

# Theme 1 Overview

## Data Discovery and Metadata

William Manley  
University of Colorado INSTAAR

Allison Gaylord  
Nuna Technologies

Craig Tweedie  
University of Texas El Paso

Ted Habermann  
The HDR Group

John Kozimor  
The HDR Group

Roberta Score  
CH2M HILL Polar Resources

Polar Connections Interoperability Workshop  
November 8, 2016

# The Five Key Questions

What is interoperability in the context of different stakeholders (e.g. scientists, Arctic residents)?

How can interoperability benefit the polar and global community?

What initiatives, standards and tools can be used to enhance interoperability?

What are the gaps in interoperability?

Who are the actors needed to enhance interoperability (individuals, organizations, governments etc.)?



Arctic Research  
Mapping Application

[armap.org](http://armap.org)

Arctic Observing  
Viewer

[ArcticObservingViewer.org](http://ArcticObservingViewer.org)



## A Challenge for Arctic Data ...

... is knowing what's where.

- Do the data exist?
- How do I find it?
- What research is being done?
- How can we better plan, coordinate, and achieve scientific objectives?

## A Vision

### Project & Data Life Cycle:



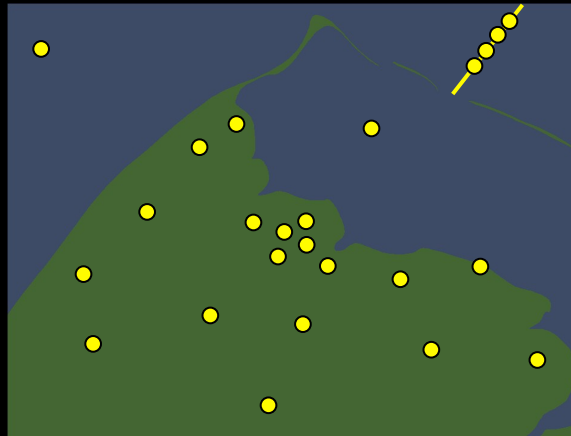
Each **project location** is a logistical base of operation.

# A Vision

## Project & Data Life Cycle:



Each **project location** is a logistical base of operation.



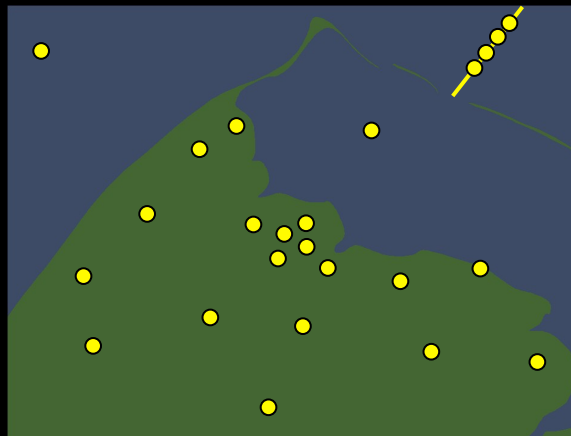
Each **data collection site** is a sensor, monitoring asset, observing platform, or wherever repeat measurements have been taken. E.g., flux towers, vegetation plots, climate stations, boreholes, stream gauges, buoys, ship tracks, etc.

# A Vision

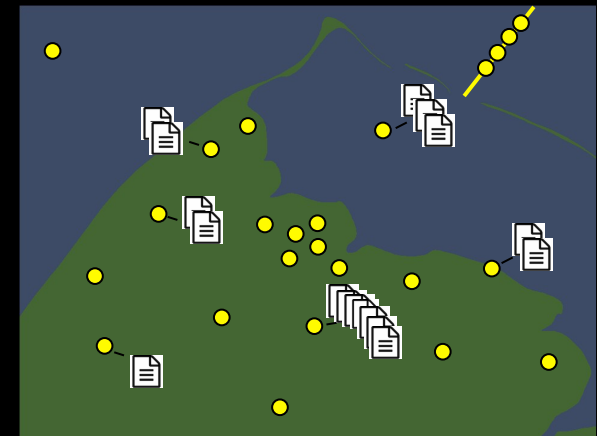
## Project & Data Life Cycle:



Each **project location** is a logistical base of operation.



Each **data collection site** is a sensor, monitoring asset, observing platform, or wherever repeat measurements have been taken. E.g., flux towers, vegetation plots, climate stations, boreholes, stream gauges, buoys, ship tracks, etc.



Each data collection site can have many **datasets**.

# A Vision

## Interoperable Applications:

Project Planning

Collection Site  
Monitoring

Dataset Usage &  
Understanding



...



...



- 1,300+ project locations
- High order
- Title, funding agency, funding program, discipline, point of contact, start and end dates, etc.
- All Arctic science
- 18 agencies & organizations
- 13,000+ data collection sites
- High spatial resolution
- Collection type, site name, elevation, science keywords, links to datasets, etc.
- Arctic Observing
- Multiple networks
- Tens of thousands of datasets
- Scientific data
- Discipline- and measurement-specific details
- All Arctic science
- Numerous projects, funding agencies, initiatives, etc.

# A Vision

## Meet User Needs:



...



...



Who is doing what, when and where?

How do we plan for logistics?

Where are medical facilities, field research stations, ship tracks, airports, etc.?

How do we best achieve the science?

Where are existing data collection sites?

Where are more sites needed?

Who operates and manages existing sites?

Which sites can I use?

Is this dataset suitable for my research?

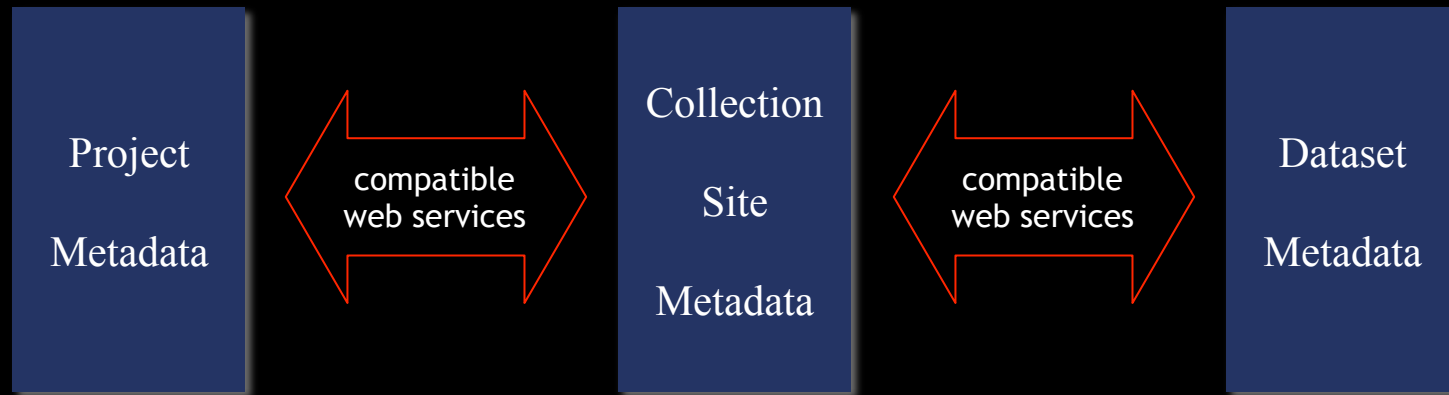
Does it cover my area for the right time period?

How was it created? What are the errors? Who do I contact with questions?

*Consider different audiences.*

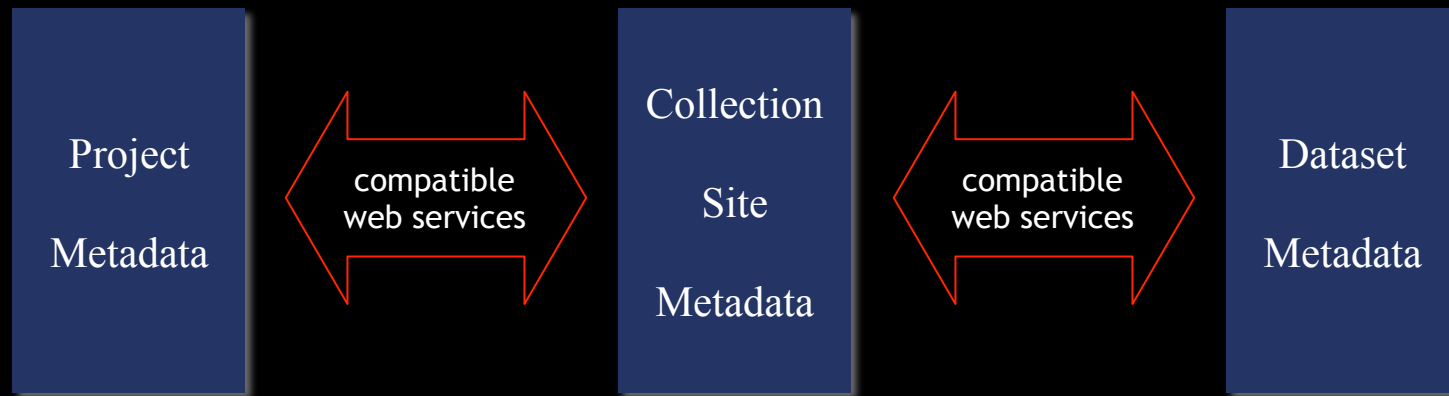
## A Vision

Connecting Information Systems through ISO Metadata and RESTful Services:



## A Vision

### Connecting Information Systems through ISO Metadata and RESTful Services:



- Metadata can be distributed across multiple organizations through web services.
- ARMAP and AOV have adopted ADIwg community standards, with some differences in implementation. Templates are available and a guide to interoperability is on the website.
- We support ISO 19115-3 and other standards.
- Overarching goals are consistency, appropriate level of completeness, and interoperability.

# AOV Guide to Interoperability

ArcticObservingViewer.org/interoperability


- Rationale
  - Template ISO XML's
  - Live Use Cases
  - Data Dictionary
  - Pick Lists
- 
- Web Services
    - ARMAP
    - AOV (*coming soon*)


## Implementation Examples

The metadata web services inherent to AOV and ARMAP are illustrated with ISO XML links in the table below. The template XML's are embedded with explanatory text, whereas the use case XML's are from live services for an NSF-funded AON project. Together they can assist with generating a workflow.

Project-Level Metadata:	-	use case
List of Collection Sites:	template	use case
Site-Level Metadata:	template	use case
Dataset-Level Metadata:	-	use case

[Home](#) [Partners](#) [About](#) [Interoperability](#) [Collaborate](#)

 **ARCTIC OBSERVING VIEWER**  
collaborative mapping of data collection sites



---

## Guide to Interoperability

Many Arctic science organizations realize that it's important to share information. Sure, a data center or monitoring network can increase its visibility by having a web page or data catalog of some sort, making it possible for end users to browse for information – metadata – that makes data more discoverable and accessible.

The problem is that there is a growing multitude of data catalogs, with the end result that the Arctic data landscape is fragmented, frustrating the end user that wants to easily find relevant data. In this context, organizations or initiatives can more successfully showcase their efforts by releasing metadata in such a way that it is broadly compatible for inclusion in various portals. In so doing, the information is highly visible for more users, and for greater impact.

Once an organization makes a decision to release metadata, the next hurdle is deciding on a path forward with implementation that maximizes compatibility with other information systems. Ideally, organizations will release metadata through web services – live data feeds between databases and applications – so that metadata is kept up to date and comprehensive. In this light, this brief guide is an attempt to facilitate the interoperability of metadata, and specifically for sets of metadata that span from projects to collection sites to datasets and back. This guide is intended for existing or potential Partners collaborating with AOV, and may be helpful as an example of successful implementation.

---

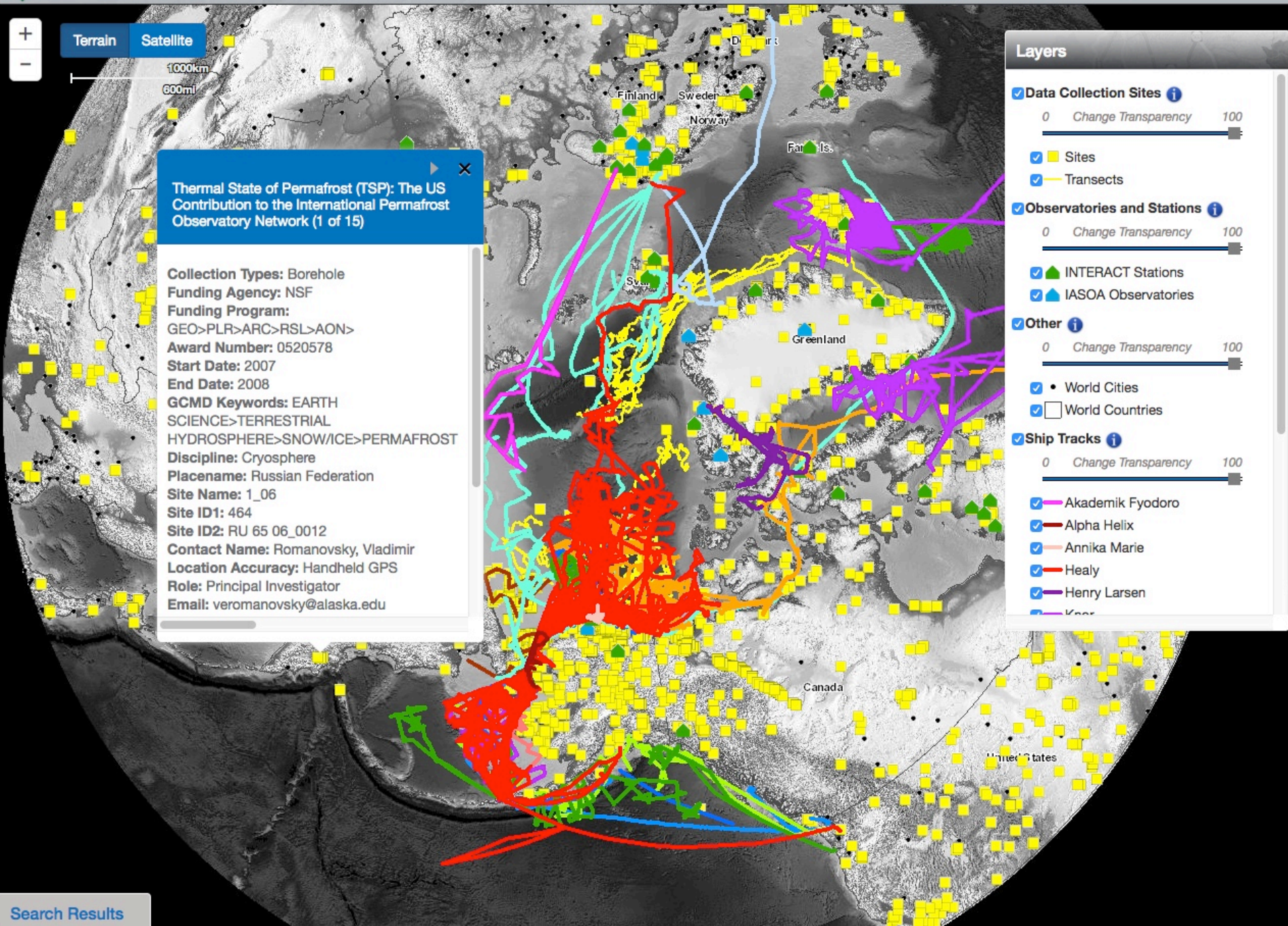
## Why Create Web Services?

The ultimate goal is that information for multiple observing networks is discoverable, authoritative, and up to date. Due credit should be given to data sources. And the information should be made accessible for use by various groups in a variety of ways for their own purposes.

In essence, what is needed is a dynamic network of distributed nodes for information sharing. This in turn relies on establishment of web services – live data feeds that conform to community-based metadata standards and compatible web service formats. Without interoperable web services, information becomes out-of-date, or requires repeated, substantial harmonizing and reprocessing. The Arctic data community is making progress on this front, notably through ADIwg, the IARPC ADCT, the IASC/SAON ADC, and other efforts or initiatives. The AOV Team is assisting with this planning.

## Why Consider the Project-Data Life Cycle?

Dynamically sharing information is becoming commonplace for dataset-level metadata, with compatible web services enabling quick federated searches across multiple data catalogs, for example. Also important is sharing project-level metadata – high-order information such as title, funding agency, project location, etc. for project tracking, program coordination, or for other purposes. And the piece in the middle that can tie it all together is



# What is interoperability?

The ability to easily share, find, assess, access, and integrate data and information.

# Why be interoperable?

Connect multiple portals

Improve discovery and access

Reproduce and re-use

Integrate for synthesis

Avoid re-inventing the wheel

Avoid laborious harvesting and reprocessing

Increase visibility and impact for your research

Obtain a comprehensive perspective for better science planning

Better achieve scientific objectives

*Consider different audiences.*

# Existing Resources

ADIWG mdTools

AOOS Research Workspace

GeoPortal

ArcCatalog metadata editor

GICat and other brokering technologies

tools through INSPIRE

templates etc. through AOV

## What are the gaps?

- **Fragmentation of the data landscape with multiple, isolated data catalogs and portals**
- **Dissimilar metadata vocabularies across systems**
- Disconnects between disciplinary or regional systems and global systems
- Paucity of easy to use tools for authoring ISO metadata
- Incomplete “buy in” from data centers, researchers, funders, policy makers, etc. to share information (compatibly)
- Metadata not fully integrated from beginning to end of a project
- **Proliferation of customized metadata implementations**
- Insufficient definitions of fields within a standard
- **Variable implementations within a given standard**
- A guide to implementing ISO without hiring an ISO expert
- Funding for policy-level and technical level, coordinated improvements to interoperability

## Good News



Value of Data & Data Sharing

# Who is involved?

Data repositories, data catalogs, observing networks, science planning organizations, workshops, data management plans, and data policies

- too many to list
- see the Arctic Data Ecosystem Map

Funding agencies

Research programs

PI's

Students

Coordinating initiatives:

US Arctic – ADIwg, IARPC ADCT

European Arctic –EU-PolarNet, INSPIRE

Circumarctic – IASC/SAON ADC, SAON CON, ASDI

Antarctic – SCAR, SOOS

US – ESIP, Unidata, DataOne, a2dc

Global – GEO, RDA, CODATA, GEO, GOOS, WMO, OGC, ISO, WDS, W3C

# Conceptual Models

Datasets

Projects / Datasets

Projects / Sites / Datasets

Programs / Networks / Projects / Platforms

Projects / Products

Projects / Assets

Projects / Data & Maps

# Hierarchical Approaches to Metadata

## Nested

project-level

site-level

dataset-level

vs

## Separate

project-level

site-level

dataset-level

*one big xml per project*

*multiple x-linked xml's*

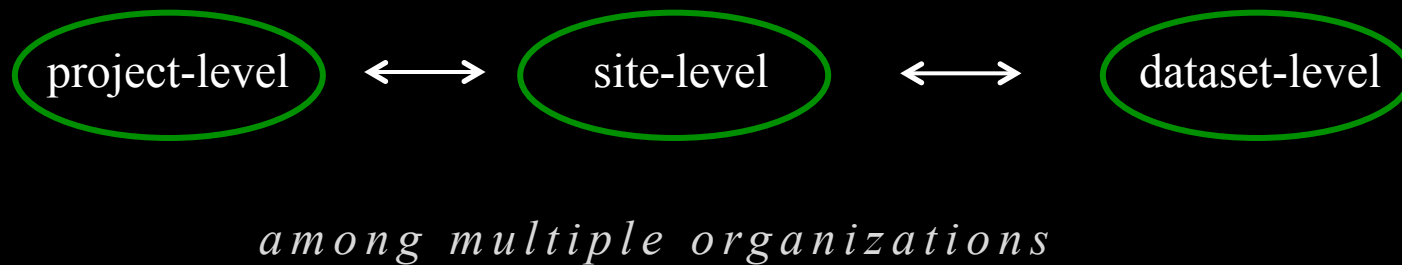
# Hierarchical Approaches to Metadata

## End to End



vs

## Distributed



# ISO or What?

DIF

EML

ISO 19115-1

SensorXML

ISO 19115-2

ISO 19110

ISO 19115-3

FGDC

Dublin Core

WMO WIGOS

Not to mention various file formats, programming languages, web service protocols, etc.

*But various tools and crosswalks can help.*

# Compatible Fields?

A fundamental challenge: mismatch of content

Project Title	↔	Project Title
Study Site Name	↔	Site Name
Project Status	<del>—</del>	
	<del>—</del>	Discipline
Platform Type	↔	GCMD Platform Keyword
Program Name	..... ? .....	Funding Agency
Network Title	..... ? .....	Initiative

# Improving Interoperability

- Share schemas, templates, data dictionaries, code lists (vocabularies), use cases, crosswalks, ...
  - Groups starting out: proceed with eyes wide open to avoid later effort
  - ADC Data Discovery and Metadata Working Group: compile or link to various approaches and especially element definitions, with recommendations
  - Establish a working group to agree on community implementations for flexible standards such as ISO
  - Establish compatible web services (including interoperable, distributed, and hierarchical approaches with x-links ...)
  - Follow the lead on successful implementations of brokering technologies, federated searches, and service crawlers (e.g. Arctic Data Explorer, GEOSS Common Infrastructure, Arctic SPD, )
  - Communicate and coordinate: ADC, IARPC, ADIwg, GEO, etc.
  - Avoid silos
- The (data) revolution is happening.*

*Thank you!*

Arctic Research  
Mapping Application

[armap.org](http://armap.org)

Arctic Observing  
Viewer

[ArcticObservingViewer.org](http://ArcticObservingViewer.org)