

Tobias T. Work¹, Cynthia L. Chandler², Andrew R. Maffei³, M. Dickson Allison¹, Robert C. Groman¹, Steve L. Gegg²
1~Biology Department, 2~Marine Chemistry and Geochemistry Department, 3~Computer and Info Services, Woods Hole Oceanographic Institution, Woods Hole MA 02543

Abstract

Content Management Systems (CMSs) provide powerful features that can be of use to oceanographic (and other geo-science) data managers. Strategic use of some Drupal6 CMS features enables three separate but complementary interfaces that provide access to oceanographic research metadata via the MySQL database: 1) a Drupal6-powered front-end; 2) a standard SQL port (used to provide a MapServer interface to the metadata and data; and 3) a SPARQL port (feeding a new faceted search capability being developed). The Biological and Chemical Oceanography Data Management Office (BCO-DMO) currently uses a MySQL backend and a ColdFusion front-end but the intent is to replace ColdFusion with Drupal 6.

In late 2006, the U.S. National Science Foundation (NSF) funded BCO-DMO at Woods Hole Oceanographic Institution to work closely with investigators to manage data generated from their research projects and to ensure those data are permanently archived at the U.S. National Oceanographic Data Center (NODC). In the past, BCO-DMO has submitted data to NODC as an email with attachments including a PDF (a manually completed metadata record) and one or more data files. This method is no longer feasible given the rate at which datasets are submitted to BCO-DMO. BCO-DMO has worked with NODC to design a more streamlined and automated system to keep up with the data that needs to be archived at NODC.

Motivations for Using Drupal

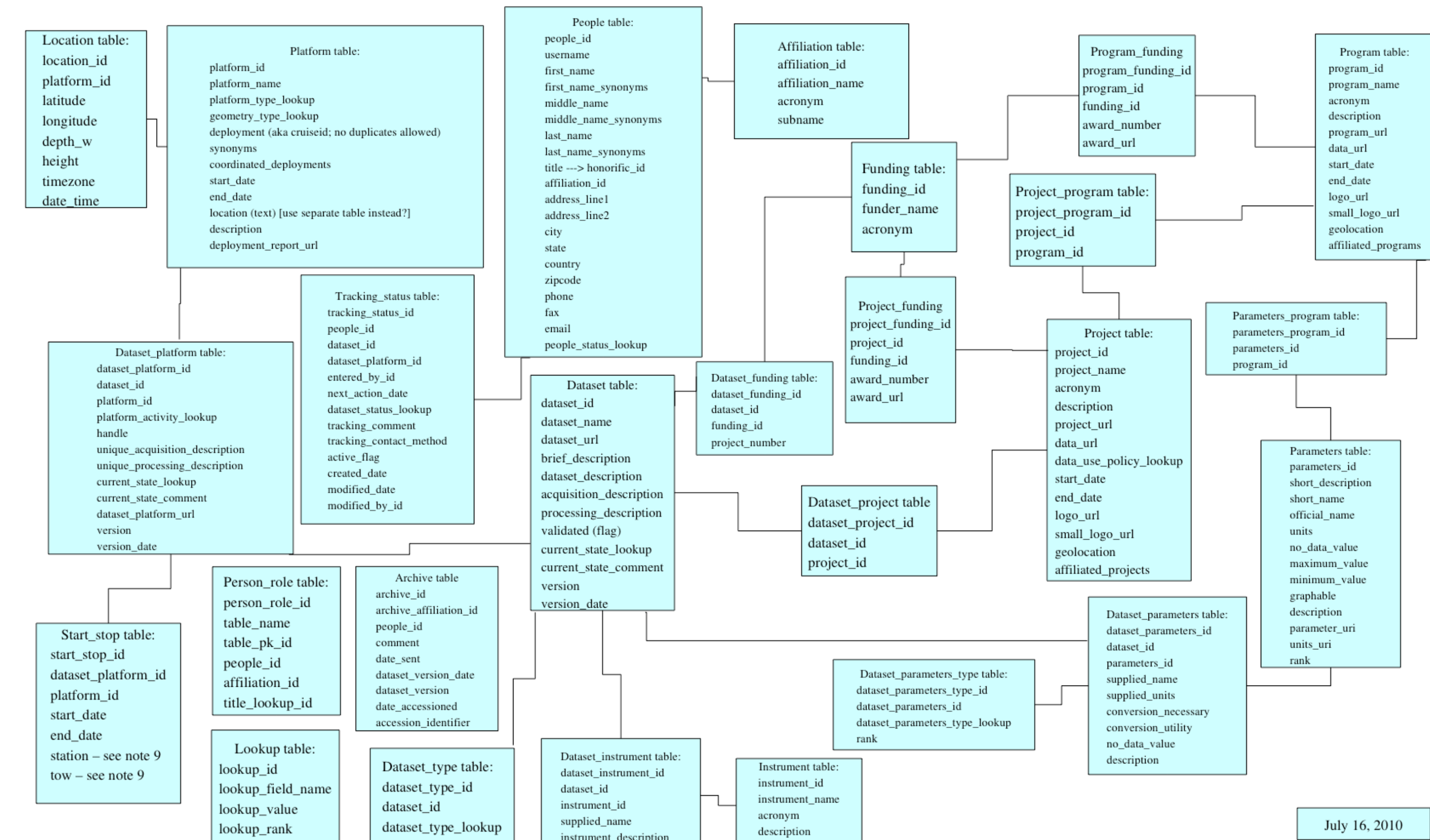
ColdFusion 5	Drupal 6/7
Proprietary	Open-source
Custom code	Modules
No Semantic Web	Semantic web support in Drupal 7

Highlighted above are just three of the advantages that BCO-DMO would gain from using Drupal instead of ColdFusion. The first of which is the fact that Drupal is Open Source. This enables BCO-DMO to have much more control over the software that is serving our data to the web. The second advantage is that because Drupal is Open Source and built with a modular nature in mind, BCO-DMO can take advantage of code written by the Drupal community and may even write custom code if it is necessary. There are literally thousands of Drupal modules, many of which would be useful to BCO-DMO. The third advantage is that Drupal 7 *looks* as if it will have semantic web capabilities built into the core. This is a major plus as we see semantic technologies as playing a very important role in the development data access, data storage and data interoperability.

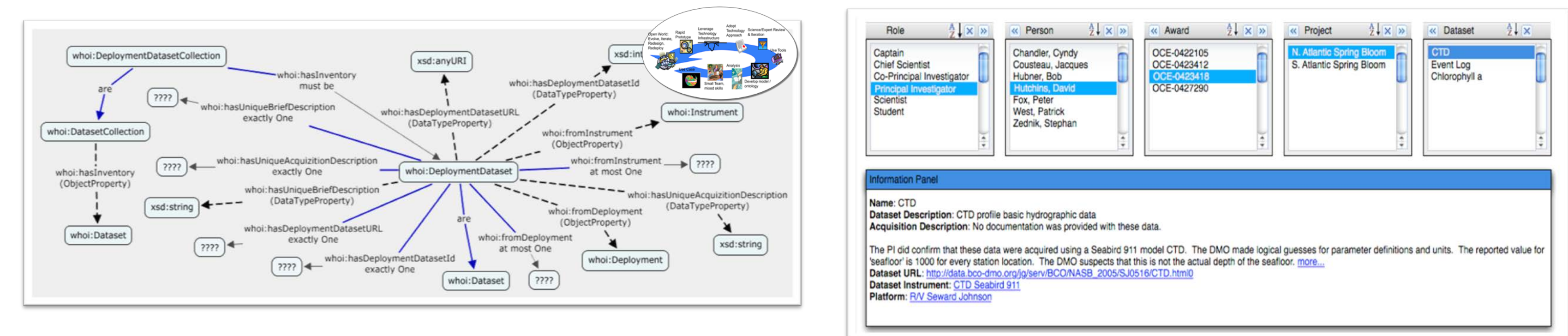
Motivations for Automated Data Submission

The motivation for a more automated data submission system is straightforward: the old way (filling out a PDF by hand and sending an email) is simply not fast enough given the volume of data that BCO-DMO is required to manage and therefore submit to NODC. The new way involves minimal interaction on behalf of the data manager and leaves more time for other data manager responsibilities.

Migrating to Drupal

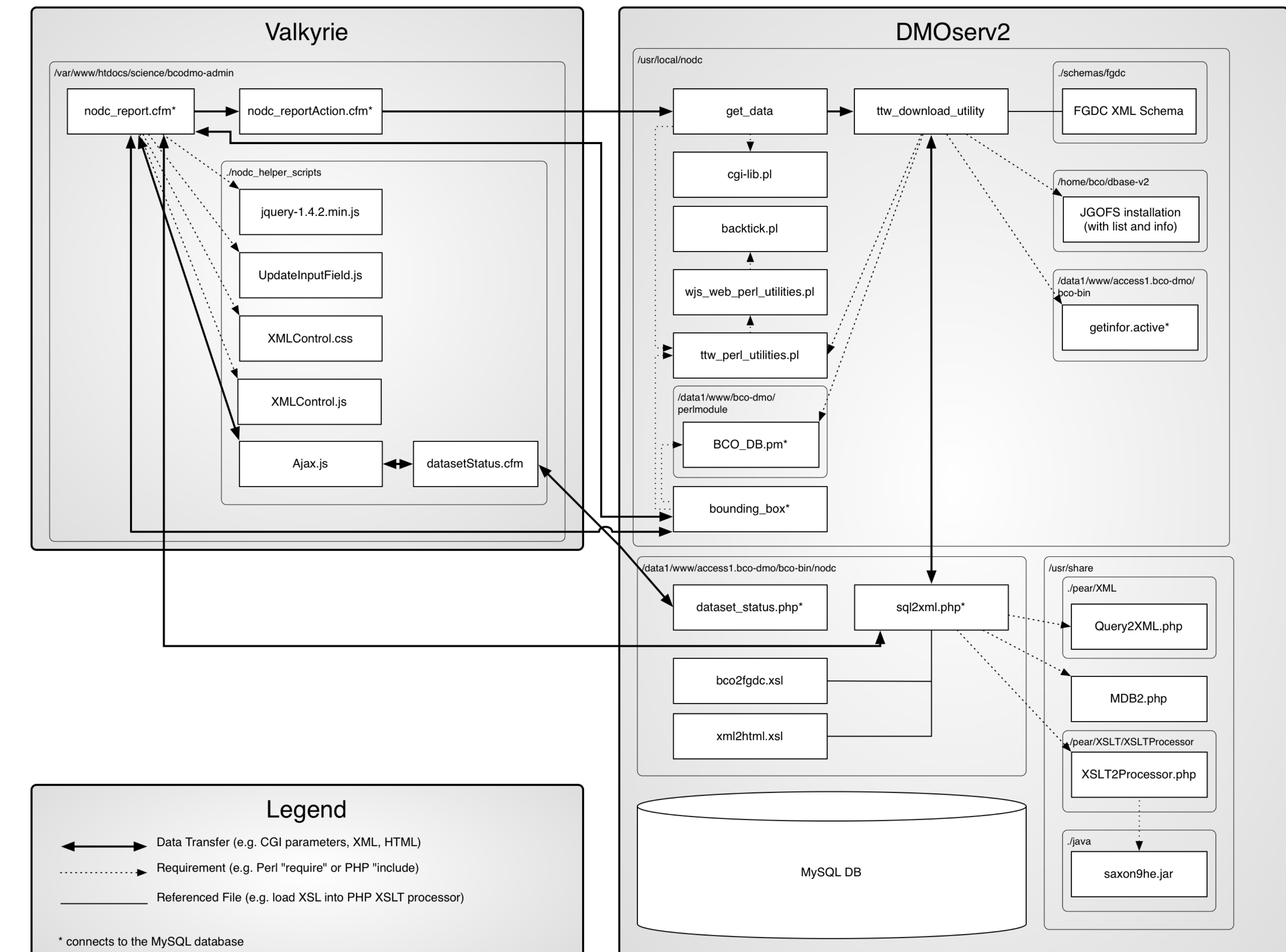


Above is a representation of BCO-DMO's current MySQL metadata database schema. Each table in the above diagram will be represented in Drupal. Once the Drupal structure is set up, the information will be imported from the existing MySQL database and linked (ideally in one step). Once the data is represented accurately in Drupal, we can take advantage of the many modules that are available from the Drupal community. We may also develop our own modules to customize Drupal for our use should certain functionalities not be available “out of the box” or from existing modules.



We plan to map RDF vocabularies to content types and fields. We will create a new BCO-DMO ontology as well as map other fields to SEAVox, FOAF, VSTO, MMI Device, SWEET, ICES ship codes, LDEO/MGDS and other vocabularies to Drupal content types and fields. This figure shows CMAP representation of DeploymentDataset portion of the BCO-DMO ontology being developed by BCO-DMO team with the help of Peter Fox at the Tetherless World Constellation, Rensselaer Polytechnic Institute

Automating Data Submission



The automated data submission system allows the data manager to archive data and metadata at the National Oceanographic Data Center (NODC) or any other data center quickly and efficiently. For NODC, the metadata is in the form of a Federal Geographic Data Committee (FGDC) compliant XML file and the data is in the form of one or more ASCII comma-delimited files.

The diagram above is a representation of the structure of the automated data submission system. The large gray boxes titled “Valkyrie” and “DMOserv2” are representative of separate servers (design constraints were such that the system could not reside on one server). The smaller white boxes are individual files and the arrows connecting them indicate either data transfer or a software dependency (more detail is available in the legend). The system uses several languages to accomplish data submission including ColdFusion, Javascript, Perl, PHP and XSLT.

Acknowledgements

This work has been funded by the National Science Foundation, the Woods Hole Oceanographic Institution, and Rensselaer Polytechnic Institute. We are grateful for the contributions of Julie Allen (ColdFusion metadata web interface) for her applications programming work and Peter Fox et al. from the Rensselaer Polytechnic Institute for their help developing a BCO-DMO ontology.