February 2013: DEIMS working group update

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

February 2013: DEIMS working group update

Mon, 10/22/2012 - 9:54am — kvanderbilt

Agenda – DEIMS Video Conference Call

Objective:A Q&A session preceded with What is DEIMS, Why, Brief status, Roadmap

Date 1/4/2013 Time From: 1 p.m To: 1.45 p.m
Date 1/5/2013 Time From: 10 a.m To: 11.00 p.m

Summary Points

What is DEIMS 2 mins DEIMS Elevator Speech. Using a popular, free and flexible content management system to connect the data to knowledge

Brief Background2008 - Started with two early adopters (SEV, LUQ) and two implementers. 2009 – ASM working group, official launch on google code. ARC, PIE and NTL join in. 1st Training pool money for a data catalog and EML module. 2010, small 2nd training. 2011 Training and pooled resources for robust launch of next gen. Many more participants.

Current StatusRFP launched at the UNM on Nov 2011. In negotiations with top of the industry Drupal contractor. UNM lawyers and procurement.

RFP Outline The Scope of work delineates an ambitious plan. Links to the actual document will be available when we have a contract. Contract to be renegotiated after initial discovery and definition. Complications with multi-site pooled resources. See appended points of the original Scope of Work -

Beyond 6 months Future plans : Pursue expansion of the tool, user base and further development. Use Drupal Association, NSF and other private foundations to advance goals.

Q & A 15 min

Supplemental material –

Scope of Work for next gen DEIMS as published in the RFP – Nov 2011

Asses the overall approach to our goal. Is the DEIMS structure optimal? What changes can be done that optimize, simplify and improve both the performance (bottlenecks) and usability, especially considering the next DEIMS implementation (say DEIMS 2.0) should be done in D7. Describe the best organization of content types, node references, and other implementation details for our data entry application. Identify current problems and design solutions.

As stated in the introductory sections, DEIMS contains a custom module (views_bonus_eml, aka Drupal2EML) that takes information from the interconnected nodes casted in 6 content types and generates an XML stream that complies with the Ecological Metadata Language schema according to this specification: http://knb.ecoinformatics.org/software/eml/eml-2.1.1/index.html. This D6 custom module is inspired on the views_bonus_export contrib. The Drupal2EML module or functionality needs to be migrated to the new DEIMS structure in D7. The contractor will modify the module as necessary for D7, taking advantage of new modules in D7 that make it easier to use and maintain. The contractor will ensure that the Drupal2EML module produces EML that meets the EML XML schema specifications and certain extra requirements that are required by the LTER network data synthesis services.

The workflow for documenting our ecological data is inefficient to use. Specifically, the data file structure and variable details can use improvements. We would like to see a "wizard" like approach in D7 for documenting variables and importing datasets, in essence, usability improvements. That is, we want to be able to upload the data file, and then document each variable while looking at a view of the data. For example, the contrib. module node import parses out CSV files and stores the information in a content type. The module uses an 8-step import wizard to guide the data entry person. We can use something similar to enter data and or metadata. The overall goal is to be document the data entity (data container) and be able to produce structured metadata in a manner that is as simple as possible.

Due to the heterogeneous nature of the research (about 200 projects per LTER site, several dozens of investigators, multiple research locations, 30+ year history),

many datasets use similar variables, sometimes with different names or slight changes in format. When we imported the datasets into DEIMS, we ended up with similar variables that described in essence the same measurable. For instance, the variable "date" should be reused over and over again. DEIMS now will show a few instances of variables with similar names as one selects the variable from the node reference autocomplete widget (pop-up akin to a dropdown list), but such list does not show all of the possible "date" like variables (See Fig. 3). We need to reduce the redundancy (perhaps at the source of the data), yet make it more usable to see all name matches in the autocomplete widget.

The contractor will create a sophisticated data catalog query page. An example of a catalog query page, which allows the user to search the data holdings at the Sevilleta LTER, can be found at: http://sev.lternet.edu/data. The new query page needs to contain a field for querying all text, and also needs to allow users to 1) select a core research theme, and 2) have the remaining dropdown lists narrowed to selections only related to that theme. In addition, the contractor must include temporal and spatial query elements in the catalog query page. We expect the contractor to have recommendations about how to structure this page.

We need to provide the ability to download a complete metadata records in a flat text file format. This may be included in the data download, or make it separate, along the data download function. This feature is similar to the ability of exporting metadata records in an XML compliant format – the difference is that the format would be plain text.

The contractor will integrate external scientific data to DEIMS - expose these entities to the contrib. module Views. For instance, Sevilleta LTER maintains a climate database in MySQL. Sevilleta LTER now has a simple PHP page (outside Drupal) that allows users to query that climate database and return data for specific timeframes and specific sites. The user can download the data in an Excel file. We want this same functionality to be built in to Drupal 7 or 8. That is, we want to be able easily build query pages on views 3 that access data external to Drupal. In addition, we would like to be able to query not only databases, but large text files that are generated from streaming sensor data. As much as possible, we would like to integrate (refer, relate) these external entities to the Drupal nodes and taxonomies. These entities may be database tables or CSV files. Possible solutions would be to use and extend the contrib module data or table wizard (tw) or entities.

The contractor will migrate all metadata from the first version of DEIMS in Drupal 6 to the new version of DEIMS in Drupal 7 (or 8) for the 3 LTER sites (Sevilleta, Luquillo and North Temperate Lakes) who have deployed DEIMS in production systems. The finished product will include three LTER websites serving content packaged in EML which will be harvested successfully in to LTER Network metadata catalog (at http://metacat.lternet.edu).

The contractor will migrate all other content (publications, news, media, stories, blogs) on 3 LTER sites (SEV, LUQ, and NTL) from the first version of DEIMS in Drupal 6 to the new version of DEIMS in Drupal 7 (or 8) . Appearance, current themes and panels will be replaced by a new adaptive theme that captures the essence of the current deployments and improves (see next point, #9). The migration would also include the dependent contributed modules and extensions to the core DEIMS content types if applicable. In summary, the contractor will ensure that each of the mentioned 3 sites has functioning and ported systems in the DEIMS incarnation.

Development and Deployments strategies – pertaining to new DEIMS installs and maintenance of deployed instances.

1. The contractor will package DEIMS for easy download and installation. An installation profile of Drupal for DEIMS may suffice. This will include a common adaptive theme for the website. Such theme serves the main content for mobile platforms as well as large screens. For example, an Omega subtheme or similar ready-to-deploy (Zen, AdaptiveTheme) adaptive themes are suitable solutions.
2. The contractor will help with the documentation. Installation and the overview of how DEIMS works (wizards and data entry workflows, query catalogs, import and export utilities, theme customization, as well as pointers to all other contrib. and core utilities)
3. A workflow for rapid deployment and maintenance is also desirable. For example, using Drupal features, or solutions that make the process of updating as efficient as possible. Specific replicates can be deployed at sites, but a central core would ensure the stability of the code system. Periodic updates (security, upgrades, core features) will be pushed to the sites. Content would be synchronized with the core instance, but details to be discussed further.

Basic species taxa information. One of the elements of the Ecological Metadata Language is a taxonomic species component. Currently DEIMS does not yet include any taxonomic-specific information. Cautionary note, in this context "taxonomy" is not a Drupal vocabulary. In this context taxonomy is a Scientific Species (live beings) Classification tool. For usability considerations, an easy to use field to tag species related to a study (or dataset), the contractor will either consume the web services of the Integrated Taxonomic Information System (http://www.itis.gov/) and devise a procedure for storing taxa-related content. A hierarchical Drupal taxonomy may accommodate this. At the very minimum, the contractor should create a simple vocabulary to tag datasets with species or common names, and such tags must be included in the EML structure on export. This service may suggest tags based on user input (Entered Tag: Wolf – pop-up or autocomplete input would show: is "Canis Lupus" the appropriate taxa? – see http://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=180596).

Some Drupal based efforts have surpassed these minimal requirements to deal with taxa matters. For example, Scratchpads and the Encyclopedia of Life's

LifeDesks . Integrating Scratchpad/LifeDesks services would be both a convenient start point and a value-added part for the site.

Consistently tagging content with relevant keywords results in a better content discovery experience. LTER is making in-roads in building a common vocabulary for content for all the network of sites. The contractor will accomplish two related tasks:

1. Create a Drupal 7 (or 8) module to consume web services for the LTER Controlled Vocabulary, enhancing the resulting EML with suggested terms.
2. Create a feed that periodically synchronizes the LTER controlled vocabulary via web services, updating a mirrored local Drupal taxonomy for the LTER controlled vocab.

Scientific units are standards used to compare scientific observables. LTER uses a repository of known units (http://unit.lternet.edu/unitregistry/). The DEIMS content type variable has an optional CCK field named "unit" where users can specify which unit the current variable node uses. The contractor will develop a module that suggests to the user a unit when he/she enters a unit in the variable "unit" field. The suggestion(s) will be based on the existing LTER units based on matches to the LTER unit repository DEIMS custom module. As an added note, currently, the custom module "drupal2eml" (views_bonus_eml) consumes the RESTful services of the LTER units module to generate and XML tidbit that is included in the EML streams generated by DEIMS. For an example of the XML served, see this query for "Liters Per Square Meter" http://unit.lternet.edu/services/unitformat/stmml/unit/name~litersPerSquareMeter

The contractor will create, port or adapt a bundle of feeds to import content straight from existing XML conformant to the EML schema. We tested such task using feeds and feeds xpath parser with some success.

EML is one of the specifications (standards) used by the broader ecological scientific communities for exchanging information about data sets. However, there are other standards, including the ISO19115 and its XML implementation ISO19139, as well as the Biological Data Profile, a profile of the Federal Geographic Data Committee for biological datasets. The contractor will create modules to export the content from DEIMS into XML compliant with these specifications.

Contractor will develop a Drupal module to consume services exposed by the PASTA API. PASTA is an architecture designed to ingest data based on EML metadata packages. The API has at least a function that evaluates the feasibility of ingesting data based on the produced EML. We would like to develop a module that informs of the "readiness" (Yes/No) function for the prepared EMLs.

The contractor will implement a solution such that manages media such as files, photos, slides and photo galleries in a unified manner consistent with the rest of the content.

Select Questions and Comments during the VTC

- Q: Timeline of work seems too fast - how are you going to cope with the rapid developments?
  A: We need a solid "discovery and definition Phase" where expectations are clearly defined. The "Agile" software development [1] will help keep everybody on track, as development will be broken in 2 weeks sprints, with attention to the goals of each sprint. The LTER DEIMS team will have unfettered access to the project management reports and site, and will integrate with the developers team as we define in the first stages of the project.
- Q: About hiring professional services – the RFP process,etc
  A: Complicated and inefficient. In this line of work, the RFP process does not aid much, yet, it is a constraint that is almost unavoidable. The Contractors are unhappy about it (See http://denver2012.drupal.org/program/sessions/no-rfps-why-requests-proposals-are-bad-business-and-how-we-can-stop-them) and the clients (us) are also unhappy, as we have to work extra to get the same service we would get as if we could hire directly some group. Ask Kristin for details.
- Q: Is the project mature enough to serve the needs of projects such as a macroecology group?

- Virtual Updates [2]

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/1112

Links:
[1] http://en.wikipedia.org/wiki/Agile_software_development
[2] http://im.lternet.edu/taxonomy/term/169