Data Sharing and Archiving

Bob Cook

ORNL Distributed Active Archive Center
Oak Ridge National Laboratory
Oak Ridge, TN

ABoVE Science Team Meeting
St. Paul, Minnesota
October 2, 2015

ABoVE Solicitation

Terrestrial Ecology Solicitation: NNH14ZDA001N-TE

- ORNL DAAC as archive for ABoVE

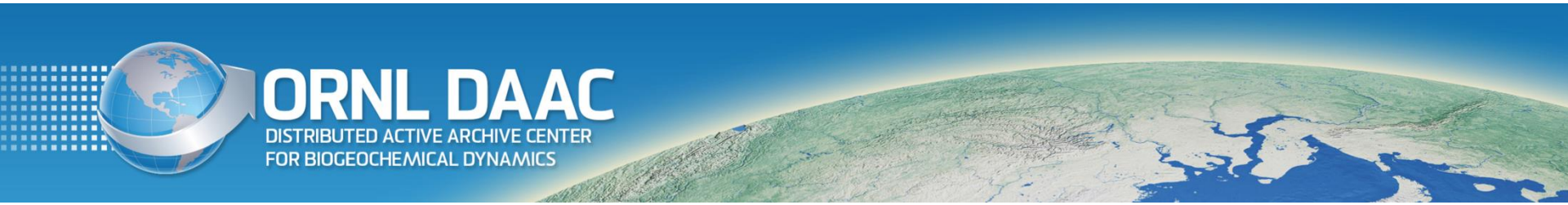
  …much of the data collected through ABoVE will be archived and distributed by the ORNL DAAC.

  NASA anticipates the possibility that some ABoVE data might be more appropriately archived at another NASA DAAC or other long-term archive…

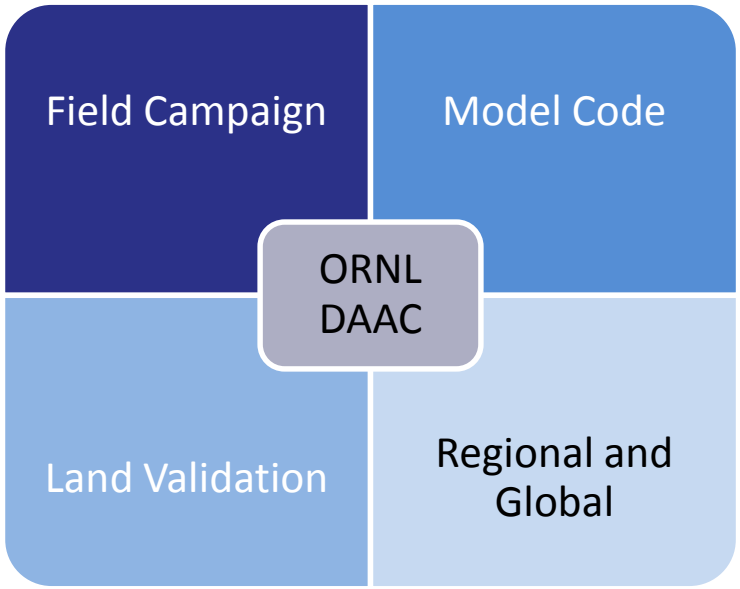
Topics

- ORNL DAAC
  - Curation
  - Archiving data from manuscripts
- Science Data Working Group

Oak Ridge National Laboratory Distributed Active Archive Center



Assemble, distribute, and provide data services for a **comprehensive archive of terrestrial biogeochemistry and ecological dynamics observations and models** to facilitate research, education, and decision-making in support of NASA's Earth science.

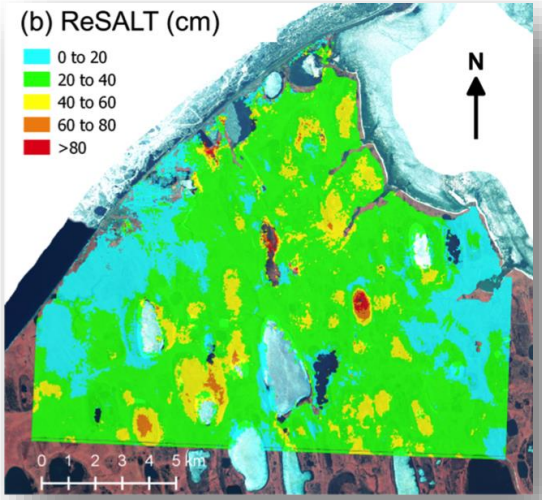


http://daac.ornl.gov

Data Products Archived

ABoVE (3 from Kevin's Project)

Liu, L., K. Schaefer, A. Chen, A. Gusmeroli, E. Jafarov, S. Panda, A. Parsekian, T. Schaefer, H. A. Zebker, T. Zhang. 2015. Pre-ABoVE: Remotely Sensed Active Layer Thickness, Barrow, Alaska, 2006-2011. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1266

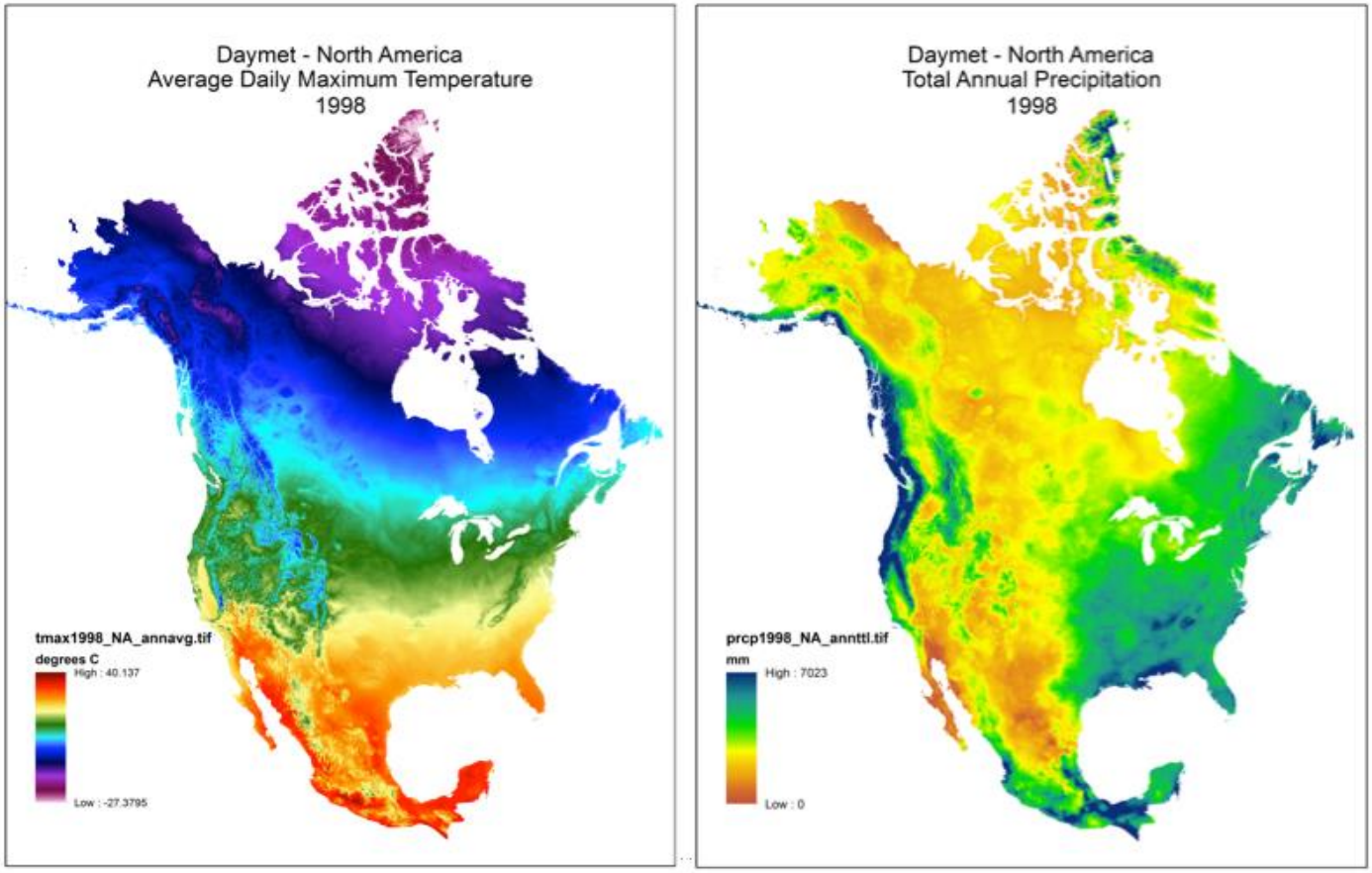


CARVE

Veraverbeke, S., B.M. Rogers, and J.T. Randerson. 2015. CARVE: Alaskan Fire Emissions Database (AKFED), 2001-2013. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1282

Daymet for North America

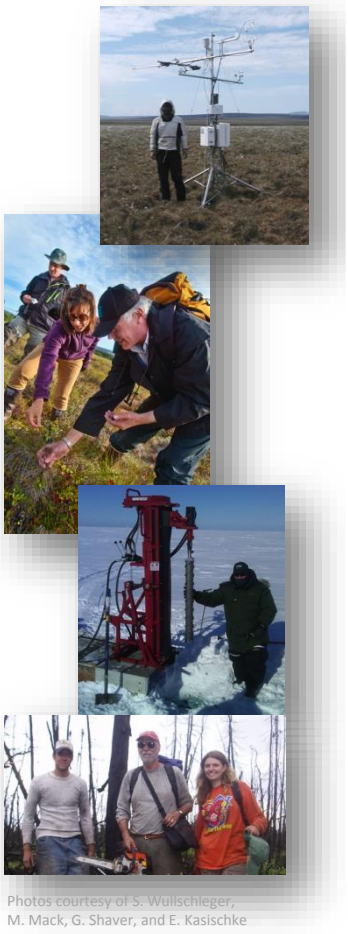
Available in Fall 2016



Daily 1-km resolution estimates of min and max temperature, precipitation, vapor pressure, snow water equivalent, and short-wave radiation as a continuous surface

- Current product: 1980 – 2014, North America south of 52N
- doi: 10.3334/ORNLDAAC/1219

Curation and Archive



Data Files

Data Information

Documents & Supplemental Files

Curation

Quality review

Prepare Documentation

Release

Archive

Data Set
- Data files
- Documentation
- Metadata

Services
- Discovery
- Access
- Visualization

Data Providers

ORNL DAAC
for a user 20 years from now

Data Users

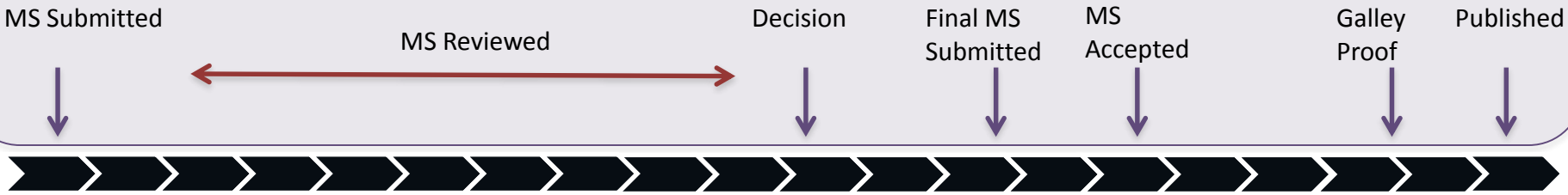
ORNL DAAC New Service

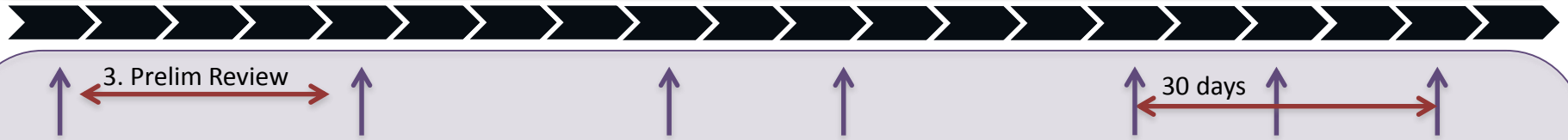
archiving data associated with an article

- Many journals now require that original data used in a manuscript be archived
- ORNL DAAC has a new workflow for manuscript data that provides an archived data product and citation
  - During the manuscript review process
- Requires close coordination between authors and the DAAC

Timeline for Archiving a Data Set for a Manuscript

Manuscript Publication at a Journal





1. Contact DAAC
2. Upload information
   - document
   - questionnaire
   - data files

4. Agreement letter
   - DAAC commitment
   - Offer temporary storage
   - Issues identified
   - Author commitment to timely response

5. Any Issues resolved
   - New files uploaded, if needed
   - Questions answered

6. DAAC Notified of Decision

7. Begin Ingest

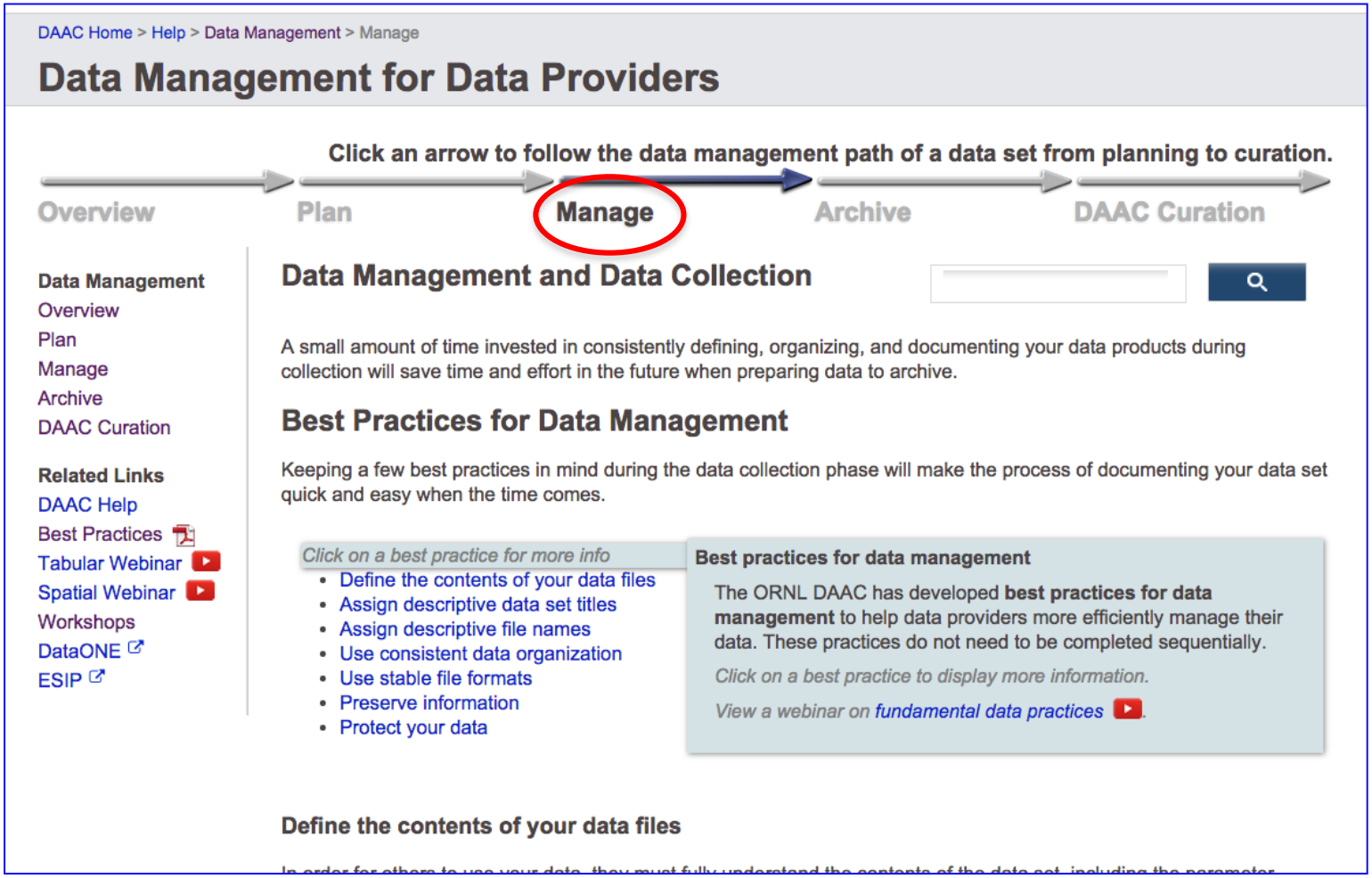
8. Citation provided and DOI Assigned

9. Data Archived and author notified

Data Set Archival at ORNL DAAC

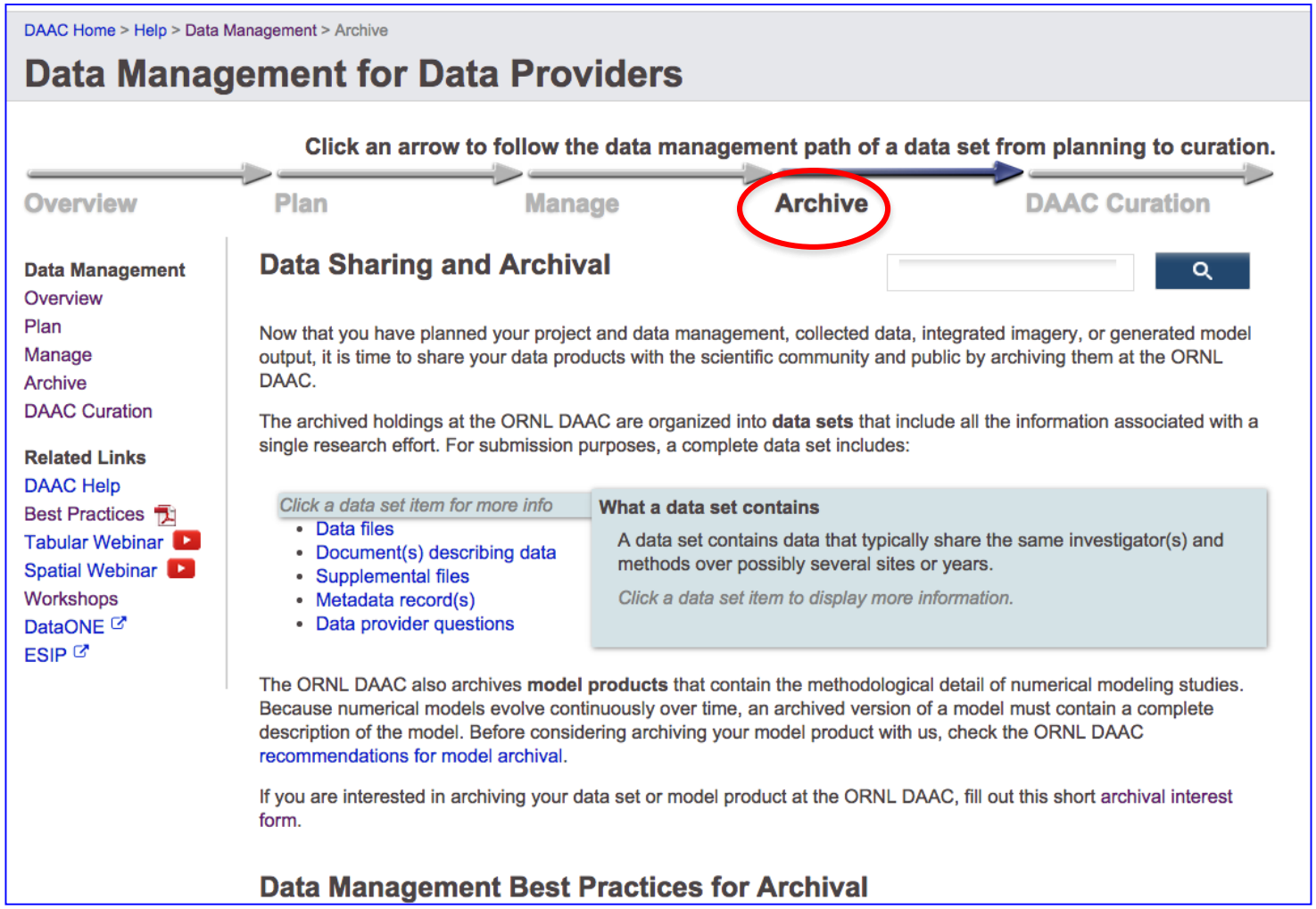
ORNL DAAC Best Practices

http://daac.ornl.gov/PI/pi_info.shtml



ORNL DAAC Best Practices

http://daac.ornl.gov/PI/pi_info.shtml



Science Data Working Group:
Example Topics

- Develop ABoVE Data Policy
- How can ABoVE facilitate sharing data to address ABoVE team research questions?
  - Develop: ABoVE projection, variable names, formats, methodologies, etc.
- Identify and make available critical existing data products.
- What data products are required to support the topics discussed at the Thematic and Crosscutting Breakout sessions?

Data Policy

- Based on NASA's Open Data Policy
- Elements (suggested)
  - What is ABoVE data?
  - Giving credit to data collectors
  - Protecting the rights of students
  - Sharing and archiving data
  - Sharing and archiving model source code
  - Acknowledging ABoVE project
- Example Policies: NACP, LBA-ECO, SAFARI 2000
- Policy should be developed by the ABoVE Science Team

Resources

- ORNL DAAC http://daac.ornl.gov
- Data Management Best Practices http://daac.ornl.gov/PI/pi_info.shtml
- Bob Cook *cookrb@ornl.gov*


@ornldaac

Training for Data Management Best Practices

1. Define the contents of your data files
2. Define the parameters
3. Use consistent data organization
4. Use stable file formats
5. Assign descriptive file names
6. Preserve processing information
7. Perform basic quality assurance
8. Provide documentation
9. Protect your data
10. Preserve your data

Curation at the ORNL DAAC

Files and documentation suitable for a user 20 years from now

- Documentation
- Metadata
- Quality Review
  - Files received as sent
  - Documentation describes files
  - Variables and units defined (community of practice)
  - Time and geospatial information standardized
  - Values are physically reasonable