

**Digital Presence Online Analysis:**  
**USGS Alaska Science Center Data Repository**

Rachel Casey

College of Computing and Informatics, Drexel University

INFO 591: Data and Digital Stewardship

Dr. Alex Poole

November 1, 2020

## **Introduction**

The USGS Alaska Science Center “provides objective and timely data, information, and research findings about the earth and its flora and fauna to Federal, State, and local resource managers and the public to support sound decisions regarding natural resources, natural hazards, and ecosystems in Alaska and circumpolar regions.” (Alaska Science Center). The Data Repository at the USGS Alaska Science Center includes research data that is publicly accessible at this website: <https://www.usgs.gov/centers/asc/data-tools>.

## **Archive vs. Repository**

On the USGS website there is a section titled “Archive vs. Repository: Is There a Difference?”, which explains that these definitions mean different things within the Federal government compared to other institutions. “In the field of data management, the terms "archive" and "repository" often are used interchangeably. Within the Federal government, however, the term "archive" is specific to the mission and activities of the National Archives and Records Administration (NARA)” (USGS Data Management). The ASC data repository is specifically labeled a repository and included on the USGS list of Trusted Digital Repositories, so USGS does not consider their website holdings to be archival. USGS defines a data repository as “a centralized place to store and maintain data. A repository can consist of one or more databases or files which can be distributed over a network. Data repositories are often managed by data curation personnel who ensure that files are managed and preserved for the long-term” (USGS Data Management). A lot of the verbiage overlaps with their definition of archiving, particularly the focus on long-term storage.

## **Data Repository Ownership and Management**

The ASC Data Repository is owned and funded by the United States Geological Survey (USGS), which is a scientific agency of the United States federal government. All datasets from the ASC is also available in the USGS Science Data Catalog, which collects and backs up metadata from select USGS centers through a weekly harvest process. Data release backups for ASC data sets are also located on in the USGS ScienceBase catalog. “This community (ScienceBase) contains copies of released data sets from the ASC. The authoritative copy of the data resides on the ASC repository. This community serves as an offsite backup and alternate download site for ASC released data” (USGS ScienceBase). Having the data sets backed up in multiple locations is a key aspect of the ASC data policy, which states that “proactive protection from loss or corruption are provided through the use of access permission, back-ups, version control, and off-site storage” (ASC Data Policy). These backups, which are hosted on USGS sites, ensure that should there be a catastrophic event in which the ASC data repository is destroyed or inaccessible, the data will not be lost since there are alternative copies maintained by trustworthy catalogs.

## **USGS Alaska Science Center Material Collection**

The Alaska Science Center policy webpage notes that “the Data Policy for ASC staff and contractors requires that research data, computational models, software and scripts, web-based tools, and information products (collectively “data”) adhere to the [following] policy statements” (ASC Data Policy). This suggests that the data in the repository was collected directly from ASC staff (which includes various researchers and scientists), contractors, and collaborators. All of the data is stored in house using ASC facilities and protocols. “Data will have a designated

Data Manager who is accountable and responsible for its long-term safe-keeping, description, management, maintenance, and availability. Established ASC facilities are to be used whenever possible for data archive, storage, and distribution” (ASC Data Policy). There are 134 staff members at the ASC including 2 IT Specialists who specifically focus on data management, Dennis Walworth and Marla Hood.

### **USGS Alaska Science Center Data Policy**

At the top of the repository page there is a link to the ASC Data Policy. This page specifies that ASC staff and contractors must comply with the policy. The introduction to the ASC data policy notes that the cost of the time, labor and monetary cost of data management is justified by the value of the data to the public (ASC Data Policy). The first part of the data policy emphasizes compliance with Federal law, DOI, USGS and ASC requirements, standards, and policies. The policy notes that all data will have a designated Data Steward and Data Manager. The Data Steward is “knowledgeable in the specific topic who is accountable and responsible for the specifications and quality of the dataset” while the Data Manager is “accountable and responsible for its long-term safe-keeping, description, management, maintenance, and availability” (ASC Data Policy). By requiring both a designated Data Steward and a Data Manager, the repository has greater oversight when it comes to overall data quality and maintenance.

The next section of the policy covers the location of the data maintenance noting that ASC facilities should be used for archiving, storing, and distributing data. Research project data are to be maintained in an ASC managed workspace where “proactive protection from loss or corruption are provided through the use of access permission, back-ups, version control, and off-

site storage, files and other documents are organized and accessible in digital or paper folders in an understandable fashion that would allow a knowledgeable peer, to continue the work with minimal startup effort in the absence of the employee, and replicate the work given the same starting material” (ASC Data Policy). Documentation is covered in the following portion of the policy including Research Data Management Plans (RDMPs), data models and diagrams that describe workflows, and complete provenance information for the data. Poole notes that “key considerations in crafting a DMP include the data to be generated and shared, funder, publisher, and internal requirements, disciplinary and domain standards and best practices, relevant policies, documentation, deposit location and provisions for backup, legal and ethical obligations, goals for sharing, roles and responsibilities of stakeholders, and resources needed” (Poole, 2016). These considerations are key to providing an overview, framework, and contextualization that is specific to each individual research project.

The data is finalized through documentation “using formal standardized metadata formats such as ISO 19115 or FGDC (Federal Geographic Data Committee), then it is provided with a valid digital object identifier (doi), and undergoes a review process to receive approval before archiving or release to the public” (ASC Data Policy). The data is then first shared with ASC staff before it is shared with the public within 1 year of the completion of the project. This timeline ensures transparency and timely publication of data sets. The data that is made is public is required to “use machine-readable and open formats, be published to the finest possible level of granularity practicable and permitted by law, and have open licenses that place no restrictions on copying, publishing, distributing, transmitting, adapting, or otherwise using the information for non-commercial or commercial purposes” (ASC Data Policy). By requiring machine-readability, open formats, and open licenses without restrictions ASC is ensuring that the data

can be reached by the widest possible audience. All of the statements listed in the ASC data policy work in facilitate to establish effective data management and adherence to USGS standards.

### **Assessment of Repository and Collection Scope**

The Data Repository at the USGS Alaska Science Center contains digital data from ASC researchers. As of October 30, 2020, there are 283 items in the repository. Of the 283 items, the majority (273) are science data sets. The other items are web applications (2), data management tools (1), GIS Data (5) and other (2). The topics include Biology and Ecosystems, Data, Tools & Technology, Geology, Land Resources, Mapping, Remote Sensing and Geospatial Data, Minerals, and Water. The repository items span from 2013 to the present day (2020). The ASC repository's main page highlights a sampling of research data collections that reflect the scope of the repository including Alaska Earth Mapping Resource Initiative, Wildlife Tracking Data Collection with Animated Maps, Heat Stress on Yukon River Chinook Salmon, Glacier and Climate Project Data, and Changing Arctic Ecosystems Data. The repository itself is easy to browse, well organized and frequently updated.

### **Primary Audience and Community Needs**

As stated in their mission, the primary audience of the Alaska Science Center data repository is Federal, State, and local resource managers. The ASC website has a page dedicated to Community Outreach and Engagement, which states that the "USGS participates in meetings and co-management councils to inform Alaska Native and rural communities of our research plans and findings, especially for studies that take place near communities or for research with findings that may be of high interest to communities." (ASC Outreach & Engagement). This

page also notes that the ASC follows the guidelines included in “The Principles for Conducting Research in the Arctic”, which was prepared by the Interagency Arctic Research Policy Committee (IARPC) and chaired by the National Science Foundation (NSF). These principles are “guidelines for the conduct of research, to better align with U.S. Arctic policy, to incorporate the latest advances in research methods, and to reflect expanded research efforts and disciplinary breadth in a rapidly changing Arctic” and were created with input from Alaska Native people and organizations (IARPC). It is important for ASC to address the needs of Alaska native people and organizations and to collaborate with them since they are indigenous to the land. ASC research output is directly relevant to these communities and their input can help ensure that the research projects are conducted in a respectful and authentic manner.

The Principles for Conducting Research in the Arctic state that “for projects involving Arctic Indigenous residents and others as research collaborators or study participants, determine in advance who collects, owns, manages, evaluates, and disseminates the data to allow projects to proceed with a shared understanding of data governance and ownership” (IARPC). This document encourages collaboration and effective communication with Alaska Native communities noting that “tribes and communities often conduct their own research...inquire about ongoing Tribal and community research and priorities and collaborate appropriately. Be aware and respectful of Indigenous Peoples’ practices and protocols for accountability” (IARPC). There is an emphasis on respecting Indigenous knowledge and culture. Researchers are encouraged to “be open to new viewpoints and be aware of and acknowledge differences and biases when discussing analysis and interpretation of data and observations with residents. Arctic Indigenous Peoples hold unique knowledge and understanding of their homelands and can offer valuable collaborative partnerships with scientists. Inclusion of Indigenous Knowledge in

research is encouraged” (IARPC). This emphasis on collaboration and acknowledgment of Indigenous Knowledge is key towards fostering a relationship of trust between the ASC and the local community whose livelihoods could be impacted by the results of ASC research.

The ASC appears to have a collaborative and mutually respectful relationship with the local community by participating in co-management councils to share information with Alaska Native community and by offering internships, mentoring and education outreach. The ASC hosts a monthly seminar series that highlights current research projects that is accessible to the public at large.

### **Authenticity and Trust**

To determine authenticity, I took a close look at a repository object titled “Measurement Data of Polar Bears Captured in the Chukchi and Southern Beaufort Seas, 1981-2017”, which is a data set (see figure 1). The entry includes the Data Download, which contained a Microsoft Excel CSV file and an XML file with the metadata for the record that is expressed in Dublin Core. The metadata follows the Federal Geographic Data Committee (FGDC) Biological Data Profile of the Content Standard for Digital Geospatial Metadata. There is a link to the larger project that the data set is a part of, which includes contact information for the researchers as well as related publications and data for context. There is a summary, followed by the author’s name, a suggested citation, and a list of referenced publications. At the conclusion of the entry there is a DataID number (315), a digital object identifier (doi:10.5066/P9TVK3PX), and the date/time which the data set was posted online and when it was last updated.

Each entry in the repository includes the USGS Trusted Digital Repository seal (see figure 2). In order to display this seal, a USGS Trusted Digital Repository must be “internally certified in the Bureau through a rigorous panel review of questionnaire answers provided by a



prospective Bureau repository or server manager. The process requires prospective repositories to ensure that USGS data are preserved, made accessible, documented, and backed up in a manner that reflects international standards for repositories” (USGS Office of Science Quality and Integrity). Approval lasts for a three-year period, after which the repository is re-evaluated to ensure that it is maintaining all requirements to be considered a USGS TDR. By using this process and the Data Seal of Repository, USGS can “ensure its repositories are robust and reliable, enabling exposure and access to USGS assets by researchers and the public” (Faundeen, 2017). This meticulous process ensures that the data in the repository is trustworthy by researchers and the general public alike.

While I trust the ASC repository and data sets, I am weary of government data stewardship, especially regarding climate change. In an article about climate data rescue efforts in *Physics Today*, Gretchen Goldman, research director for the Center for Science and Democracy at UCS (Union of Concerned Scientists) speaks about this concern noting that “with a president (Trump) who doesn’t appear to respect scientific information, one abuse could be data mysteriously disappearing from websites, or government scientific websites may suddenly have misinformation” (Feder, 2017). It is critical that this data remains publicly accessible and managed in a trustworthy manner.

### **Analysis and Conclusion**

While the ASC data sets are available to fellow researchers, they are also available to the general public. “If data are to be made widely available and used by people with no personal knowledge of their creators, and for different purposes than those for which they were created, then trust must derive from how the data are managed and documented” (Ray, p. 2). The

detailed documentation provided by the ASC for each data set helps to contextualize the research for people who may access the data but are not familiar with the science behind the project. Another way that the ASC makes their data more digestible for the general public is through their monthly seminar series which attendees to either join the seminar in person or to view virtually as a webinar.

Access to provenance is also a key establishing trust in authenticity of repository holdings. “Because digital preservation techniques such as migration inevitably alter the data, authenticity has to be demonstrated by paying attention to such characteristics as provenance (where the data came from) and context (the circumstances surrounding the creation, receipt, storage, or use of data and their relationship to other data)” (Oliver & Harvey, p.56, 2016). The ASC data policy requires “full provenance information that gives credit to original sources, describes how data were acquired for use, and identifies authors and actions for modifications, transformations, and improvements to the data” (ASC Data Policy). By requiring full provenance information, ASC is ensuring full and accurate record that includes information and context for every version of a dataset from creation to its current form.

The robust data policy at the Alaska Science Center provides thorough guidelines to ensure that repository data is managed in a trustworthy and sustainable way. The fact that they have a relationship with the local Indigenous community and collaborate with them through research and various initiatives increases confidence in the ASC as a data steward that strives for authenticity. The careful stewardship of ASC data through its trusted digital repository assists and complements the responsible stewardship of the land and environment by ASC scientists and local resource managers.

## Works Cited

- Alaska Science Center. (n.d.). Alaska Science Center Community Outreach and Engagement. Retrieved October 28, 2020, from [https://www.usgs.gov/centers/asc/science/community-outreach-and-engagement?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/asc/science/community-outreach-and-engagement?qt-science_center_objects=0#qt-science_center_objects)
- Alaska Science Center. (n.d.). Alaska Science Center Data Policy. Retrieved October 28, 2020, from [https://www.usgs.gov/centers/asc/science/alaska-science-center-data-policy?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/asc/science/alaska-science-center-data-policy?qt-science_center_objects=0#qt-science_center_objects)
- Faundeen, J. (2017). Developing Criteria to Establish Trusted Digital Repositories. *Data Science Journal*, 16, p.22. DOI: <http://doi.org/10.5334/dsj-2017-022>
- Feder, T. (2017). Climate-data rescue efforts gear up. *Physics Today*, 70(3), 31–33. <https://doi.org/10.1063/PT.3.3491>
- Interagency Arctic Research Policy Committee. 2018. Principles for Conducting Research in the Arctic. Washington D.C. Retrieved October 28, 2020, from [https://www.iarpccollaborations.org/uploads/cms/documents/principles\\_for\\_conducting\\_research\\_in\\_the\\_arctic\\_final\\_2018.pdf](https://www.iarpccollaborations.org/uploads/cms/documents/principles_for_conducting_research_in_the_arctic_final_2018.pdf)
- Oliver, G. & Harvey, R. (2016). *Digital Curation*: Vol. Second edition. ALA Neal-Schuman.
- Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation*, 72(5), 961-986. doi:<http://dx.doi.org.ezproxy2.library.drexel.edu/10.1108/JD-10-2015-0123>
- Ray, J. (2013) *Research Data Management: Practical Strategies for Information Professionals*, edited by Joyce M. Ray, Purdue University Press, 1-23. ProQuest Ebook Central, <https://ebookcentral-proquest-com.ezproxy2.library.drexel.edu/lib/drexel-ebooks/detail.action?docID=3120304>.
- USGS Data Management. (n.d.) Archive vs. Repository: Is There a Difference?. Retrieved October 27, 2020, from [https://www.usgs.gov/products/data-and-tools/data-management/archive-vs-repository-there-a-difference?qt-science\\_support\\_page\\_related\\_con=0#qt-science\\_support\\_page\\_related\\_con](https://www.usgs.gov/products/data-and-tools/data-management/archive-vs-repository-there-a-difference?qt-science_support_page_related_con=0#qt-science_support_page_related_con)
- USGS Office of Science Quality and Integrity. (n.d.) USGS Trusted Digital Repositories (TDR). Retrieved October 28, 2020, from <https://www.usgs.gov/about/organization/science-support/office-science-quality-and-integrity/trusted-digital-repository>
- USGS ScienceBase. (n.d.) USGS Alaska Science Center - Data Release Backups. Retrieved October 28, 2020, from <https://www.sciencebase.gov/catalog/item/56b3ee22e4b0cc79997fb64b>

## Appendix

Figure 1 – Example ASC Data Repository Record



**USGS**  
science for a changing world

**Alaska Science Center**

HOME Science Publications Data and Tools Connect

ASC Data Repository

**Measurement Data of Polar Bears Captured in the Chukchi and Southern Beaufort Seas, 1981-2017**

**Data Download** [196KB] | [Metadata](#) | [AK Science Portal](#) | [Project Site](#) | Date Range: 1981 - 2017

This dataset includes measures collected on polar bears captured in the Chukchi and Beaufort Seas, 1981-2017 by the U.S. Geological Survey and U.S. Fish and Wildlife Service. Data collected include body length, body mass, axillary girth, skull width and tail lengths. Bears were also aged as described in the methods. For some bears, an adipose tissue sample was collected and percent lipid content was measured, percent body fat was measured via bioelectrical impedance analysis, and/or recent feeding behavior was assessed via gut palpitation or blood urea and creatinine levels, all of which are further described in the methods.

Author(s): Rode, K. D.

**Suggested Citation:**  
Rode, K. D., 2020, Measurement data of Polar Bears captured in the Chukchi and southern Beaufort Seas, 1981-2017: U.S. Geological Survey data release, <https://doi.org/10.5066/P9TVK3PX>.

Referenced Publication(s):  
Rode, K. D., T. C. Atwood, G. W. Thiemann, M. St. Martin, R. R. Wilson, G. M. Durner, E. V. Regehr, S. L. Talbot, G. K. Sage, A. M. Pagano, and K. S. Simac. 2020. Identifying reliable indicators of fitness in polar bears. PLoS One 15(8):e0237444. doi:10.1371/journal.pone.0237444

DataID: 315 | doi:10.5066/P9TVK3PX | Date Posted Online: 2020-07-31 | Last Updated: 2020-07-31 12:33:03

Figure 2 – USGS Trusted Digital Repository Seal

