

Virtual Update Notes July 6 & 7, 2009 - Network Information System (NIS) and Data Access Server (DAS)



Published on *LTER Information Management* (<http://im.lternet.edu>)

Home > Virtual Update Notes July 6 & 7, 2009 - Network Information System (NIS) and Data Access Server (DAS)

Virtual Update Notes July 6 & 7, 2009 - Network Information System (NIS) and Data Access Server (DAS)

Fri, 06/26/2009 - 3:30pm — sremillard

Network Information System (NIS) and Data Access Server (DAS)

See presentation link below.

Participants:

Monday (7/6): Mark Servilla (moderator, LNO), Don Henshaw (AND), Duane Costa (LNO), Gastil-Buhl (MCR), Jason Downing (BNZ), Jonathan Walsh (BES), Margaret O'Brien (SBC), Suzanne Remillard (AND), Barbara Benson (NTL), Dave Balsiger (NTL), Kristin Vanderbilt (SEV), John Porter (VCR), Emery Boose (HFR), James Brunt (LNO)

Tuesday(7/7): Mark Servilla (moderator, LNO), Karen Baker (PAL, CCE), Nicole Kaplan (SGS)

Monday Notes:
by Emery Boose
(see John Porter notes attached below)

See power point on IM website for more details.

Two topics: NIS (network information system) and DAS (data access server)

Network Information System

LNO supplemental grant will get NIS development back on track. Previously only core LNO activities were funded (not including NIS). This funding is economic stimulus money.

New funding will make it possible to buy back some of Duane's time from other projects (starting Aug. 2009). Duane will also continue work on Dryad through 2010 (connected to Metacat development).

Four new hires at LNO: NIS analyst / programmer (entry level), technical intern, web development intern, information manager. New projects (EcoTrends, Pasta, etc) are generating need for data management at LNO.

Information manager position expected to be permanent. Stimulus funding is for 5 years (through 2013); will need to seek funding to continue other positions.

Pasta re-kindled. See updated diagram in power point. Value-added applications include LTER data catalog, EcoTrends data portal, etc.

Starting to plan and examine resources. See timeline in power point. Includes new hires and fractional times for each person. First couple of months are project planning. Timeline will change as project matures. Leverage new resources as much as possible over next few years.

See description of each module in power point: description, effort, resources, dependencies, duration.

Metacat. Set up Metacat so it can run independently.

Data Registry. Use EML documents to register datasets in Pasta framework. Test and validate data files against EML documents.

Parser-Loader. Use Data Management Library (information in data registry) to add new data.

Data Cache. Relational database that stores datasets uploaded from sites. Use predetermined method to cache datasets. Pull from sites into Pasta framework as needed by application.

Workflow Engine. Workflow manager for different types of workflows (Kepler, R, etc).

Derived Data. Another relational database to store derived products. Accessible to other applications. Not necessarily different from data in Data Cache.

Data Access Server. Proxy server provides auditing and notification capabilities.

EML Factory. Automate generation of metadata for derived data products. Include provenance metadata.

Extracat. Provides fast search capability using external search index. Currently used for EcoTrends web portal. Improved performance in many cases.

Access Interface. Consistent, straightforward, standard approach. Possibly via REST.

Data Catalog. Decouple from Metacat, simplify some current functions. Include new LTER controlled vocabulary with auto-completion. Optimal data structures. Initially for LTER network but extensible for other users.

Current vision is to incorporate site data into Pasta system. Persistent archive in case sites go offline. Similar to ClimDB and HydroDB but without extensive preparation (units, scaling, etc). Sites can continue to create datasets in the way that makes most sense locally. Add ability to exclude data as appropriate.

Goal for end 2010 / beginning 2011 is working prototype of Pasta system.

ChemDB is a good example with diverse methods and datasets. A science effort may be required to identify valid derived products. Sites may have to do some of the data preparation. Problem of maintaining large numbers of R scripts, etc. But Pasta would also be a good way to make derived products available.

LTER-centric in the sense that high-quality EML is required.

Portal. Extensibility to current data catalog. Sites can reuse Java application for other uses.

Testing, Validation, and Operations. Testing of individual modules by developers. Beta testers of entire system. Data itself will need to be managed by information manager.

Total personnel requirement is still under study. Current timeline is through 2011.

NISAC has been involved over last three years since initial conception. These latest effort estimates being compiled for the development of an implementation plan have not been seen by NISAC but will be on the next meeting agenda. LNO hopes to create a solid design document for next meeting with NISAC.

Data Access Server

Sites vary in their mechanisms for tracking data downloads. DAS provides a proxy server.

Replace data URL with proxy URL to require user registration. The audit database will be accessible to site IM. Email notification to IM and dataset contact. Email to end user identifying dataset and providing link to metadata.

Integrate DAS into Pasta framework itself.

ARC and GCE are early adopters. DAS provides more complete information on users. See project planning document available online.

System is ready to accept new sites. But still a Beta application. Test functionality for each new site. Use IM website / admin function to register data URLs and notification information.

Plan to add monthly reporting function.

Ask each site to register a couple of datasets before ASM? Sites can register selected URLs (not all or nothing).

Current form is relatively simple to fill out. Information is harvested from the EML. Not necessary to add information for each dataset.

See power point slide 42 (DAS administration) for registration information.

Data URL needs to be updated. Root component (head) is replaced with DAS address. Specific component (tail) remains the same.

Tuesday Notes:

Attachment	Size
NIS-DAS-Presentation for VWC [1]	2.02 MB
John Porter notes from Monday, July 6th VWC [2]	63.37 KB

- Virtual Updates [3]

- Copyright © 2012 Long Term Ecological Research Network, Albuquerque, NM - This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Please contact us with questions, comments, or for technical assistance regarding this web site.

Source URL: <http://im.lternet.edu/node/471>

Links:

[1] <http://im.lternet.edu/sites/im.lternet.edu/files/NIS-DAS-WaterCooler-20090706-07.pdf>

[2] http://im.lternet.edu/sites/im.lternet.edu/files/IM_VTC_2009-07-06.pdf

[3] <http://im.lternet.edu/taxonomy/term/169>