

# Keeping Up with Terminus Changes – A Data Curation Case Study

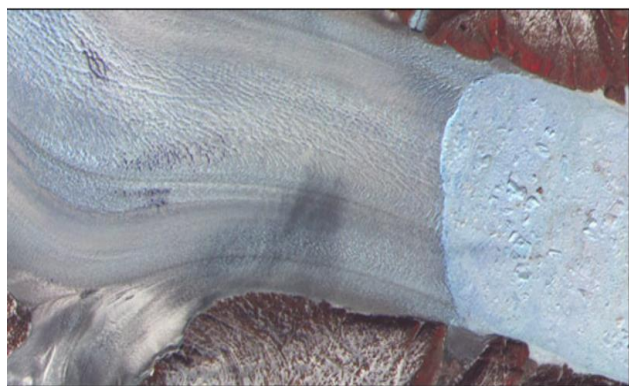
Chung-Yi Hou (hou@illinois.edu)

Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign

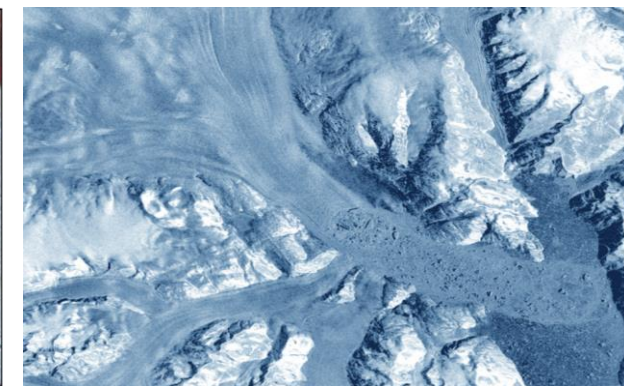
## Introduction

The dataset I curated was from the Advanced Cooperative Arctic Data and Information Service (ACADIS)<sup>1</sup>, and it was based on the study of terminus area changes over time using satellite images for the Helheim and Kangerdlugssuaq glaciers from Greenland.

To understand, identify, and address the curation needs of the dataset, I came up with a 4-step strategy that helped me in creating and implementing the Data Curation Plan.



Satellite Image of Helheim Glacier<sup>2</sup>.



Satellite Image of Kangerdlugssuaq Glacier<sup>3</sup>.

## Aim for the Data Curation Plan

When I received the dataset, there were a total of 3 files: a text formatted ReadMe file, one CSV file containing Helheim glacier's data, and the other in the same format for the Kangerdlugssuaq glacier. The project and the data collection had been completed in 2010, and the dataset's metadata was last updated in March, 2013. The dataset documentation shows that the dataset results not only impact glaciology studies, but also influence the development of instrumentation and processing techniques applicable to other science disciplines, such as volcanology.

As a result, the critical goals for the datasets are:

- Develop a curation plan to provide continuing maintenance and to uphold the dataset's provenance in order to enable long-term use of the dataset.
- Provide strategies to enhance the dataset from the current Research Collection type to the Resource Collection type based on National Science Board's definition<sup>4</sup> to extend the dataset's scientific contributions.

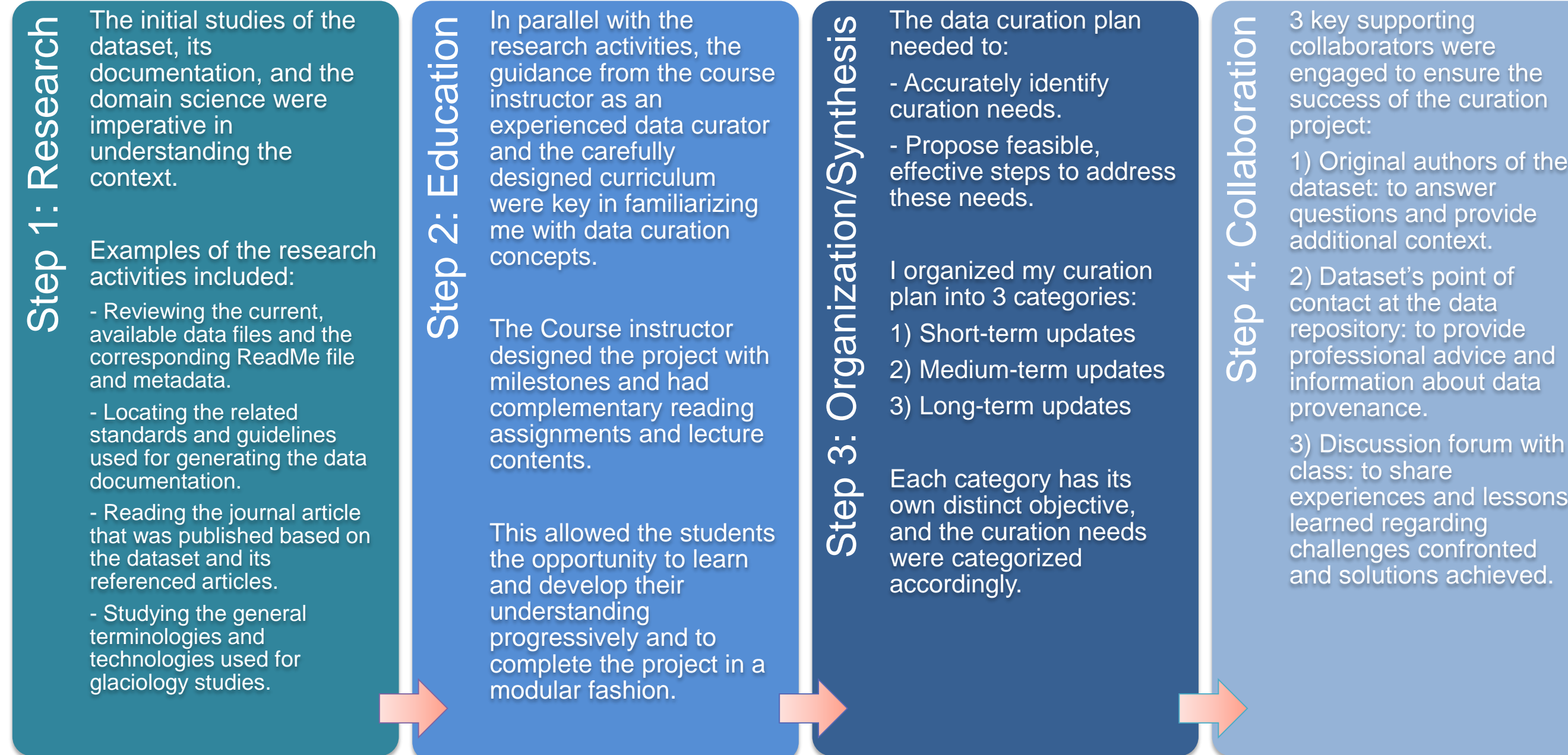
## References

1. Advanced Cooperative Arctic Data and Information Service. (2013). *Terminus position time series: Helheim and Kangerdlugssuaq glaciers, Greenland* [Data file]. Retrieved from [https://www.aoncadis.org/dataset/terminus\\_position\\_time\\_series\\_helheim\\_and\\_kangerdlugssuaq\\_glaciers\\_greenland.html](https://www.aoncadis.org/dataset/terminus_position_time_series_helheim_and_kangerdlugssuaq_glaciers_greenland.html)
2. Allen, Jesse. (January 20, 2006). *Retreat of the Helheim Glacier, Greenland*. [NASA Image]. Retrieved from <http://earthobservatory.nasa.gov/IOTD/view.php?id=6207>
3. ESA. (July 8, 2011). *Satellite's Final Images Focus on Changing Glaciers*. [Satellite image]. Retrieved from [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Space\\_for\\_our\\_climate/Satellite\\_s\\_final\\_images\\_focus\\_on\\_changing\\_glaciers](http://www.esa.int/Our_Activities/Observing_the_Earth/Space_for_our_climate/Satellite_s_final_images_focus_on_changing_glaciers)
4. National Science Board. (2005). The elements of the digital data collections universe and Appendix D. In *Long-lived Digital Data Collections* (pp. 17-23 and 73-78). Retrieved from <http://www.nsf.gov/pubs/2005/nsb0540>

## Acknowledgments

The author would like to thank course instructor Ruth Duerr of the National Snow and Ice Data Center (NSIDC), course teaching assistant Tiffany Chao of Illinois, Gordon Hamilton and Kristin Schild of University of Maine, Toni Rosati of NSIDC, and the Graduate School of Library and Information Science at Illinois for providing and supporting this curation project learning opportunity.

## Method for Creating and Implementing a Data Curation Plan



## Results: Samples of Completed Curation Tasks from Short-term Updates

- The ReadMe file accompanying the dataset was updated so that it has the correct file format (ASCII), the recommended filename naming convention, and the required headers and contents per ACADIS' ReadMe File Template guidelines.
- The current ACADIS metadata fields were mapped to ISO19115:2003 to establish the correlation between the 2 metadata standards and to demonstrate the additional resources needed in order for the ACADIS metadata to become ISO19115:2003 compliant.

### Original ReadMe File<sup>1</sup>

Data are supplied in ascii comma-separated format.  
Column headings are:  
Year = year  
Day = day of year  
Cum day = cumulative day since 01/01/2001  
Zeroed area = relative change in terminus area since minimum observed terminus position  
HLM: day 234, 2005 (August 22)  
KGL: day 62, 2006 (March 3)  
Width normalized = linear change in terminus position (area change divided by terminus width)  
relative to minimum observed terminus position  
HLM: day 234, 2010 (August 22)  
KGL: day 62, 2006 (March 3)

### Updated ReadMe File<sup>2</sup>

On the Same File Name:  
Helheim 1111  
Kangerdlugssuaq 1111  
Metadata File Series: Helheim and Kangerdlugssuaq glaciers, Greenland  
Document Information:  
This README file was created on 2013-03-13 as a class project for Chung-Yi Hou, a student of University of Illinois at Urbana-Champaign's LIS599C01, Professor  
Other File Contacts:  
Principal Investigator: Gordon Hamilton (gordon.hamilton@maine.edu)  
Data Center: Kristin Schild (kschild@nsidc.org)  
Summary Description:  
Use by the Principal Investigator Gordon Hamilton from the University of Maine (see Award #090801), the Collaborative Research project High Resolution View, the year, each terminus position was manually digitized according to the procedure described by Moon and Sougin (2008).  
Background Information:  
The study of the Helheim and Kangerdlugssuaq glaciers (interglacial, glaciological), and general observations to build an understanding of flow in  
Based on the data, the authors did not find a simple relationship between air-temperature and sea surface temperature to the spatial and temporal variability  
Temporal Coverage:  
Beginning Date = 2001-08-24  
Ending Date = 2010-08-20  
Platform/Program/Instrumentation:  
Data were collected using the European Space Agency's (ESA) satellite Terra, launched on December 18, 1999. Terra has been collecting data about its  
Area = Area is a major international earth science satellite mission centered at 60N. Launched on May 4, 2002, the satellite has six different earth-orbitary

### Current ACADIS Metadata<sup>1</sup>

Terminus position time series: Helheim and Kangerdlugssuaq glaciers, Greenland

Metadata	Contribute [meta]Data
Description:	We constructed high temporal resolution time series of terminus position for Helheim and Kangerdlugssuaq glaciers, two major outlet glaciers of the Greenland Ice Sheet. Both glaciers terminate at fjord-like inlets in deep fjords. Calving front positions for each glacier are derived from ~daily resolution satellite images collected by the MODIS Moderate Resolution Imaging Spectroradiometer sensors onboard the Terra and Aqua satellites. We use the MODIS snow product which has a 250 m spatial resolution and is constructed from MODIS bands 1-2 in the visible part of the spectrum. Visible images are available for cloud-free days during periods of solar illumination, corresponding to near-daily coverage between 1 May and 31 October each year, and ~weekly coverage for the remainder of the year. Each terminus position was manually digitized according to the procedure described by Moon and Sougin (2008). Journal of Geophysical Research: Earth Surface in which the calving terminus is a polygon bounded on the sides by parallel lines approximating the lateral margins, and at the upglacier end by an arbitrary line well inland of the minimum observed terminus position. Repeating this procedure for each image yields the change in terminus area with time. We divide the area change by polygon width at each terminus position to obtain a width-averaged linear distance of advance/retreat. Uncertainty is ~250 m. For more details, see Schild and Hamilton (2013), Journal of Glaciology.
Date Created:	2013-03-13 06:21:01
Date Last Updated:	2013-03-13 11:25:10
Author(s):	Schild, K.R. and G.S. Hamilton
Data Center Contact(s):	Liz Schlapik, schlapik@nsidc.org
Metadata Contact(s):	Gordon Hamilton, gordon.hamilton@maine.edu
ICMD Science Keyword:	Cryosphere > Glaciers/Ice Sheets > Glaciers
ISO Topic:	geoscientificInformation
Data Format(s):	Other ASCII

### Mapping of ACADIS to ISO19115:2003<sup>2</sup>

Appendix G: Metadata Mapping between ACADIS and ISO19115:2003

ACADIS Metadata Fields (Fields with Asterisks* are Required)	ISO19115:2003 O = Optional M = Mandatory C = Conditional Type = entry type
*Project	CI_Citation collectionId (O) Type: Free Text
*Title	CI_Citation title (M) Type: Free Text
*Short Name	CI_Citation alternateTitle (O) Type: Free Text
*Description	MD_DataIdentification purpose (O) Type: Free Text
*Author	MD_DataIdentification credit (O) Type: Free Text
*Location Keyword(s)	CI_Citation siteId (O) Type: Free Text MD_DataIdentification descriptionKeywords (O) Type: MD_ScientificKeyword (M) Type: Free Text MD_DataIdentification extent (O) Type: Free Text MD_DataIdentification geographicElement (O) Type: Free Text MD_DataIdentification extentTypeCode (O) Type: Boolean MD_DataIdentification extentTypeCode (O) Type: MD_Identifier or ES_Identifier MD_Identifier authority (O) Type: CI_Citation

<sup>1</sup>Note: The complete files currently can be viewed under the following links: ReadMe File (<https://drive.google.com/file/d/0B-0foVY86qLeTGSxWChZkxvN3M/edit?usp=sharing>); Mapping of ACADIS to ISO19115:2003 (<https://drive.google.com/file/d/0B-0foVY86qLeTGSxWChZkxvN3M/edit?usp=sharing>). The files are a part of the data curation class assignment but have also been submitted to ACADIS for review. The approved updates will be published by ACADIS and can be found using the link under Reference 1 when the updates are made available.

## Conclusions

Although I was not able to complete all the curation activities set out in my data curation plan within the allotted course time, I was able to develop the time management skills to set curation task priority with respect to available resources. I also learned the fundamental data curation skills needed to research, organize, synthesize, and collaborate in order to address the curation needs. The hands-on curation project was a significant learning experience for me to work as a student-curator, and as a result, I was able to gain practical data curation skills to enhance my professional development.

## Results: The Final Data Curation Plan

**Short-term Updates – To complete and make compliant the current documentation and data files.**  
Status: Completed and submitted to ACADIS for review.

- Provided updated information for “Short Name” and “Related Resource” metadata fields per ACADIS' Data Provider's Guide.
- Updated the ReadMe file per ACADIS' ReadMe File Template guidelines.
- Mapped ACADIS' current metadata fields to ISO19115:2003 metadata standard.

**Mid-term Updates – To strengthen the collection's provenance for scientific reproducibility and reusability.**

Status: More than 50% of the items identified have been collected and provided to ACADIS for review. However, further investigations and discussions are needed.

- Depending on the further availability and accessibility of the Principle Investigator and his co-author, additional discussions can provide further dataset context and information to be included with the dataset.
- Discussion with ACADIS is needed to determine the feasibility and appropriate methods for integrating these information.

- Moderate Resolution Imaging Spectroradiometer (MODIS) raw data (image files).
- Instrument information regarding MODIS sensor.
- Instrument information regarding Terra and Aqua satellites.
- Margin of error or uncertainty for digitization and analysis process.
- Repeatability test procedure used for image digitization.
- Computer platform, software, algorithms, and/or procedures that were used during image digitization and data analysis.
- Links to the auxiliary data of the collection (Weather record, ocean temperature, cyclic tidewater dynamics, atmospheric/oceanic forces, mélange behavior, melt seasons length, fjord bathymetry, ice thickness, sub-glacial topography, and earthquake records).
- Simulation models.
- Researchers' observations or notes during the project.
- Other resources, such as projects, research groups, and journals, with related interests.
- Future, related research areas and topics.

**Long-term Updates – To further enhance the collection's inter/intra-disciplinary discovery and uses.**

Status: Current databases from which the dataset can be located have been identified. Additional discussion with ACADIS is needed to determine further, available resources.

- Review and verify the metadata mapping between ACADIS and ISO19115:2003.
- Review ISO19115:2003 and add additional metadata fields from ISO19115:2003 to ACADIS' metadata requirements.
- Determine if an identification scheme should be used for the data collection.
- Determine linkages to additional databases.