

Cyberinfrastructure and Collaboratory Support for the Integration of Arctic Atmospheric Research

BY SANDRA STARKWEATHER AND TANEIL UTTAL

The urgent research questions posed by global environmental change test more than the minds of scientists, but also the vitality of cyberinfrastructure and collaboration support required to advance research integration. Research integration refers to efforts that involve researchers from different scientific fields (e.g., atmospheric science, terrestrial ecology, social science) and research approaches (e.g., in situ observations, modeling, physical process studies) combining their work in value-added ways (Averyt 2010). Advancing critical research integration across institutions and projects demands innovative approaches that effectively leverage partner resources. Cyberinfrastructure encompasses technological solutions to enabling research integration, such as sharing data and computing resources across institutions, while collaboratories, which can take on a variety of forms (Bos et al. 2007), are an essential and often overlooked infrastructure for interpersonal collaboration. A collaboratory for research integration serves to network independently funded experts and provide them with the resources they need to work toward collaborative objectives, without regard to their physical location. For the International Arctic Systems for Observing the Atmosphere (IASOA, see the article by

Uttal et al. in this issue), an international consortium of 10 independently funded atmospheric observatories encircling the Arctic, the vision to integrate long-term, distributed atmospheric observations into the complex picture of Arctic change was hindered by both incompatible data management infrastructures and a lack of collaboration infrastructure. This article presents the IASOA response to these obstacles: a new data portal for unified discovery of and access to long-term, in situ Arctic observations (www.esrl.noaa.gov/psd/iasoa/dataatagance), and the facilitation of open, international scientific collaboratories (www.esrl.noaa.gov/psd/iasoa/science) that tackle pan-Arctic atmospheric research challenges.

DATA-AT-A-GLANCE. A first-order challenge of integrating environmental observations collected by diverse scientists and institutions is identifying and cataloging all of the data—a process known in data science as “data discovery.” Data portals are often recommended as the cyberinfrastructure of choice for data discovery as they can virtually unify the holdings of incompatible data management systems into a single search interface through the use of standardized metadata (i.e., documentation of data). Well-described metadata goes beyond addressing discovery questions like “What is out there?” and “Where do I get it?”. It also includes documentation on file formats, processing procedures, and uncertainty estimates. A well-designed data portal can greatly reduce the time and headaches associated with tracking down usable data; it reveals critical gaps in observing networks and opportunities for addressing new questions with previously unknown data resources. The purpose of creating an IASOA data portal was to provide such an interface for all of the long-term (greater than three years), in situ Arctic atmosphere and surface observations collected across the ten IASOA observatories.

AFFILIATIONS: STARKWEATHER—CIRES/Earth System Research Laboratory, Boulder, Colorado; UTTAL—NOAA/Earth System Research Laboratory, Boulder, Colorado
CORRESPONDING AUTHOR: Sandra Starkweather, CIRES/Earth System Research Laboratory, 325 Broadway, Boulder, CO 80305

E-mail: sandy.starkweather@noaa.gov

DOI:10.1175/BAMS-D-14-00144.1

©2016 American Meteorological Society

Category	Alert	Barrow	Cherskii	Eureka	Ny-Alesund	Oliktok	Pallas-Sodankyla	Summit	Tiksi	Villum
Aerosol	●	●		●	●		●	●	●	
Physical, Optical	●	●		●	●		●	●	●	
Physical, Primary	●	●					●		●	
Inorganic, Trace elements							●			
Atmospheric State	●	●	●	●	●		●	●	●	●
Cloud Properties	●	●	●	●	●		●	●	●	●
Macrophysical	●	●		●	●		●	●	●	
Microphysical		●		●				●		
Optical and Radiative Properties		●		●				●	●	
Cryosphere	●	●	●	●				●	●	
Greenhouse Gas	●	●	●	●	●		●	●	●	
Ozone	●	●		●	●		●	●	●	
POPs					●		●			
Precipitation Chem							●			
Radiometric	●	●	●	●	●		●	●	●	
Reactive Gas	●	●		●	●		●	●	●	
Surface Properties	●	●	●	●					●	

Fig. 1. Data-at-a-Glance is the landing page for IASOA’s data access portal. It provides a succinct visual summary of all of the data holdings across the IASOA observatories (data not yet available from Oliktok Point). Measurement categories (green triangles) can be expanded into levels of increasing detail. In this case, “Aerosol” and “Cloud Properties” have been expanded. Clicking on a blue dot directs the user to a set of search results that link to original datasets.

The IASOA data portal, “Data-at-a-Glance” (Fig. 1), is organized into a succinct matrix (measurements¹ by observatories) that provides a snapshot of both all the measurements made at a given location and the pan-Arctic coverage for a related set of measurements. Selecting a measurement category at a site of interest—for example, aerosol-optical properties from Barrow, Alaska—will bring up a list of relevant search cards for unique datasets—in this case, 13 results (3 shown in Fig. 2). Each card emphasizes expedited access (ideally “two clicks”) to actual data files hosted at a variety of archives and summarizes key information such as principal investigator and the length of the data record.

In spite of its value, a high level concern of creating Data-at-a-Glance was how such a small, research-driven consortium like IASOA would create and maintain all of the required metadata—hundreds, if not thousands, of unique descriptions. The good news was that it didn’t have to. IASOA was able to use

a highly leveraged approach to design and populate its data access portal based on automatically assembling much of the relevant metadata from existing online collections, a process known in data science as metadata “harvesting.” Most IASOA observatories are already active partners in global networks with robust data management capabilities such as Global Atmospheric Watch (GAW) and the Baseline Surface Radiation Network (BSRN). These networks support data archives with tools to author and search their metadata collections. Community data management is emerging for individual Arctic researchers through efforts like the Advanced Cooperative Arctic Data and Information Service (ACADIS), where researchers have access to tools to author metadata that is then included in the ACADIS catalog. Many of these repositories also support web-accessible means for automatically downloading their metadata. Through identifying and encouraging commonalities in the structured metadata already in use by these repositories—based on the ISO 19115 standard—IASOA automated the harvest of metadata for more than 800 datasets within the first 6 months of its launch.

To maintain a complete and up-to-date catalog of relevant datasets (likely more than 1,200), IASOA will

¹ The IASOA measurement vocabulary was aggregated from existing atmospheric science vocabularies (e.g., World Meteorological Organization) to support diverse queries in terms that are well understood by atmospheric scientists.

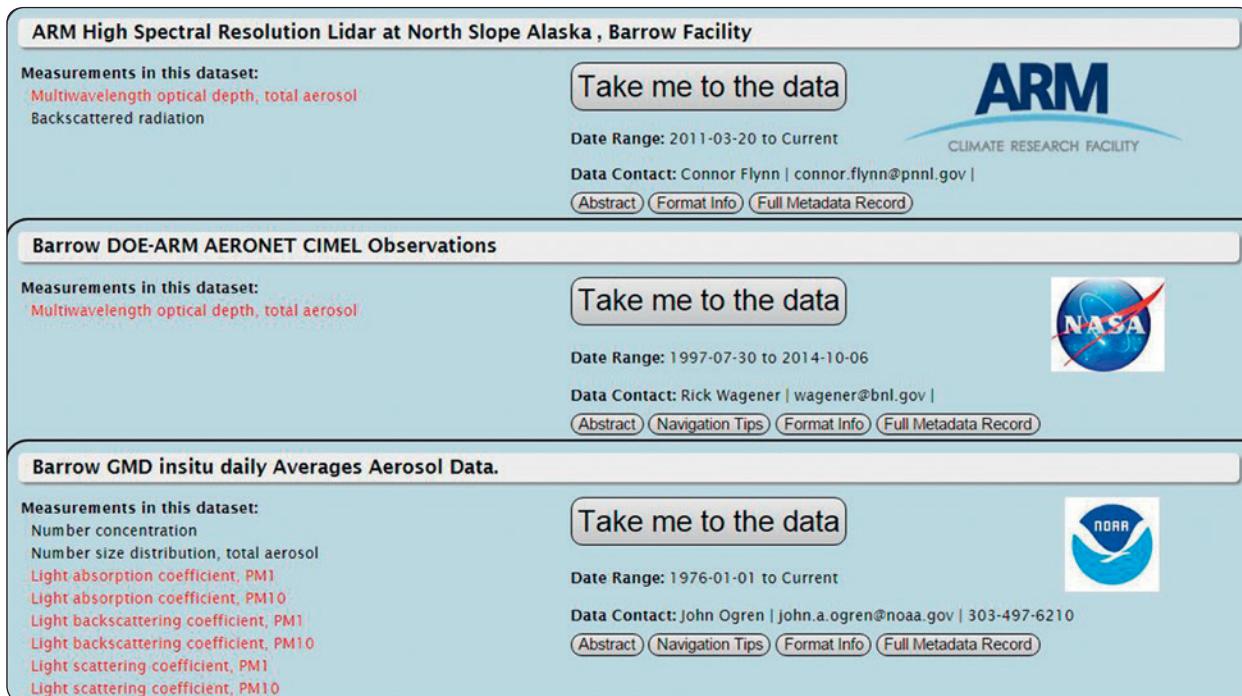


FIG. 2. Example search cards from the IASOA data portal emphasize expedited access to datasets (“Take me to the data”) while also providing access to full dataset documentation (“Full Metadata Record”). This search example also illustrates how IASOA’s data portal unifies the holdings of multiple data centers—in this case DOE-ARM, NASA, and NOAA.

continue to expand its automated harvesting across 10 global network repositories and as many large institutional and project-level repositories, though some obstacles persist. The disparity in datasets available for harvest across IASOA observatories and measurement categories is one indicator of the pressing need for investments in cyberinfrastructure within contributing institutions (Table 1). Many sponsor agencies that contribute to IASOA lag the global networks in their data management practices. While they provide long-term data storage, they often have yet to create structured metadata or to make it available for harvesters. IASOA has begun working with some sponsor agencies to create metadata, but resource constraints impeded progress. A further roadblock to harvesting is the incompatible metadata formats encountered across repositories. For example, many U.S. government agencies adopted the Federal Geographic Data Committee (FGDC) standard for metadata in the late 1990s, while international organizations like GAW have adopted ISO 19115. It is therefore valuable when the repositories can provide metadata in a range of common formats. IASOA advocates within partner organizations for greater

standardization and access. Keyword vocabularies present another source of incompatibility across IASOA partners. NASA’s Global Change Master Directory (GCMD) metadata standard vocabulary has 14 keywords to describe aerosol properties, whereas the GAW vocabulary has more than 90. The Climate and Forecast (CF) convention standard names are growing in popularity for their high specificity, going beyond mere keywords to describe nonambiguous physical parameters with canonical units. There are currently 138 distinct CF standard names for aerosol properties. IASOA found that a specific benefit of the ISO 19115 standard for metadata was its flexibility in combining multiple keyword vocabularies; this is one reason the consortium will continue to promote its use.

It is encouraging that even a small portal like Data-at-a-Glance has enjoyed responsive support from the large repositories, many of which crave closer connections with user communities and expanded opportunities to share data collections. The U.S. Department of Energy-Atmospheric Radiation Measurement Program (DOE-ARM) program was willing to add the ISO standard to its metadata of-

ferings (previously only offered in FGDC) in order to be included in the portal. The GAW program was willing to add fields to its ISO profile and expand its keyword vocabulary. ACADIS has both added metadata fields to its ISO profile and is considering ways to diversify its keyword vocabularies, following the IASOA model. Collaboration with these repositories is an extremely valuable, mutually beneficial process. The large repositories benefit from learning more use scenarios; they become more flexible and expand the reach of their catalogs. IASOA benefits from letting large data centers create, manage, and maintain metadata.

SUSTAINING FACILITATED NETWORKS OF EXPERTS. IASOA recognized, however, that scientists need more than cyberinfrastructure to integrate their research with others; they also need infrastructure that facilitates interpersonal collaboration. This is particularly imperative when collaborators are as physically distributed (often across 12 time zones) as IASOA’s pan-Arctic network of experts. Following the official closing of the last International Polar Year in 2012, the IASOA steering committee reviewed a

host of freshly published national and international science plans to identify six topics where urgent progress was needed to combine in situ observations and improve understanding about the Arctic atmosphere. IASOA is providing critical support in these areas through its open topical working groups (www.esrl.noaa.gov/psd/iasoa/science), which operate similarly to the “Distributed Research Center” collaborative model described in Bos et al. (2007). This type of collaborative tackles the most challenging form of research integration: aggregating independently funded research efforts at a distance with the goal of co-creating new knowledge. Because of this challenge, all IASOA working groups are supported by an Implementation Scientist—a novel position created to directly support collaborative research development and facilitate working group progress.

Today, three of six topics are progressing through vibrant collaborative working groups, which tackle a range of community challenges during their regular teleconferences. For example, experts share and compare Arctic-specific best practices for monitoring to address issues like instrument frosting and low signal-to-noise ratios. Further, they deliberate unified

TABLE 1. Count of datasets currently available through IASOA data access portal by category and observatory. More than 1,000 datasets have been collected across the 10-observatory consortium (data not yet available from Oliktok Point) in measurement categories of interest to Arctic atmospheric and system science. The disparity of harvested datasets across observatories and measurement categories is one indicator of the pressing need for investments in data management and cyberinfrastructure within the member institutions and global networks.

Measurement Category	Alert	Barrow	Cherskii	Eureka	Ny-Alesund	Pallas-Sodankyla	Station Nord	Summit	Tiksi
Aerosol	9	22	0	3	7	13	0	7	9
Atmospheric State	6	72	2	3	5	6	3	11	9
Cloud Properties	1	117	0	7	2	1	0	16	3
Cryosphere	0	51	2	1	0	0	0	1	2
Greenhouse Gas	35	91	1	6	17	15	0	54	14
Ozone	8	10	0	10	13	12	0	10	6
POPs	0	0	0	0	1	7	0	0	0
Precipitation Chemistry	0	0	0	0	29	2	0	0	0
Radiometric	5	50	2	7	2	3	0	23	10
Reactive Gas	25	24	0	14	38	51	0	19	1
Surface Properties	0	99	5	4	0	0	0	0	8



correction schemes for data acquired across diverse instrument configurations. They also leverage resources, such as sharing error estimates developed at highly instrumented facilities with smaller facilities. The emerging outputs of the groups include the development and documentation of interoperable data products and peer-reviewed publications that analyze these observations in the context of the Arctic atmospheric system. The Implementation Scientist helps these experts identify areas where progress can be made, seeks out new expertise and resources to address identified obstacles, and entrains new perspectives through information stakeholders. Because the Implementation Scientist is also responsible for the design and development of Data-at-a-Glance, critical connectivity exists between data products developed by the working groups, their metadata development, and accessibility for the broader research community.

IASOA working groups are open to all collaborators with a stake in putting Arctic observations to use. The Implementation Scientist accelerates progress through the introduction of new expertise in modeling, coupled process studies, data management, or decision-relevant services like sea ice forecasting. For example, the Aerosol Working Group has grown from 6 observational scientists from 3 countries in 2013 to 19 collaborators from 8 countries in 2015, including modelers involved in intercomparison projects and data managers from the World Data Center for Aerosols (WDCA). During this time, their efforts grew from comparing data from three observatories (see 2013 Arctic Report Card: www.arctic.noaa.gov/reportcard/) to developing and documenting a series of standardized data products for aerosol optical properties from seven observatories. These additional experts provide critical guidance on factors that increase the usability and application of this valuable observational data. The participation of data managers from the WDCA has streamlined the inclusion of these new data products to the archive, while the introduction of atmospheric chemistry modelers has prompted stimulating discussions on the types of model outputs that will be most readily compared to these in situ observations. Similarly, the Radiation Working Group has grown from a few investigators with roughly comparable datasets into an energetic collaboration of more than 20 investigators who are comparing and explaining long-term, pan-Arctic trends in albedo and explaining regional differ-

ences in cloud radiative forcing. An additional group formed in 2014 to address surface-atmosphere coupling with an objective to improve understanding of Arctic-wide energy, moisture, and carbon fluxes. This group is highly interdisciplinary and is progressively defining common objectives between atmospheric, terrestrial, and cryospheric scientists. Developing interoperable flux data products across the diversity of observing infrastructure (e.g., tower heights that vary from 3 to 50 m) at IASOA sites will be a particular challenge for this group to address, as will be identifying best practices for consistent site characterization when standardized approaches emphasize midlatitude ecosystems. Specific outputs of IASOA working groups can be explored in detail in the article by Uttal et al. in this issue.

Importantly, all of the IASOA working groups are investigator-driven, so each takes on a unique flavor and focus as the groups self-assign objectives. Early career researchers have provided a valuable means for conducting much of the work to meet these objectives; an important dividend of this has been the substantial mentoring they receive through the group meetings from experts around the world. The Implementation Scientist, who participates in all groups, can identify opportunities for cross-group synergies and ways to improve data accessibility through “Data-at-a-Glance.” Without such facilitated integration, collective decision making, and co-creation of new knowledge, undocumented and individually processed observations would not realize their full value and would potentially be unavailable to the broader research community. Arctic observations are too valuable to risk that fate.

SUMMARY. Developing a comprehensive understanding of the Arctic atmospheric system is broader than any individual investigation, discipline, or even national effort. Research infrastructure both within and across disciplines and institutions is required to address challenges at this scale. IASOA’s easy-to-use cyberinfrastructure plays a critical role in unifying the globally distributed collection of pertinent observations for discovery and access. The design of “Data-at-a-Glance,” automatically populated with harvested metadata, is readily maintained and directly supports the IASOA vision for pan-Arctic research synthesis. Collaboration infrastructure is required as well, but often overlooked. Direct support for research facilitation and collaborative development through the role of the Implementation

Scientist has expedited implementation of IASOA's mission and contributed to robust and productive collaboration. It must be emphasized that all forms of infrastructure require investment. The challenge that lies ahead for IASOA is developing sufficient multi-institution support to sustain and grow this international effort.

ACKNOWLEDGMENTS. We thank T. Habermann, M. Parsons, and A. Milan for their help with guiding these concepts for IASOA's cyberinfrastructure.

FOR FURTHER READING

- Averyt, K., 2010: Are we successfully adapting science to climate change? *Bull. Amer. Meteor. Soc.*, **91**, 723–726, doi:10.1175/2010BAMS2906.1.
- Bos, N., A. Zimmerman, J. Olson, J. Yew, J. Yerkie, E. Dahl, and G. Olson, 2007: From shared databases to communities of practice: A taxonomy of laboratories. *J. Comput. Mediat. Commun.*, **12**, 652–672, doi:10.1111/j.1083-6101.2007.00343.x.
- <http://iasoa.org>

NEW FROM AMS BOOKS!

A Scientific Peak: How Boulder Became a World Center for Space and Atmospheric Science

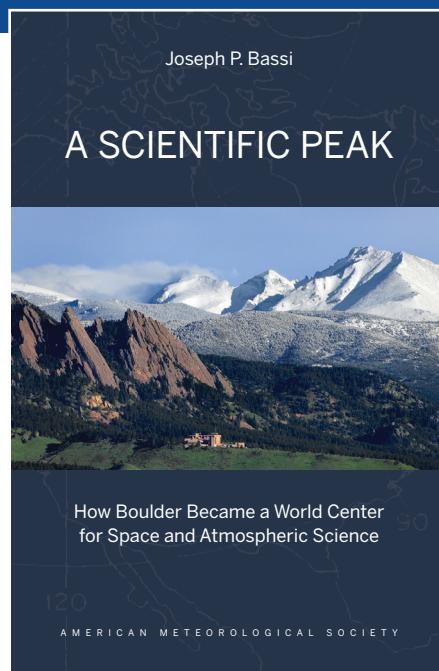
Joseph P. Bassi

Once a Wild West city tucked between the Rocky Mountains and the Great Plains, Boulder is now home to some of the biggest names in science, including NCAR, NOAA, and NIST.

Why did big science come to Boulder? How did Boulder become the research mecca it is today?

A Scientific Peak is a fascinating history that introduces us to a wide variety of characters, such as Walter Orr Roberts, and the serendipitous brew of politics, passion, and sheer luck that, during the post-WWII and Cold War eras, transformed this “scientific Siberia” into one of America’s smartest cities.

© 2015, 264 pages, paperback
print ISBN: 978-1-935704-85-0 eISBN: 978-1-940033-89-1
List price: \$35 AMS Member price: \$25



AMS BOOKS

RESEARCH APPLICATIONS HISTORY

> bookstore.ametsoc.org