ARM

CLIMATE RESEARCH FACILITY

ARM Data Center Experience in Linking Big Data

Giri Palanisamy

ARM – www.arm.gov

Climate Change Science Institute

Oak Ridge National Laboratory, USA

Presented at: CENDI-NFAIS

December 11, 2012

Climate Change Science Institute Data Centers at a Glance

http://climatechangescience.ornl.gov/

	ARM Archive	ESGF*	CDIAC	DAAC
Sponsor	DOE	DOE	DOE	NASA
Type of data	Atmospheric processes, cloud dynamics	Climate model datasets, climate observations for intercomparision	Atmospheric gases, emissions data, AmeriFlux	Biogeochemical dynamics, carbon cycle, FLUXNET
Products	> 3000	32512	1269	986
Archive Size	> 300 TB	850 TB	~ 400 GB	~ 1 TB + 60 TB of MODIS data
Users/year	~1,500	~25,000	~300,000	~10,000
Year Started	1991	2000	1982	1993

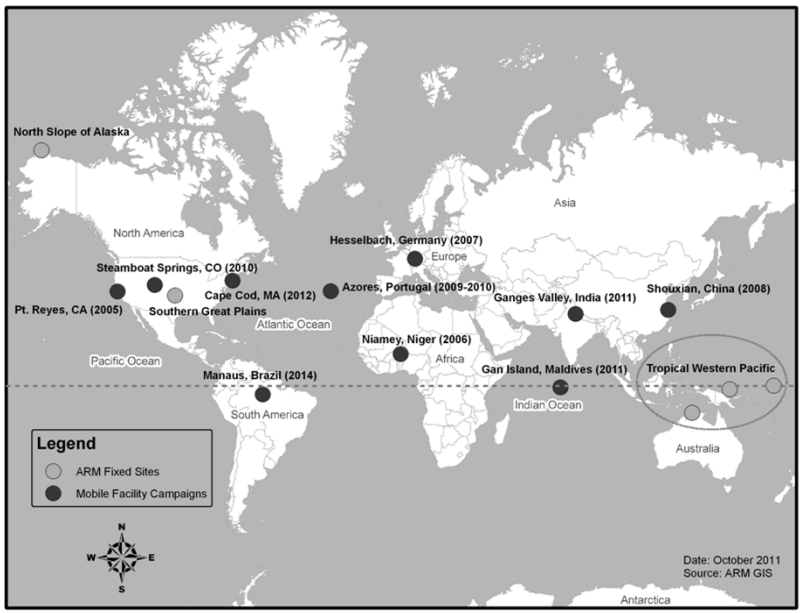
* - Federated data archived at ORNL, PCMDI, BADC, NCI, NERSC, WDCC and JPL Gateways, ORNL numbers are displayed inside the parenthesis

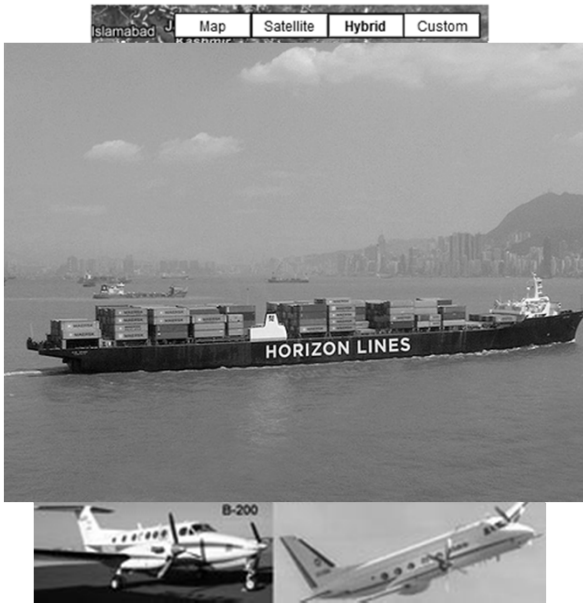
- Broad range of users: modelers, field scientists, educators, NGO's, general public
- Broad range of atmospheric, land process, and ocean data
- Additional projects within and between data centers

ARM CLIMATE RESEARCH FACILITY

U.S. DEPARTMENT OF ENERGY | Office of Science

ARM permanent sites provides long-term data. Mobile sites and aircraft increase diversity.





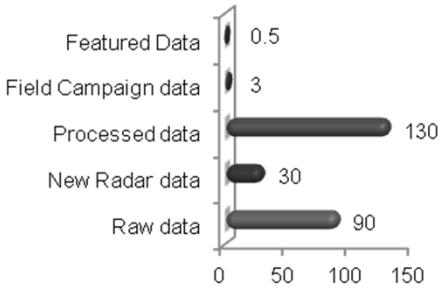
- Southern Great Plains (1993)
- North Slope of Alaska: Barrow (1998) and Atqasuk (1999)
- Tropical Western Pacific: Manus (1996), Nauru (1998), and Darwin (2002)
- First ARM Mobile Facility (2005); Second ARM Mobile Facility (2010)
- ARM Aerial Facility (2007)

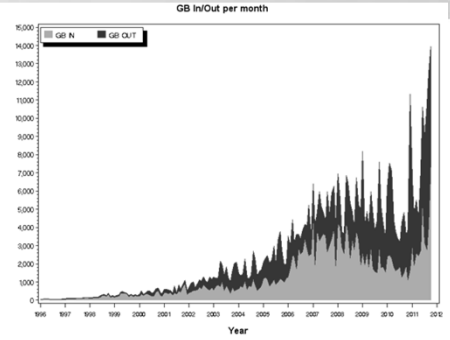
ARM CLIMATE RESEARCH FACILITY

U.S. DEPARTMENT OF ENERGY | Office of Science

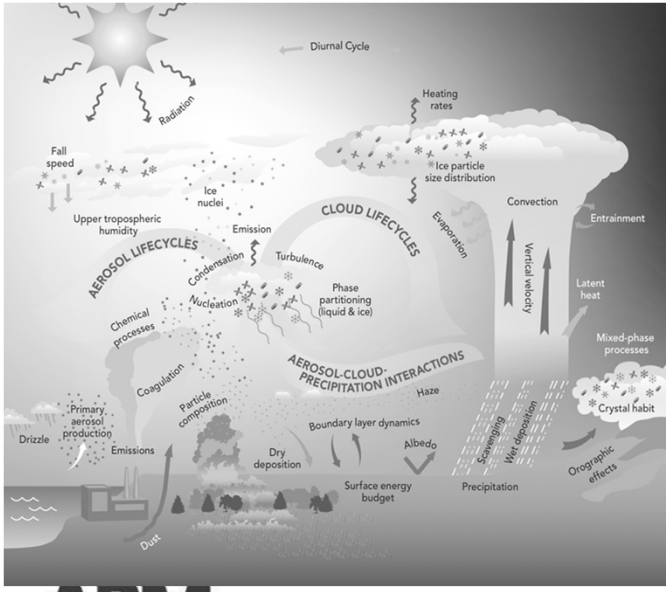
Atmospheric Radiation Measurement(ARM) - Data Archive

- DOE national user facility for global climate change research
- Data from routine operations and field experiments are stored and distributed through the Archive

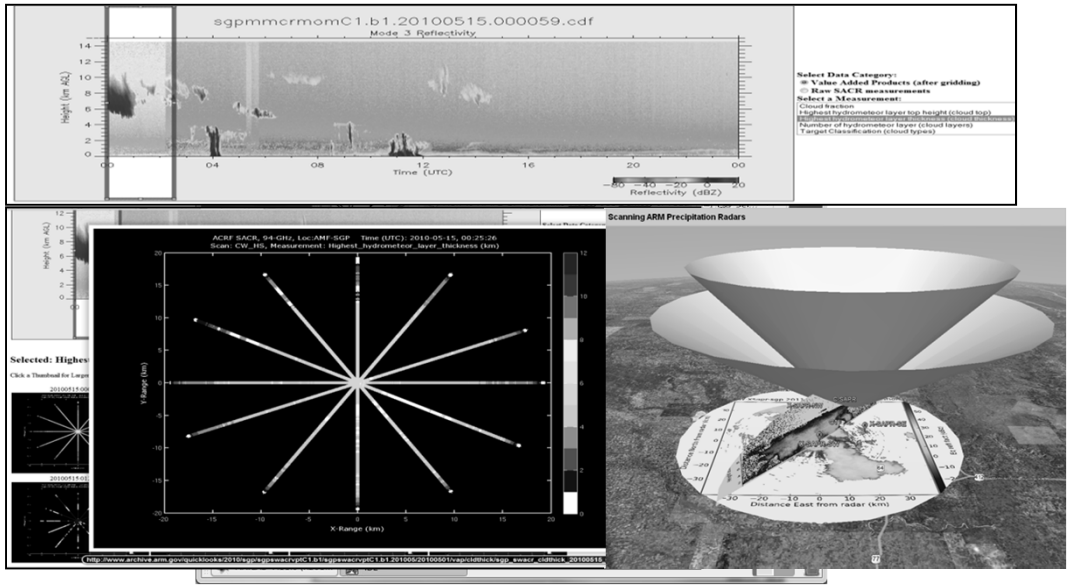




ARM Measurements



Advanced Visualizations using ARM Cluster



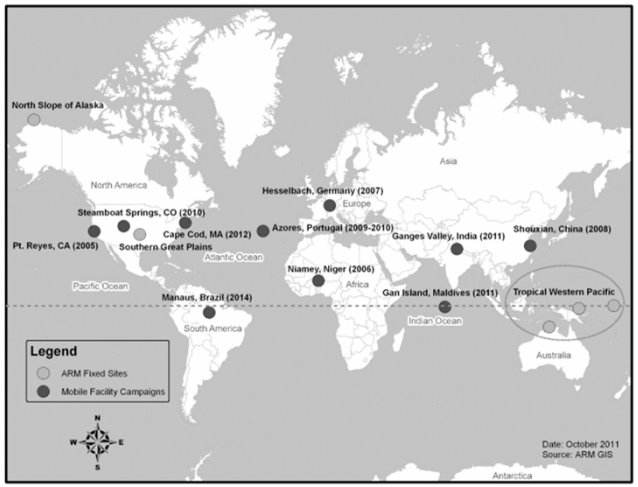
ARM CLIMATE RESEARCH FACILITY

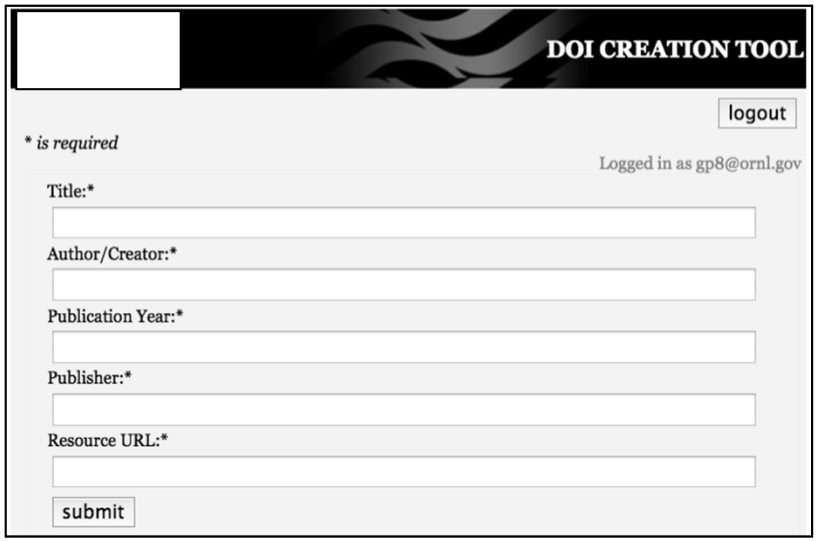
U.S. DEPARTMENT OF ENERGY | Office of Science

ARM Data Citation Service

Goal

- Allow users to cite the exact ARM data used in their research/publication
- Allows future data users and the project to easily track the data used in various articles





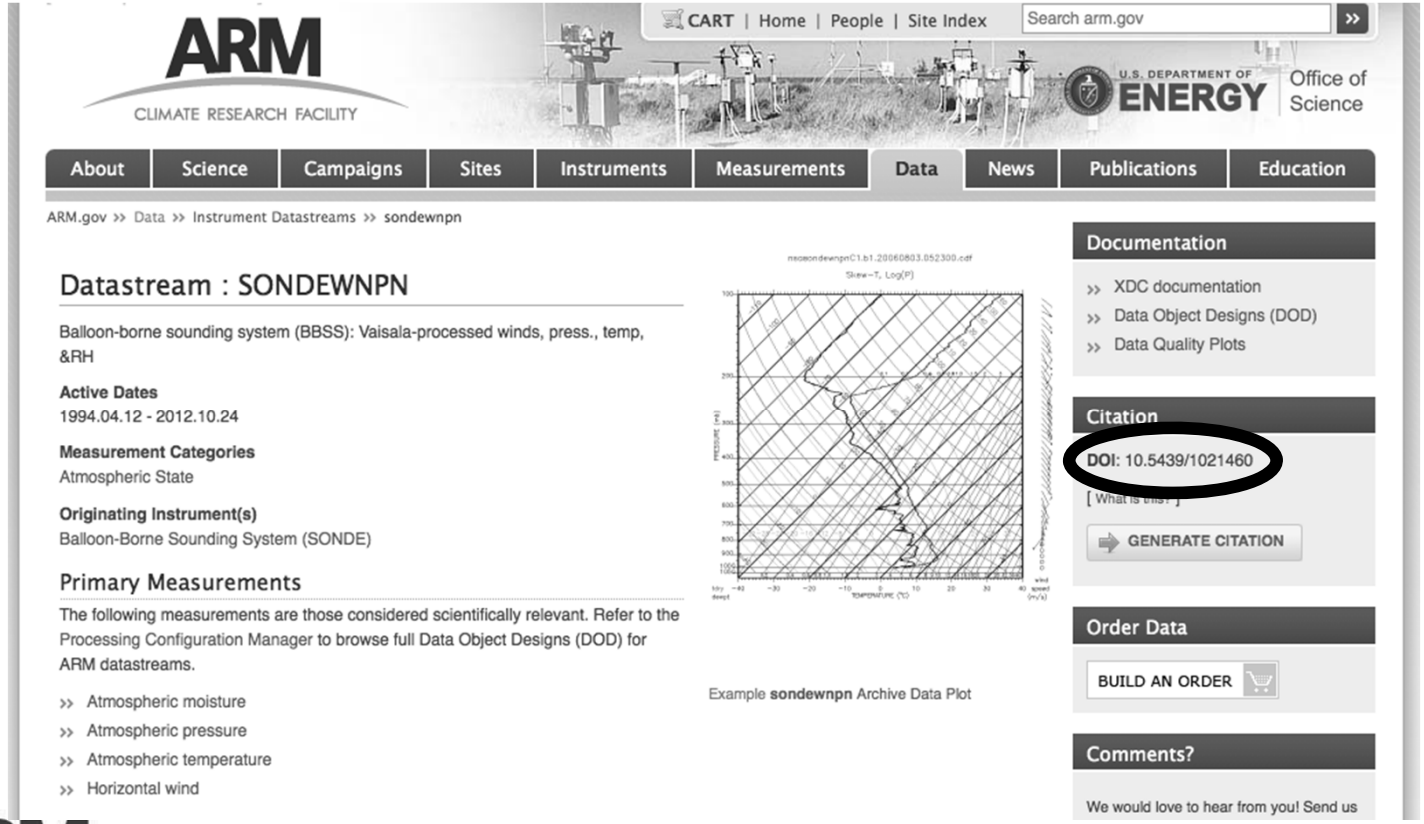
ARM Data Citation Service

The Challenge

- Millions of data files from over 3000 data products
- Most of them are continuous data products
- Large user community and complex use of data
- Data is also published via other portals

ARM Data Citation Service

- **Using DOE Office of Scientific and Technical Information (OSTI) DOI service (Mark Martin)**
- **DOIs are assigned at the data collection level**



ARM - Recommended Citation Format

Several citation formats are possible using DOIs. ARM encourages users to include the following information when citing ARM data:

- *Author*
- *Original publication date*
- *Update period, if applicable (daily, monthly, quarterly, yearly, etc.)*
- *Dataset name*
- ***Dates used****
- ***Locations**** *(latitude/longitude, site name, and facility identifier)*
- *Editor(s) or compiler(s)*
- *Place of publication*
- *Publisher*
- ***Date accessed****
- ***DOI****

**** Needed for future replication of data requests***

More Information: http://www.arm.gov/data/docs/doi-guidance

ARM CLIMATE RESEARCH FACILITY

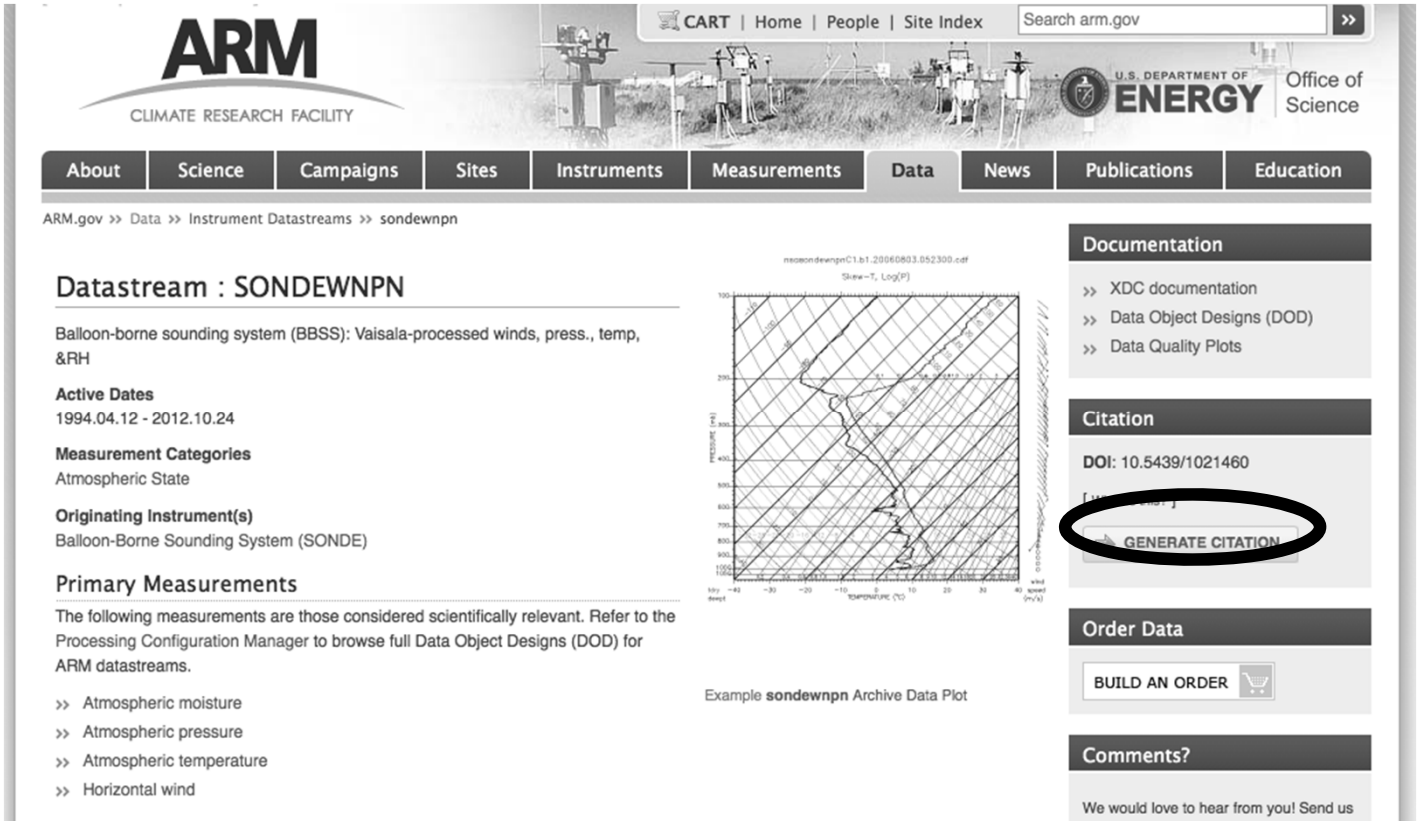
ARM - Citation Format Example

Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1994, updated daily. SONDEWNPN. Oct. 2010–March 2011, 36° 36' 18.0" N, 97° 29' 6.0" W: Southern Great Plains Central Facility (C1). Compiled by R Coulter, J Prell, M Ritsche, and D Holdridge. Oak Ridge, Tennessee, USA: ARM Data Archive. Data set accessed 2011-04-13 at http://dx.doi.org/10.5439/1021460.

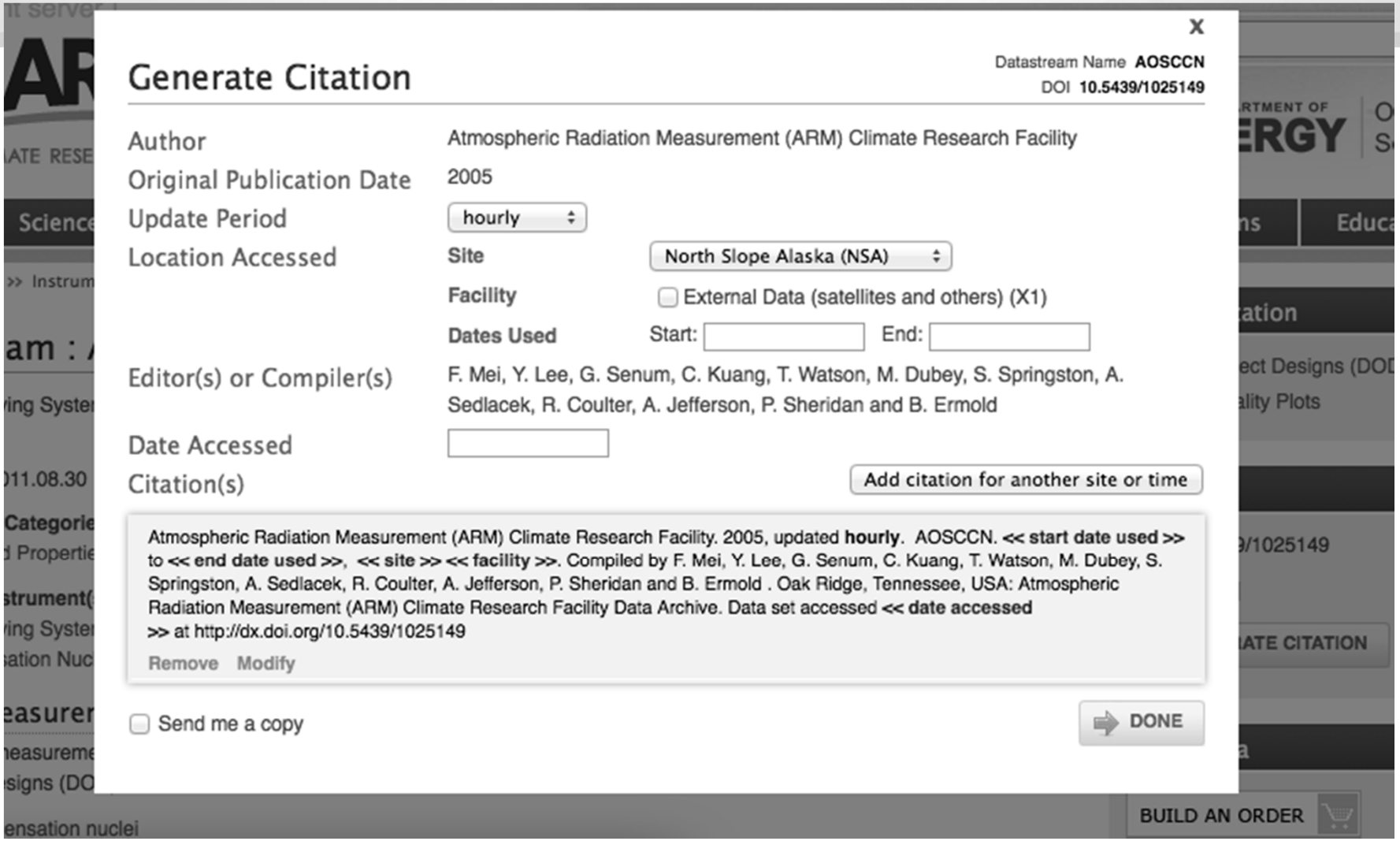
More Information: http://www.arm.gov/data/docs/doi-guidance

ARM CLIMATE RESEARCH FACILITY

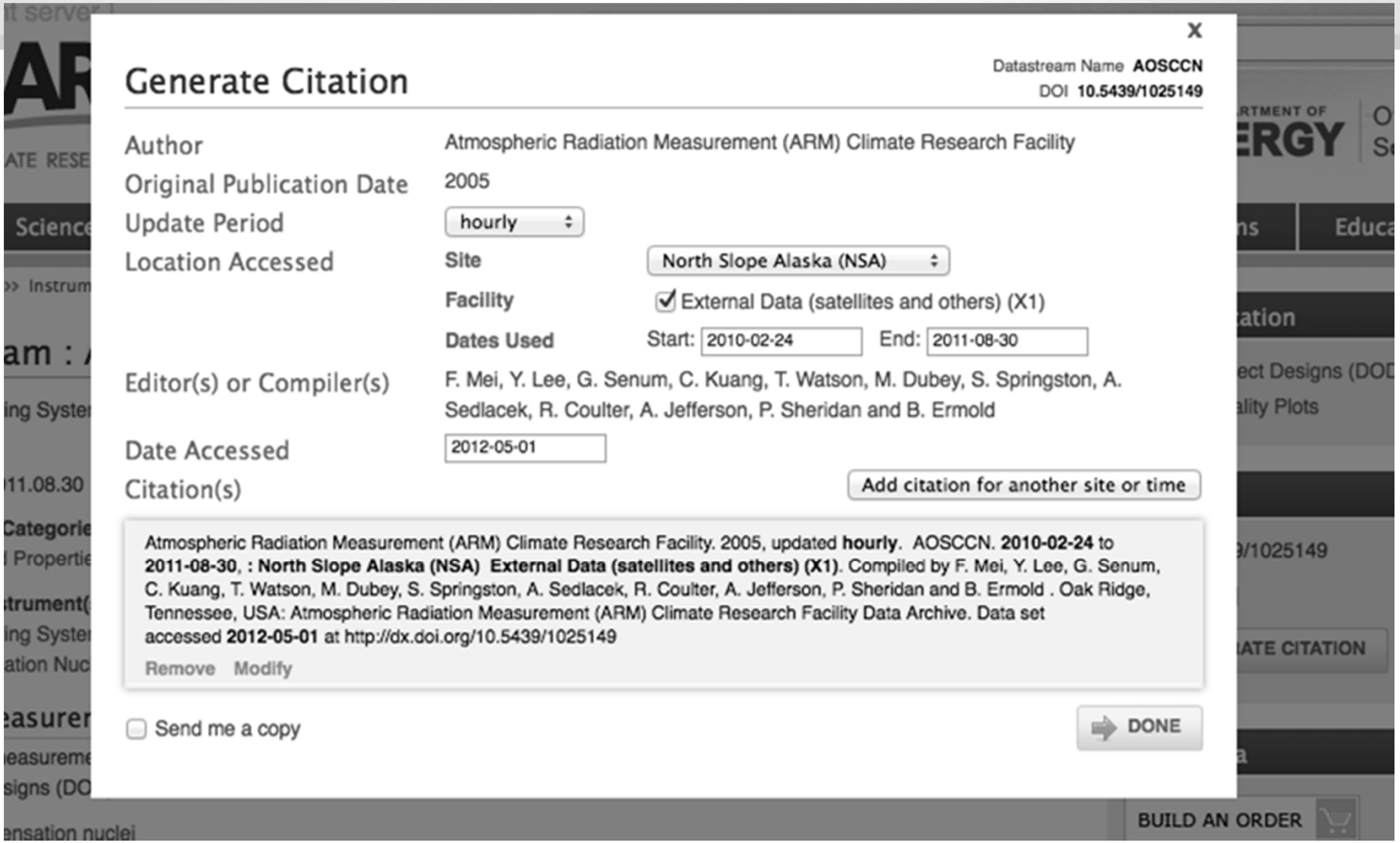
ARM - Citation Resolver Service



ARM - Citation Resolver Service



ARM - Citation Resolver Service



Showing potential users where data have been cited

NPP MULTI-BIOME: NPP AND DRIVER DATA FOR ECOSYSTEM MODEL-DATA INTERCOMPARISON

Below are files for t

Project: NET PRI
NPP (net primary pr
ecosystem types. Th
carbon cycle and to

Data Set: NPP MUL
The Ecosystem Mod
net primary producti
measurements and
This Data Set has a

- npp_doc.pdf
- NPP_TM196

You will need Adobe

Citation:
Olson, R. J., J. M. O
Intercomparison. Da
Tennessee, U.S.A. d

Publications Using

Data Set Files: (47.
All Data Taken At La

NPP MULTI-BIOME: NPP AND DRIVER DATA FOR ECOSYSTEM MODEL-DATA INTERCOMPARISON

The following 12 publications cited the product above:

Cuntz, M., P. Ciais, G. Hoffman, and W. Knorr. 2003. A comprehensive global three-dimensional model of 18-O in atmospheric CO2. Journal of Geophysica No. D17, doi:10.1029/2002JD003153.

Roxburgh, S. H., D. J. Barrett, S. L. Berry, J. O. Carter, I. D. Davies, R. M. Gifford, M. U. E. Kirschbaum, B. P. McBeth, I. R. Noble, W. G. Parton, M. R. Rau Roderick. 2004. A critical overview of model estimates of net primary productivity for the Australian continent. Functional Plant Biology 31(11): 1043-1059.

Zhao, M., F. A. Heinsch, R. R. Nemani, and S. W. Running. 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. Environment 95: 164-176.

Zaehle, S., and S. Sitch. 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochemical Cycles 19: art-

Hickler, T., I.C. Prentice, B. Smith, M.T. Sykes, and S. Zaehle. 2006. Implementing plant hydraulic architecture within the LPJ Dynamic Global Vegetation M Ecology and Biogeography 15:567-577.

Naegler, T., and I. Levin. 2006. Closing the global radiocarbon budget 1945-2005. Journal of Geophysical Research-Atmospheres 111.

Zhao, M. 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological rea Geophysical Research 111:G01002.

Van der Werf, G.R., J.T. Randerson, L. Giglio, G. J. Collatz, P.S. Kasibhatla, and A.F. Arellano, Jr. 2006. Interannual variability in global biomass burning em to 2004. Atmospheric Chemistry and Physics 6:3423-3441.

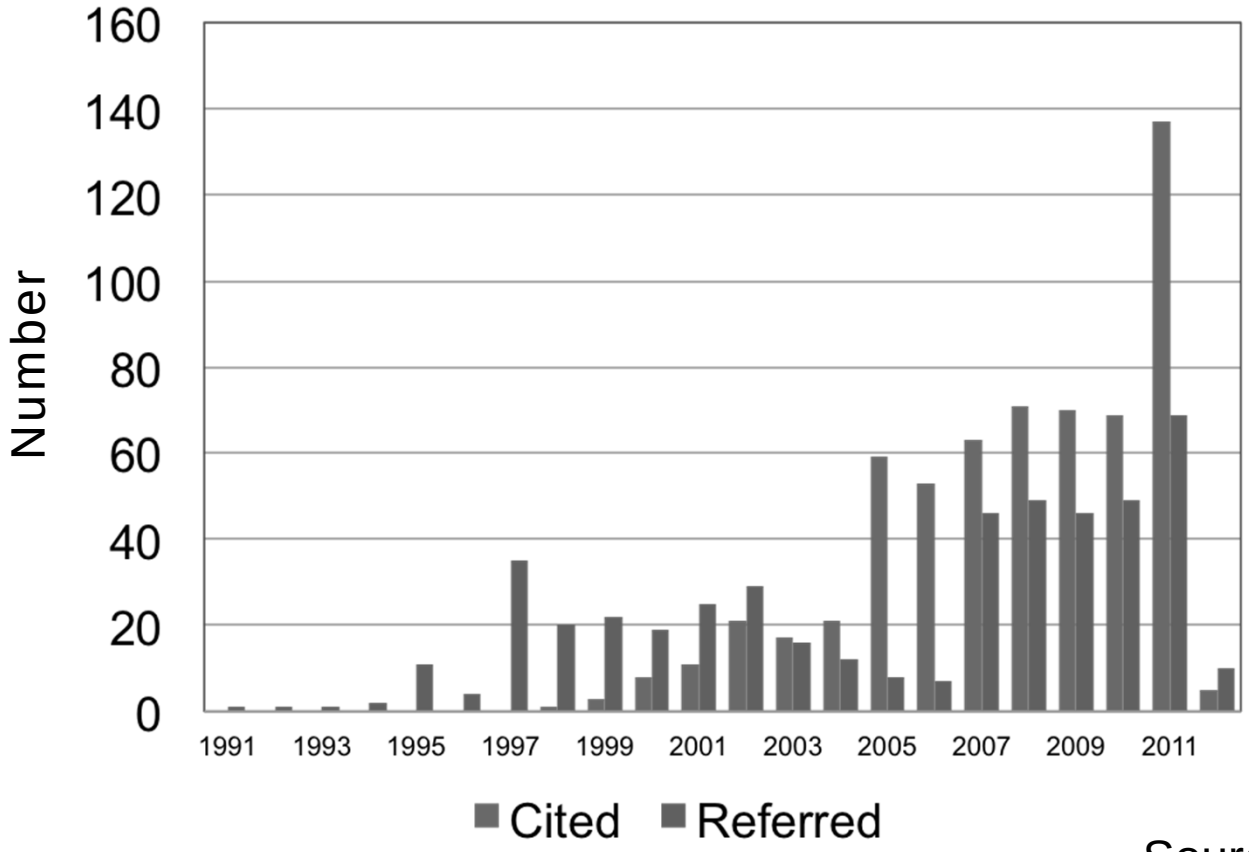
Luyssaert, S., I. Inglima, et al. (2007). "CO2 balance of boreal, temperate, and tropical forests derived from a global database." Global Change Biology 13(1

Randerson, JT.,Hoffmanw, FM.,Thorntonz, PE.,Mahowald, NM.,Lindsayz, K.,Leez, Y.,Nevison, CD.,Doney, SC.,Bonanz, G.,St ckliww, R., Global Change Bi Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models, Online

Refuge, CLW., USGS (2009). Greater Platte River Basins: Science to Sustain Ecosystems and Communities

Example of Scientific Impact

Data Products used in literature



Source: Bob Cook, ORNL

ARM CLIMATE RESEARCH FACILITY

U.S. DEPARTMENT OF ENERGY | Office of Science

Data specific DOI Documents

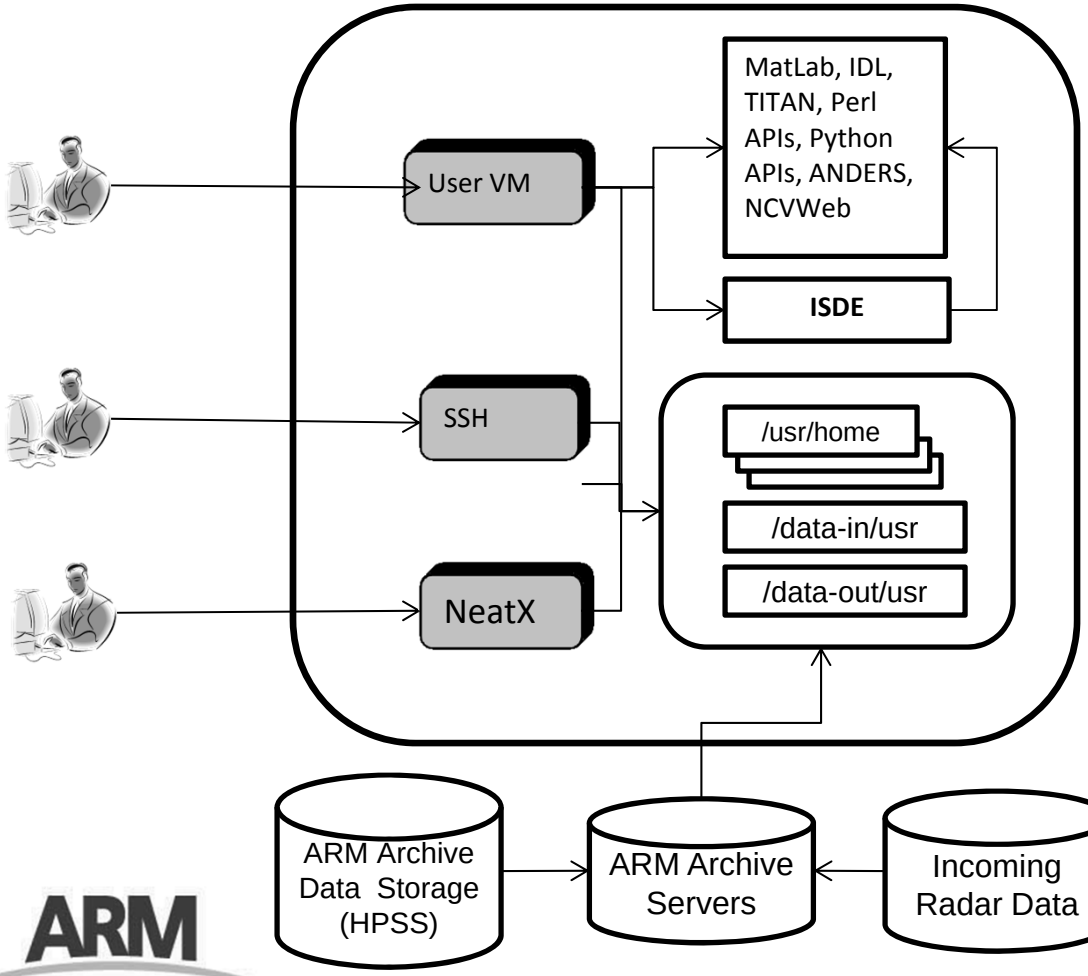
- http://www.arm.gov/data/docs/doi-guidance
- http://www.cdlib.org/services/uc3/ezid/identifiers.html
- http://ipydis.org/data/citations.html

Challenges and next Steps

- DOI strategy for data processed using ARM Big Data Clusters
- Work with publishers to link the DOI/data, register the visualization widget?
- Get metrics for citations and references from the publishers

ARM Computing Cluster

a community resource for ARM data visualization and analysis



ARM CLIMATE RESEARCH FACILITY

U.S. DEPARTMENT OF ENERGY | Office of Science

Next Steps

- Work with publishers to link the DOI/data, register the visualization widget?
- Get metrics for citations and references from the publishers

Comments/Questions?

Giri Palanisamy

palanisamyg@ornl.gov

www.arm.gov

ARM CLIMATE RESEARCH FACILITY

U.S. DEPARTMENT OF ENERGY | Office of Science