication based on experiences, fit-for-purpose analysis, and links would allow for an increased usage of data that is available via standard services, a key principle for efficient data usage.

The Semantic Web is still in its infancy, and for the time being, humans remain at key positions for data discovery, exploration, and application. Their requirements lie to a large extent on the communication side (metadata, service quality and type, fit-for-purpose, etc.), paired with technology that streamlines access to data and services. A fact that can only be matched by stringent usage of internationally adopted standards and best practices.

In order to improve the general user experience both on the data provider and the consumer side, we recommend that future initiatives should focus on the following aspects:

- **Discovery of data:** The discovery of data is still an issue. Though there are already catalogs available with thousands of data sets being registered, finding the right data is still a major challenge. Experiences have shown that there are lots of data sets, but the discovery process becomes lost in web pages with data descriptions that link to other websites that eventually produce dead ends instead of data. To improve this process, three main aspects need to be tackled:
 - **Annotation, vocabularies, and linked data:** Annotation systems, both human- and machine-based, are required to identify data that has been used for specific purposes. If humans could mark data and share their experiences, other could gain from these experiences and thus improve their own results. Both human- and automated annotation could build on linked data principles, where publications link the underlying data sets, or users describing their work on (portal) web pages link the original data, styles, schemas, and other relevant aspects.
 - **Crawling based approaches:** The current catalog approach, though it facilitates orchestrated catalog hierarchies in principle, turns out to be usually used in

isolation. I.e. each portal features their own catalog (if at all). Users then need to interact with a high number of catalogs, often through Web forms because the API endpoints are not directly exposed. This slows down the discovery process enormously. If these catalogs would at least provide their data in a way that search engines could fully harvest the catalog content, discovery would improve. As this is not likely to happen soon, other approaches such as direct harvesting of data services shall be further investigated, ideally combined with automated data analysis mechanisms to get fine granular insights on the actual service offerings.

- **Service availability and reliability:** Data providers often seem to underestimate the usefulness of their services. Otherwise it can be hardly explained that so many service URLs change without proper forwarding mechanism put into action. One possible approach to improve this situation are proper backlink mechanisms that show data providers what the data has been used for (in publications, other website, research, leisure, exploitation planning, governmental planning etc.). Currently, data providers often need to study the access logs of their Web servers to get insights into the user statistics, which does not go far enough. In addition to the backlink mechanisms, service operators shall be enlightened on the importance of stable URLs and unique identifiers.
- **Access to data through standardized service interfaces:** It became clear many times that the integration time for data served at standardized interfaces using standardized data models is a fraction of the time required to integrate data served in proprietary formats or embedded in Websites and reports provided as *pdfs*. It is all of our responsibility to urge data owners to make their data available at standardized interfaces, ideally suchlike OGC WFS or WCS that support access to the underlying data (compared to e.g. WMS, which only provides raster maps).
- **Open Data, Usage Policies, and Citations:** Analog to the usage of standards, it is a community responsibility to increase the number of openly available data sets. This development shall be complemented with new mechanisms to deal with usage policies and citations. At the moment, there is often little understood value for data providers to make their data continuously available at open interfaces. Citation mechanisms and backlinks play an important role in this discussion, as they can be used as arguments for continued support for data on the Web.
- **Sustainability:** Sustainability is a key element for any successful Spatial Data Infrastructure. The aspect aggregates many if not all of the elements described above. We assume that a key element will be the implementation of a new communication model in combination with reliable links to resources, available at standardized interfaces that implement open access policies.

Appendix A: Complete list of datasets analyzed for the project

Name	Sea Ice Age: Measures Arctic Sea Ice Characterization Daily 25km EASE-Grid 2.0
Description	This data set, part of the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) program, provides a daily record of Arctic sea ice characteristics for the years 1979 through 2012 derived from passive microwave brightness temperatures. Parameters include the location of sea ice cover, sea ice age, day of melt onset, and status of melt onset. Data are gridded in the 25 km Equal-Area Scalable Earth Grid (EASE-Grid) 2.0 and provided as netCDF files.
URL	http://nsidc.org/data/docs/measures/nsidc-0532/
Owner	NSIDC
Reason if not used	

Name	Electronic Navigational Chart (ENC) basemap
Description	Vector files of chart features and available in S-57 format.
URL	http://www.charts.noaa.gov/InteractiveCatalog/nrnc.shtml#mapTabs-2
Owner	NOAA
Reason if not used	/

Name	NOAA Marine observations data
Description	KML with network links for live updates
URL	http://www.ndbc.noaa.gov/kml/marineobs_by_pgm.kml
Owner	NOAA
Reason if not used	/

Name	exactAIS Arctic Archive
Description	Satellite AIS data with ship positions and tracks for the Arctic region, in GeoJSON
URL	https://www.arcgis.com/home/item.html?id=3d40a9b3538d42bb8b7a1c289b675de1
Owner	ESRI - ExactEarth
Reason if not used	/

Name	Red List of Threatened Species - Marine Mammals spatial data
------	--

Description	The data is available as ESRI Shapefiles format and contains the known range of each species. Ranges are depicted as polygons, except for the freshwater HydroSHED tables. The Shapefiles contain taxonomic information, distribution status, IUCN Red List category, sources and other details about the maps.
URL	http://www.iucnredlist.org/technical-documents/spatial-data
Owner	International Union for Conservation of Nature (IUCN)
Reason if not used	/

Name	Fish occurrence data
Description	Georeferenced occurrence records about all life on Earth, in CSV format
URL	http://www.gbif.org/occurrence/search?HAS_COORDINATE=true&HAS_GEOSPATIAL_ISSUE=false
Owner	Global Biodiversity Information Facility (GBIF)
Reason if not used	/

Name	Ecologically or Biologically Significant Marine Areas (EBSAs)
Description	The EBSAs are special areas in the ocean that serve important purposes, in one way or another, to support the healthy functioning of oceans and the many services that it provides. It is available in GeoJSON.
URL	https://www.cbd.int/ebsa/
Owner	UN
Reason if not used	We already used LMEs to outline important ecosystems. Too many outlines would have made the demonstration unclear.

Name	FAO Statistical Areas for Fishery Purposes
Description	FAO Major Fishing Areas for Statistical Purposes are arbitrary areas, the boundaries of which were determined in consultation with international fishery agencies on various considerations, including (i) the boundary of natural regions and the natural divisions of oceans and seas; (ii) the boundaries of adjacent statistical fisheries bodies already established in intergovernmental conventions and treaties; (iii) existing national practices; (iv) national boundaries; (v) the longitude and latitude grid system; (vi) the distribution of the aquatic fauna; and (vii) the distribution of the resources and the environmental conditions within an area. It is available in GML and SHP.
URL	http://www.fao.org/geonetwork/srv/en/main.home?uuid=ac02a460-da52-11dc-9d70-0017f293bd28
Owner	FAO GeoNetwork
Reason if not used	We already used LMEs to outline important ecosystems. Too many outlines would have made the demonstration unclear.

Name	Microsoft Bing Maps aerial imagery
Description	Worldwide imagery provided by the Microsoft Bing Maps web service.

URL	https://www.bingmapsportal.com/
Owner	Microsoft

Name	Porcupine Caribou Herd
Description	GPS collar tracking data of the Porcupine Caribou Herd. The dataset was provided under a Terms of Use agreement between OGC and the Porcupine Caribou Technical Committee. Data was limited to those GPS-collar individuals in the Porcupine Caribou Herd with more than 15 months of continuous data. One relocation per day per animal between 1985 and January of 2016. Just to be clear that its not one location per animal per day from all animals tracked during this time. The dataset had originally been put together for the purpose of range analysis.
URL	https://carma.caff.is/images/_Organized/CARMA/About/Conferences/Carma8/Suitor_PCH%20monitoring_36by44P.PDF
Owner	Porcupine Caribou Technical Committee

Name	Arctic Circumpolar Distribution and Soil Carbon of Thermokarst Landscapes, 2015
Description	This data set provides the distribution of thermokarst landscapes in the boreal and tundra eco-regions within the northern circumpolar permafrost zones. This dataset provides an areal estimate of wetland, lake, and hillslope thermokarst landscapes as of 2015. Estimates of soil organic carbon (SOC) content associated with thermokarst and non-thermokarst landscapes were based on available circumpolar 0 to 3 meter SOC storage data. Olefeldt, D., S. Goswami, G. Grosse, D.J. Hayes, G. Hugelius, P. Kuhry, B. Sannel, E.A.G. Schuur, and M.R. Turetsky. 2016. Arctic Circumpolar Distribution and Soil Carbon of Thermokarst Landscapes, 2015. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1332
URL	http://dx.doi.org/10.3334/ORNLDAAC/1332
Owner	ORNL DAAC

Name	Long-Term Arctic Growing Season NDVI Trends from GIMMS 3g, 1982-2012
Description	Normalized Difference Vegetation Index (NDVI) data for the arctic growing season derived primarily with data from Advanced Very High Resolution Radiometer (AVHRR) sensors onboard several NOAA satellites over the years 1982 through 2012. Guay, K.C., P.S.A. Beck, and S.J. Goetz. 2015. Long-Term Arctic Growing Season NDVI Trends from GIMMS 3g, 1982-2012. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1275
URL	https://doi.org/10.3334/ORNLDAAC/1275
Owner	ORNL DAAC

Name	General Bathymetric Chart of the Oceans Grid 2014
Description	Worldwide bathymetry
URL	http://www.gebco.net/data_and_products/gridded_bathymetry_data/
Owner	GEBCO

Name	Viewfinder Panoramas SRTM DEM
Description	Worldwide 3-arc seconds elevation data derived from SRTM and other sources
URL	http://www.viewfinderpanoramas.org/dem3.html
Owner	Jonathan de Ferranti

Name	Arctic DEM
Description	High resolution elevation data covering the entire Arctic circle
URL	http://pgc.umn.edu/arcticdem
Owner	Polar Geospatial Center – University of Minnesota

Name	Anchorage DTM
Description	0.5m digital terrain model for the municipality of Anchorage
URL	http://maps.dggs.alaska.gov/elevationdata/#-16646306:8652684:9
Owner	Municipality of Anchorage / Alaska

Name	Natural Earth
Description	Public domain vector and raster map themes (1:10,000,000)
URL	http://www.naturalearthdata.com/downloads/
Owner	Natural Earth (public domain)

Name	OpenStreetMap
Description	World wide high resolution street maps and other vector data
URL	http://openstreetmap.org
Owner	OpenStreetMap contributors

Name	Daily sea ice concentration
Description	This data set is generated from brightness temperature data and is designed to provide a consistent time series of sea ice concentrations spanning the coverage of several passive microwave instruments. The data are provided in the polar stereographic projection at a grid cell size of 25 x 25 km.
URL	http://nsidc.org/data/nsidc-0051
Owner	National Snow and Ice Data Center

Name	Land Air Mean Temperature
Description	Monthly global gridded high resolution station (land) data for air temperature and precipitation from 1900-2014.
URL	https://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html
Owner	University of Delaware

Name	Sea Air Temperature
------	---------------------

Description	Sea Air Temperature
URL	https://www.esrl.noaa.gov/psd/repository/entry/show?entryid=synth%3Ae570c8f9-ec09-4e89-93b4-babd5651e7a9%3AL25jZXAucmVhbmFseXNpcy5kZXJpdmVkJ3N1cmZhY2UvYWlyLm1vbi5tZWFuLm5j
Owner	NOAA

Name	Sea Surface Temperature
Description	Sea Surface Temperature
URL	https://www.esrl.noaa.gov/psd/repository/entry/show?entryid=synth%3Ae570c8f9-ec09-4e89-93b4-babd5651e7a9%3AL2ljb2Fkcy8yZGVncmVIL2VuaC9zc3QubWVhbi5uYw%3D%3D
Owner	NOAA

Name	World Database on Protected Areas
Description	World Database on Protected Areas
URL	https://www.protectedplanet.net/c/world-database-on-protected-areas
Owner	UN Environment & IUCN

Name	Blue Marble Next Generation
Description	Worldwide satellite imagery mosaic
URL	http://mirrors.arsc.edu/nasa/
Owner	NASA Earth Observatory

Name	LANDSAT-8
Description	Satellite imagery
URL	http://earthexplorer.usgs.gov/
Owner	USGS

Name	Northern Canada Shapefiles
Description	Vector data in northern Canada
URL	http://geogratis.gc.ca/api/en/nrcan-rncan/ess-sst/7e388083-6b66-5e0e-a264-a3c0eb98a2f0.html
Owner	Natural Resources Canada

Name	ESRI ArcGIS – World Topographic Map (Basemap)
Description	Includes boundaries, cities, water features, physiographic features, parks, landmarks, transportation, and buildings
URL	https://services.arcgisonline.com/ArcGIS/rest/services/World_Topo_Map/MapServer (ArcGIS)
Owner	ESRI

Name	Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS) Integrated Catalogue
Description	Satellite Imagery from various satellite sources
URL	http://cwic.wgiss.ceos.org/discovery?REQUEST=GetCapabilities&SERVICE=CSW (CSW)
Owner	Committee on Earth Observation Satellites (CEOS)

Name	GeoMet WMS
Description	Provides access to Environment Canada's Meteorological Service of Canada (MSC) raw numerical weather prediction (NWP) model data layers (using Ocean Currents layer)
URL	http://geo.weather.gc.ca/geomet/?lang=E&service=WMS&request=GetCapabilities (WMS)
Owner	Environment Canada

Name	NGA Arctic Open Data through ArcGIS
Description	Arctic Summit - Shipping and Hydrography – shows the shipping lanes and routes in the Arctic
URL	https://ngamaps.geointapps.org/arcgis/rest/services/Arctic_Summit/Shipping_and_Hydrography/MapServer?f=json (ArcGIS)
Owner	NGA

Name	NGA Arctic Open Data through ArcGIS
Description	Arctic Summit – Airfields – shows the various permanent and temporary airports/airfields in the Arctic.
URL	https://ngamaps.geointapps.org/arcgis/rest/services/Arctic_Summit/Airfields/MapServer?f=json (ArcGIS)
Owner	NGA

Name	Conservation of Arctic Flora and Fauna (CAFF) Arctic Biodiversity Data Service (ABDS) WMS
Description	Protected Areas layer
URL	http://geo.abds.is:80/geoserver/ows?REQUEST=GetCapabilities&SERVICE=WMS (WMS)
Owner	CAFF

Name	Conservation of Arctic Flora and Fauna (CAFF) Ecologically and Biologically Significant Areas (EBSA) WMS
Description	Shipping incidents and accidents layer
URL	http://geo.abds.is/geoserver/ebsa/wms?REQUEST=GetCapabilities&SERVICE=WMS (WMS)
Owner	CAFF

Name	Canadian Ice Services – Ice Data
------	----------------------------------

Description	Shapefile data downloaded from University of Colorado at Boulder and merged into Compusult's current ice database using custom scripts.
URL	http://wms-icepolys.compusult.net/ServiceDBWMS/DBWMS/ICEBERGS?REQUEST=GetCapabilities&SERVICE=WMS (WMS)
Owner	Canadian Ice Services

Name	Federal Aviation Administration (FAA) Data
Description	Real-time flights data is pushed to Compusult through Aircraft Situation Display to Industry (ASDI)
URL	https://wms-faa.compusult.net/ServiceDBWMS/DBWMS/FAA?REQUEST=GetCapabilities&SERVICE=WMS (WMS)
Owner	FAA

Name	Automatic Identification System (AIS) ship data
Description	Ship locations provided through volunteers around the globe who collect and share the data via data stream.
URL	https://wms-ais.compusult.net/ServiceDBWMS/DBWMS/AIS?REQUEST=GetCapabilities&SERVICE=WMS (WMS)
Owner	Global (data served via WMS by Compusult)

Name	Aviation Routine Weather Reports (METAR) data
Description	Data for airports and permanent weather observation stations worldwide is downloaded by Compusult from NOAA's FTP site hourly.
URL	https://wms-metar.compusult.net/ServiceDBWMS/DBWMS/METAR?REQUEST=GetCapabilities&SERVICE=WMS (WMS)
Owner	NOAA

Appendix B: Revision History

Date	Release	Editor	Clauses	Descriptions
June 21, 2017	F. Houbie	0.1	all	initial version
June 23, 2017	I. Simonis	1.0	all	content revised, restructured, complemented
July 18, 2017	I. Simonis	1.1	all	all revised based on input from participants
July 18, 2017	I. Simonis	1.2	Annex A	data sets added