Citing Data Sets in the Literature: ORNL DAAC Practices

Robert Cook, Suresh SanthanaVannan, and Daine Wright
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN

MODIS Science Team Meeting
Columbia, MD
April 30, 2014
Presentation Outline

- Metrics of Impact
- Citing Data
- MODIS Data Citations
- Challenges
• ORNL DAAC archive for the NASA’s Terrestrial Ecology Program
  – Biogeochemical dynamics

• Data Holdings (1,085 data sets)
  1. Field Campaigns
  2. Land Validation
  3. Regional and Global Data
  4. Model Archive

• Provide MODIS Land Product Subsets
  – Makhan Virdi’s Poster
Motivation for Data Citations

• User Working Group in 1998 requested a metric for evaluating the impact of the ORNL DAAC on Science

• Use data product citations to indicate how many DAAC data sets have been used in peer-reviewed papers, dissertations, or policy reports
  – Data citation indices
  – Analogous to article citation indices
ORNL DAAC Data Citation Policy (est. 1998)
DOIs add in 2007

Data Product Citation Policy

Citation Policy
To acknowledge the scientists who have provided products, we request that you include a bibliographic citation to all ORNL DAAC products that you use in your publications. Such citations will help others find the products and see how they have been used.

Citation information is provided in the documentation that accompanies all our data products. If you have questions about how to cite ORNL DAAC data products or services, please contact the ORNL DAAC User Services Office (USO).

An editorial "Citations to Published Data Sets" describes the rationale and advantages for data set citations.

The content of a Data Product Citation should include as much of the following information as appropriate:
- contributing investigators/authors
- year of publication
- product title
- medium (for items other than downloaded files; e.g., CD, DVD, tape, etc.)
- online location (i.e., URL)
- publisher
- publisher's location
- date accessed
- digital object identifier

Citation Style

• On-Line Data Set

• Web Page

• MODIS Subset
Characteristics of an identifier (DOI)

• Persistent
  – Registered with The DOI System [http://dx.doi.org/](http://dx.doi.org/) through DataCite

• Actionable
  – [http://dx.doi.org/10.3334/ORNLDAAC/1086](http://dx.doi.org/10.3334/ORNLDAAC/1086)

• Specific
  – Links to the data set

• Complete
  – Links to data and the information needed to understand and use the data
    • Data set landing page
Purpose of Data Citations

- Connect publications to their underlying data
- Facilitate data and science transparency and reproducibility
- Give scientists, data centers, and funders credit for producing and curating data sets
- Understand how data are used
- Track the products that derive from data
**Title:** SAFARI 2000 MODIS L3 Albedo and Land Cover Data, Southern Africa, Dry Season 2000

**Project(s):** SAFARI 2000

**Investigator(s):** MOODY, E.G.  
KING, M.D.  
PLATNICK, S.E.  
SCHAAF, C.  
GAO, F.

**Data Set Documentation:** http://daac.ornl.gov//S2K/guides/S2K_MODIS_L3_albedo_guide.html

**Access Restrictions:** PUBLIC

**Data Set Location:** The Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC)

**Data Center Contact:** ORNL DAAC User Services Office Oak Ridge National Laboratory Oak Ridge, Tennessee 37831 USA FAX: +1(865)574-4665 - ornldaac@ornl.gov Phone: +1(865)241-3952

**Data Center URL:** http://daac.ornl.gov/


**Download Data Sets:** SAFARI 2000 MODIS L3 Albedo and Land Cover Data, Southern Africa, Dry Season 2000

**Parameter Description:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sensor</th>
<th>Source</th>
<th>Term</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAND COVER</td>
<td>MODIS</td>
<td>TERRA</td>
<td>LAND USE/LAND COVER</td>
<td>LAND SURFACE</td>
</tr>
<tr>
<td>REFLECTANCE</td>
<td>MODIS</td>
<td>TERRA</td>
<td>SURFACE RADIATIVE PROPERTIES</td>
<td>LAND SURFACE</td>
</tr>
</tbody>
</table>
Citations to data sets indicate data reuse: Strack et al used five ORNL DAAC data sets


Articles may “refer” to the data set

were assimilated using the WRF model constrained by the NCEP reanalysis product. The anthropogenic NO\textsubscript{x} and VOCs emissions are obtained from the inventory by Zhang et al.\cite{43} The biomass burning emissions are taken from the Global Fire Emissions Database, Version 2 (GFEDv2.1; available at http://daac.ornl.gov/). The lightning NO\textsubscript{x} emission is parametrized as by Choi et al.\cite{44}
Estimating Scientific Impact: Steps Taken

- Review on-line services (e.g., Web of Science, Elsevier’s Science Direct) to see if ORNL DAAC or DAAC Projects are mentioned or if data have been cited in papers
  - full text search, reference list

- Develop a database of articles (article DOI) and the data sets used in those articles
Use of ORNL DAAC Data in the Literature

Data Sets and Projects Used in Publications (Cited and referred, 1994 - 2013)

Cited
Referred

Referred indicates a call out in the text
Show users: how data used and where data cited

GLOBAL FIRE EMISSIONS DATABASE, VERSION 3.1

Download Data

Below are files for this data set. Click on a file link to display that file on your browser. The Companion Files are small but the Data Set Files can be quite large.

<table>
<thead>
<tr>
<th>Year</th>
<th>Citation</th>
<th>DOI</th>
</tr>
</thead>
</table>

Citation:

Publications Using This Data Set

Download Data Set Files: (1665.4 MBytes in 12 Files)
All Data Taken At Latitude: 89.50N, Longitude: 179.50E

You may order only the files you are interested in by checking the 'Add to Cart' box and then the 'Add' button below. You may also click on the file link to see the file and to save it to your computer if you wish. You will need to display and save any companion files listed above. You may also order the complete data set by checking the 'Add Data Set' button.
Assessing the coupling between surface albedo derived from MODIS and the fraction of diffuse skylight over spatially-characterized landscapes

Miguel O. Román, Crystal B. Schaaf, Philip Lewis, Feng Gao, Gail P. Anderson, Jeffrey L. Privette, Alan H. Strahler, Curtis E. Woodcock, Michael Barnsley

Choose an option to locate/access this article:

Get Full Text Elsewhere

DOI Search Results

Citation

Data Set
The following data set has been cited or referred to in this article.

SAFARI 2000 Surface Albedo and Radiation Fluxes at Mongu and Skukuza, 2000-2002

ORNL Distributed Active Archive Center
Biogeochemical dynamics data
Presentation Outline

- Metrics of Impact
- Citing Data
- MODIS Data Citations
- Challenges
MODIS Data Citations

• In Collection 6, each MODIS product will have a registered DOI

<table>
<thead>
<tr>
<th>MODIS Product</th>
<th>DOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Photosynthesis</td>
<td>10.5067/MODIS/MOD17A2.006</td>
</tr>
<tr>
<td>Annual NPP</td>
<td>10.5067/MODIS/MOD17A3.006</td>
</tr>
</tbody>
</table>

• DOI embedded in each HDF-EOS tile
• N.B. Citations have not been established
Possible example of citation:
Based on NSIDC Citation Policy

_DRAFT for general discussion_


- MODIS Data Products
  - Need a policy for use of DOI and citation
  - Need to identify authors for citations
Presentation Outline

- Metrics of Impact
- Citing Data
- MODIS Data Citations
- Challenges
Challenges

• Ensuring consistent use of data citations
  – By authors, editors, and journals

• Tracking data usage, citations, and citation index
  – Publishing groups need to include data centers and
data products in their compilations
Web Resources

• Data Citation Background
  https://www.force11.org/datacitation
  http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations

• ORNL DAAC Search by article DOI
  http://daac.ornl.gov/doi_search_page.shtml

• Data Citation Policies
  http://nsidc.org/about/use_copyright.html
  http://daac.ornl.gov/citation_policy.html
  https://lpdaac.usgs.gov/about/citing_lp_daac_and_data
  http://podaac.jpl.nasa.gov/CitingPODAAC