

Long-term Plan for the VIP ESDR at the LP DAAC



Dave Meyer, USGS

LP DAAC Project Scientist

Stacie Doman Bennett, SGT, Inc.

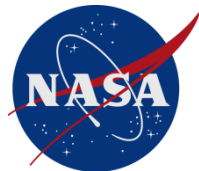
LP DAAC Scientist

Data Preservation - Goals

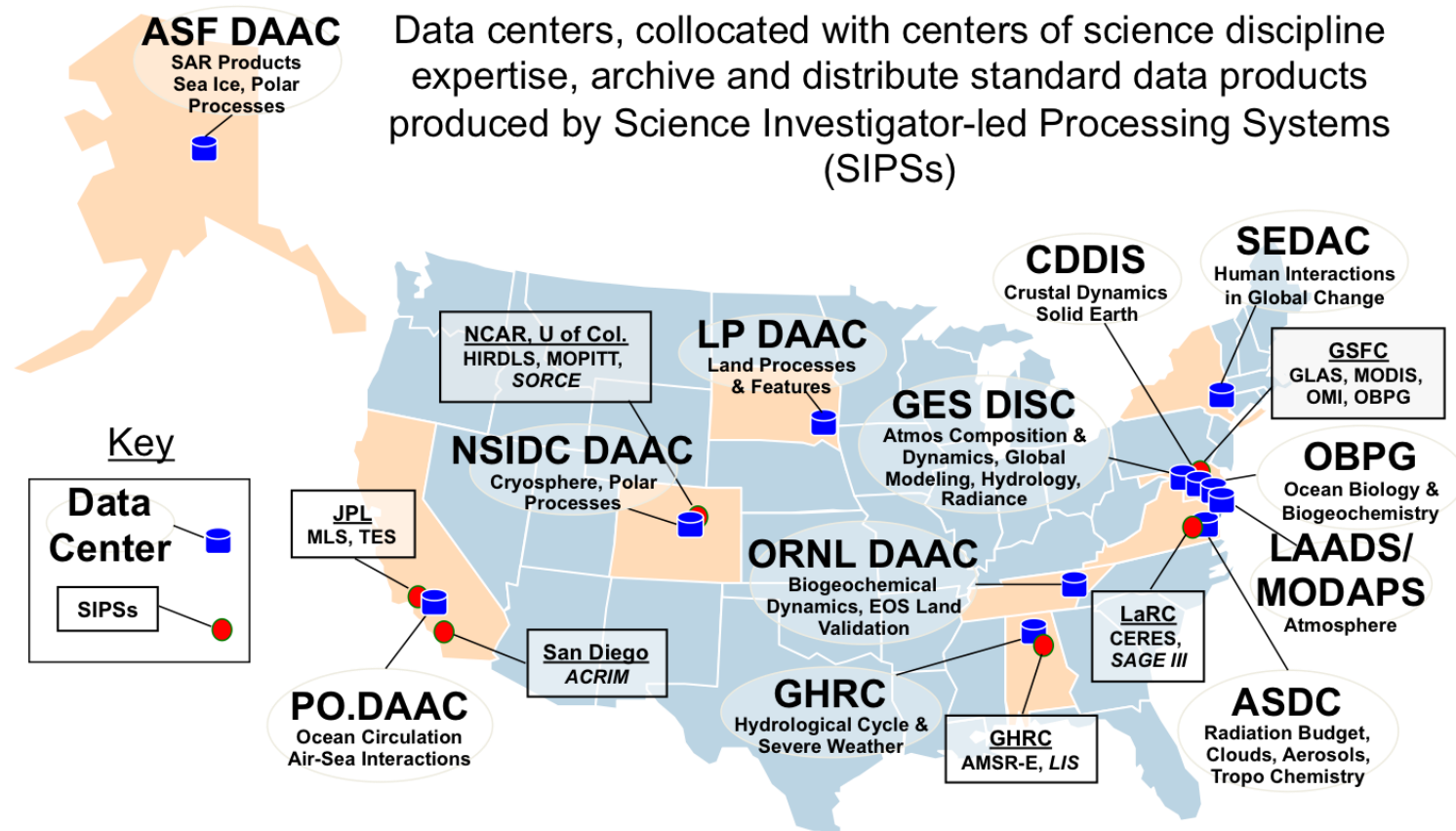
- Preserve NASA's Earth Science data for future generations
- Three aspects of preservation
 - *Maintaining bits with no loss as they move across systems and media, as well as over time*
 - *Ensuring readability over time*
 - *Providing for long-term understandability/reproducibility*
- While NASA is not a “permanent archive” agency,
 - It maintains a “research archive” for as long as data are used for scientific research or until responsibility is transitioned to permanent archives
 - USGS is a permanent archive, MEaSUREs preservation TBD

Data Preservation – Goals

- Ensure readability over time
 - Maintain format-dependent read software tools -OR-
 - Eliminate dependence on specialized software libraries
 - Develop machine- and human-understandable documentation of internal details of file structures to enable future users to write read software
- Providing for long-term understandability/reproducibility
 - Maintain documentation and ancillary data associated with products.
 - Reproducibility – not originally “in scope” for 2006 MEaSUREs, but transfer of relevant S/W and ancillary information should be discussed.
 - Other – what about specialized clients, such as the VIP Data Explorer?



NASA Earth Data and Information System (ESDIS) Project



ESDIS “Community” Solicitations

- Research, Education and Applications Solutions Network (REaSON)
 - Forty-two projects awarded in 2003-2004
 - Integration of data products, information systems and services
 - e.g., TOPS, GLCF, SERVIR
- Advancing Collaborative Connections for Earth System Science (ACCESS)
 - Enhance/improve existing distributed NASA-funded Science Info Systems
 - Solicitations most years (projects TRL > 5)
 - e.g, GRACE, HDF/OpenDAP, NACP, Sensor Web
- Making Earth Science Data Records for Use in Research Environments (MEaSUREs)
 - Focus on Earth Science Data Records (consistent science records over multiple missions)
 - Twenty-nine projects funded in 2007.

MEaSURES ESDRs at the LP DAAC

- 2006 MEaSURES Awards (to be distributed by the LP DAAC in FY13)

PI	Institution	Title
Kamel Didan	U. Arizona	Vegetation Phenology and Enhanced Vegetation Index Products from Multiple Long-Term Satellite Data Records
David Roy	SD State U.	Web-enabled Consistent Large Area Landsat Data Streams and Derived Surface characterizations MODIS-Landsat Data Fusion for the Terrestrial User community
John Townsend	U. Maryland	Earth Science Data Records of Global Forest Cover Change
Mike Kobrick	JPL	The Definitive Merged Global Digital Topographic Data Set

- 2012 MEaSURES Awards

PI	Institution	Title
Sean Buckley	JPL	NASADEM: Creating a New NASA Digital Elevation Model and Associated Products
Matthew Hansen	U. Maryland	Vegetation Continuous Fields ESDR for the AVHRR and MODIS Records: 1981 - Present
David Roy	SD State U.	Global Long-Term Multi-Sensor Web-Enabled Landsat Data Record
Prasad Thenkabail	USGS	Global Cropland Area Database (GCAD30) Through Landsat and MODIS Data Fusion for the Years 2010 and 1990 and its Dynamics over Four Decades Using AVHRR and MODIS

USGS-EOS Role: Distribution

- Assist PI with documentation, metrics, outreach, workshops
- Reviews – community review from ORNL/LP DAAC User Working Groups
- Visual discovery interface
 - Deferred as redundant with the U. Az VIP Data Explorer, focus switched to developing web services after transfer.
- Data storage
 - As part of migration, V2/V3 moving to LP DAAC distribution system

VIP V3 Activities

- Ensure LP DAAC storage requirements for V2 and V3
- Set up LP DAAC HTTP site for download access to VIP V3
- Transfer VIP V3 collection from UAZ to LP DAAC (date pending)
- Update LP DAAC Web site with VIP V3 product page
- Public release of V3
 - including news item on LP DAAC home page
- Prepare LP DAAC User Services for VIP user support
- V4?

Core vs. Community products

CORE (“Standard” products)	COMMUNITY (such as MEasUREMENTS)
Projects Subject to Programmatic Review	Projects Competitively Selected
Substantive NASA Oversight	‘Light Touch’ Oversight w/ significant Community Involvement
Tight Integration of Data System Tools, Services and Functions	Community-based Tools and Services Loosely-Coupled
Employ Well Established Information Technologies	Employ ‘Edgy’ or Emerging Technologies

MEaSURES VIP User Services

Data Center Service	Level 1	Level 2	Level 3	Level 4	MEaSURES Level
User Services	None	Referral	General	Custom	2/3
Ingest	None	Ingest to archive	Ingest with QC	Custom	2 (3?)
Processing	None	Simple monitor	Simple Monitor	Custom	3
Distribution	None	Monitor	Manual	Custom	4
Archiving	None	Unsupported to minimal	Metadata only	Primary archive	1
Production software	None	n/a	Follow internal s/w standards	Custom	1
Documentation	None	Metadata only	Metadata & provider	Comprehensive	3

Publishable Persistent Data Identifiers

- Want unique and lasting **data** identifiers for publication
 - More frequent and consistent citation of EOSDIS datasets
 - To find the NASA data used in research
 - Regardless of where it is moved to or who becomes responsible for it
 - *The LTA corollary*: To find the documentation for archived data
 - To enable metrics collection on cited datasets
- Digital Object Identifiers (DOIs) have emerged as the most accepted data identifier in the publishing community.

Digital Object Identifier system

- The DOI® system and the Handle System: Internet infrastructure
 - Internet resolution service for unique and persistent identifiers of digital objects
 - Owned by **International DOI Foundation (IDF)**– www.doi.org . The system was developed from the publishing industry (circa 2000).
- Consists of two part alphanumeric string doi:[prefix]/[suffix]
 - E.g., 10.5067/123; Prefix 10 is the DOI registry identifier; 5067 is the Registrant (ESDIS)
 - Suffix alphanumeric string identifies the data item as decided by the Registrant
- Citation and Location information is maintain at the DOI registry by an IDF Registrant Agent (RA) through a subscription provider (EZID)
 - Citation and location information can be updated as frequently as desired by the Registrant subscription holder
 - Desire is for one DOI per data item; but registry does not preclude multiple registrations/publishers of 'similar' data
 - Can migrate existing DOIs to new or different Registrant Agent and/or owner at any time

Examples of Data DOIs in Citations

- ESIP citation provider guideline using doi:
 - Doe, J. and R. Roe. 2001. *The FOO Data Set. Version 2.3*. The FOO Data Center. <http://dx.doi.org/10.xxxx/notfoo.547983>. Accessed 1 May 2011.
- Hypothetical example using doi with GES DISC citation:
 - Chung-Lin Shie, Long Chiu, Robert Adler, I-I Lin, Eric J. Nelkin, and Joe Ardizzone, 2010. *Surface Turbulent Fluxes, 1x1 deg Monthly Grid, Set1 and Set2*. Edited by A. Savtchenko. Greenbelt, MD: Goddard Earth Sciences Data and Information Services Center, Accessed <date>. doi: 10.5067/MEaSURES/OceanFluxes/data1.
- Example using doi with NSIDC citation:
 - Jackson, T. J., and M. H. Cosh, 2003: SMEX02 watershed soil moisture data, Walnut Creek, Iowa. National Snow and Ice Data Center, Boulder, CO, digital media. doi: 10.5067/SoilMoisture/SMEX02/data1.
- Example using doi for station data
 - König-Langlo, Gert and Hatwig Gernandt. 2006. *Compilation of radiosonde data from the Antarctic Georg-Forster station of the German Democratic Republic from 1985 to 1992*. Bremerhaven, Germany: Alfred Wegener Institute for Polar and Marine Research. Data set accessed 2008-05-22. [doi:10.1594/PANGAEA.547983](http://dx.doi.org/10.1594/PANGAEA.547983).

Implementing DOIs for EOSDIS

- Select data providers for pilot that are doing final reprocessing
 - HIRDLS, GLAS, AMSR-E, and **selected MEaSUREs projects**
 - Develop guidelines for DOI suffix, citation & location information.
 - Want DOIs to be attractive to users, solicit feedback from PI's and DAAC UWG members
- Add DOIs to DAAC product citation web (“landing”) pages
 - Imbed DOIs into product metadata at next reprocessing or ingest
- Add DOIs to Global Change Master Directory and ECHO through metadata updates
- Add DOI metadata to NTRS for searchable documentation
- Setup metrics collection from journal citation reports

Pilot DOI Examples

Suffix Model String	Example
[mission]/[instrument]/data[1-n]	doi: 10.5067/Aura/HIRDLS/data1234 doi: 10.5067/ICESat/GLAS/data1234 doi: 10.5067/Aqua/AMSR-E/data1234
[campaign]/[measurement group]/data[1-n] [campaign]/[platform group]/data[1-n]	doi: 10.5067/BOREAS/Airborne/data1234
[program]/[measurement group]/data[1-n] [measurement group]/[data[1-n]	doi: 10. 5067/MEaSURES/OceanFluxes/data1234 Doi: 10:5067/MEaSURES/SnowExtent/data1234

Proposed for VIP:

doi: 10.5067/MEaSURES/VegIndexPhenology/data1234

DOI “landing pages” MODIS example

The screenshot shows the USGS LP DAAC website. The header includes the USGS logo, NASA LP DAAC logo, and navigation links like HOME, ABOUT, PRODUCTS, GET DATA, TOOLS, USER COMMUNITY, and CUSTOMER SERVICE. The main content area is titled "Vegetation Indices Monthly L3 Global 0.05Deg CMG" and "MYD13C2". It contains a description of the MODIS vegetation indices and a world map showing the data distribution. The left sidebar lists various data products and links.

► Overview

► Layers

▼ Links

[User Guide \(University of Arizona Web Site\)](#)

[Algorithm Theoretical Basis Document \(PDF\)](#)

[Accuracy \(MODIS Land Validation Web Site\)](#)

[Basic Info](#)

[Policies](#)

[Get Data](#)

[Help](#)

Data Access Tools

[Data Pool](#): The Data Pool (On-line Archive) provides access to all MODIS products, and a rolling archive of ASTER level-1B products that cover the U.S. and its Territories.

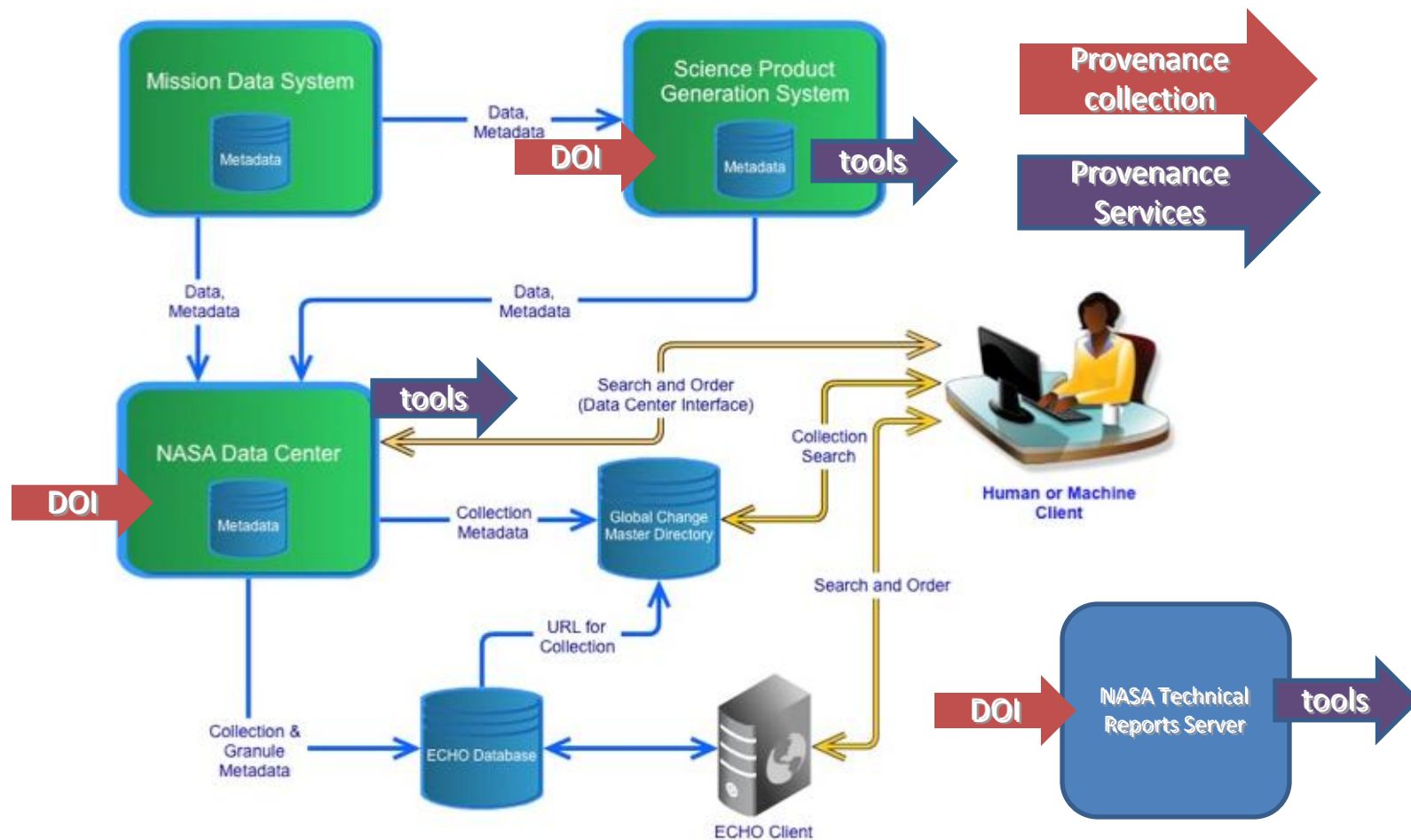
[Reverb](#): This tool provides access to a complete data record of all MODIS and ASTER products available from the LP DAAC.

The [Get Data](#) Section of this Web site also provides access to these interfaces.

LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER



Implementation in Interoperable Architectures



Metadata flows in NASA Earth Science Data Systems