Framework for Building Collaborative Research Environment

Ranjeet Devarakonda[1], Giriprakash Palanisamy[1] and Inigo San Gil[2]

[1] Oak Ridge National Laboratory
[2] University of New Mexico

Abstract—**Wide range of expertise and technologies are the key to solving some global problems. Semantic web technology can revolutionize the nature of how scientific knowledge is produced and shared. The semantic web is all about enabling machine-machine readability instead of a routine human-human interaction. Carefully structured data, as in machine readable data is the key to enabling these interactions. Drupal is an example of one such toolset that can render all the functionalities of Semantic Web technology right out of the box. Drupal's content management system automatically stores the data in a structured format enabling it to be machine. In this article we will discuss how Drupal promotes collaboration in a research setting such as Oak Ridge National Laboratory (ORNL) and Long Term Ecological Research Center (LTER) and how it is effectively using the Semantic Web in achieving this.**

Index Terms—**Drupal framework, Semantic Web, Collaboration in science research, Web frameworks for Collaboration, Oak Ridge National Laboratory, US DOE, NASA, SPRUCE, NGEE Arctic, WENDI, HIPPO, ESDORA.**

I. INTRODUCTION

As we all know collaboration at the department level or at the institutional level is becoming more and more common in the recent times. In a "Memorandum for the Heads of Executive Departments and Agencies" released by the White House on June 21 2010, the Obama Administration stresses on the importance of collaboration at the national and regional levels to link, leverage and align resources and also to make taxpayer dollars as cost-effective as possible [1].

A. What are factors contributing to Collaboration in science research?

Is data sharing a good thing?

Sharing data with colleagues the broader scientific community and public is highly desirable and will result in greater advancement of science. Below are some benefits of data sharing:

- Open science and new research
- Data longevity
- Data reusability
- Greater exposure to data
- Generation of value added products
- Verification of published works
- Possibility for future research collaborations
- More value for the research investment
- Increased citations on shared data related publications

Collaboration can vary in level ranging from very substantial to almost negligible. Factors promoting to collaboration range from basic necessity of someone insight to organizations budget cuts to the need of scientific recognition [2] or popularity [3]. Researchers at work at one physical location have the opportunity to see each other and collaborate by simply exchanging research material [4]. In other cases, researchers from different organizations may collaborate by exchanging or transferring material via the Web. There are many technologies as simple as Blogs and Wikis for creating online collaborative environments.

B. Problems of collaboration in scientific endeavors

Data sharing is easier when data is stored in a format that can be locatable, retrievable and understandable, more importantly, it should be in a form that will continue to be accessible as technology changes [5]. Scientific data, in its most general context across multiple disciplines, include measurements and observations of natural phenomena for the purpose of explaining the behavior of or testing hypotheses about the systems. Examples of such data include observational data captured in real-time by sensors, surveys, and imaging devices; data from laboratory instruments, for example, gene sequences, chromatograms, and characterization of samples; simulation data generated from models where the model is equally important with the input and output data such as for climate and economic models; and derived or compiled data that are the result of text and data mining, and compiled and integrated databases from multiple sources. Scientific data are generally rich and diverse and sharing this kind of information is challenging [6]. It will bring substantial benefits to scientists if an integrated collaboration environment can combine these data sources into a single easy-to-use, intuitive environment.

An article in 'ON Magazine' talks about how the web is revolutionizing the study of climate change [7]. Author comments that data collected for over 20 years is being put to good use by climate scientists who conducting collaborative studies with other side of the world, by holding webinars, and sharing data via the Web. In the recent years Web has grown in complexity that it is longer considered a medium for human interpretation and use. New technologies like Semantic web infrastructures are solving the problems by providing a framework for automatic web service interoperation [8] and facilitate mash up-like information sharing. Web 2.0 technologies outlines in "What is Web 2.0" [9] based websites like Wikipedia[1] and

---

1 http://wikipedia.org

HousingMaps[2] allow for an easier distributed collaboration.

There are several general challenges that present themselves when attempting to create a collaborative research setting, both social and technological in nature. Social challenges may include researchers who are reluctant to learn a new system or prefer to meet face to face. Similarly, technological challenges come in the form of operating system incompatibilities or researchers in settings with limited bandwidth. Any of these problems makes implementing an environment designed to facilitate the interactions of researchers more challenging.

Additionally, when attempting to address these challenges, a project leader often has to work under several constraints. A common constraint is the cost. Budgets for research projects often give short shrift to elements of the project that may be seen as ancillary and while viewing a collaborative environment in this vein may be short sighted, it is not un-common. Small budgets remove many of the commercial choices as options. Furthermore, smaller budgets also lead to fewer available technical resources that are not directly involved in the primary research goal. This constraint implies that it is not feasible to employ potential collaboration environments whose requirements and maintenance go beyond a certain complexity.

II. BACKGROUND

A. Drupal History

Drupal is an open source web-based content management system (CMS) [10]. Drupal is licensed under the GNU General Public License implying that its source code and derivatives are freely downloadable and customizable. In some cases this can also imply that it may not be a very stable application, but from its humble beginnings more than 10 years ago, it has undergone a rapid development to become one of the most flexible and scalable content management systems available. Drupal usage is estimated to be about 1% of the web, with over half a million sites running on this content management system. Drupal is designed from the ground up to quickly and easily allow a user to collect and share information in a variety of formats. Because of its modular design and a thriving user community, Drupal has capabilities for seamlessly deploying websites that interact with and employ all of the current web technologies like Semantic Web, RSS feeds and social networking like Twitter and Facebook. Additionally, Drupal has an extensible framework that allows experienced developers to implement additional custom features not already offered.

B. Drupal as part of Web 2.0

Drupal CMS is built upon the open source community contributed add-ons called *modules*. Each module is built to render a specific functionality to your web framework. These allow users to extend, build, and customize Drupal's core functionality per your requirements. Currently there are over 7500 [10] contributed modules to download. However, not all modules might suite your business need right out of the box, they might require an extension or sometimes a complete revision. Some key features of Web 2.0 such as interoperability, user-centered design specification can be readily answered by some modules.

C. Ease of Implementation

Ease of infrastructure implementation is another important constrains' in the science community. In most cases, scientists like to spend more time and efforts in developing the actual content than in its beautification and presentation. In most cases, Drupal modules are built for out the box use and could be plugged-in or removed from the CMS simply by a mouse click. Also, since most of the modules are community driven, bug-fixes and enhancements are taken care of by a group of active developers.

So, what are its benefits to a scientist?

By scientists concentrating on creating the content and community developers working on the CMS enhancements per Web 2.0, Drupal is quickly transforming into a powerful collaborative framework solution in the scientific research. New and improved user-centered interfaces based on Web 2.0 can be implemented right out of the box. Other interoperability issues with scientific data are also being solved by Drupal.

D. Semantic Web support

In achieving the goal of creating a collaborative environment for research findings that have divergent interpretations, it is important to choose a platform that is based on Semantic Web technology so that data is stored in a machine interpretable format as well as reuse existing knowledge bases.

New sets of languages are being developed in the scientific community for making the content accessible to machines [12]. Drupal is one such tool that makes it easier for the people to create machine-readable content and make it widely available over the Web. Drupal CMS intelligently stores the information in a well-defined machine-readable format [13]. It hides the complexity of the structure and elements of Semantic Web from the end user while yet rendering its benefits. There are modules that have the capability to expose structured information to the Web as a Resource Description Framework (RDF) or in an Ontology language (OWL) without requiring extensive implementation knowledge on the Semantic Web. Also, there are some modules such as *Linked Data*, by which existing RDF data from the Web could be hooked up to your Drupal site [14]. The Drupal - foundation of Software Collaboration Framework (SCF)[3] is OpenSource software toolkit which is freely available as part of the Drupal installation. SCF is semantic web based toolkit which includes many modules supporting scientific communities in publishing, annotating, sharing and discussing content.

E. Collaborative modules in Drupal

Drupal provides a powerful collaboration framework and, with its plethora of extensions, a vast range of possibilities. Content Construction Kit (CCK) allows user to add content and to create new content types using a web interface. In Drupal, each item of content is defined as a *node* and each node is created by a specific *content type*. Content type can also be defined as a skeleton for CMS, where you specify the field types for your page, example: Title, description, menu location, etc. A website can contain multiple type of content, such as news items, blog posts, polls, etc.

---

2 http://www.flickr.com

3 http://sciencecollaboration.org/

In Drupal there are numerous modules that can promote a collaborative environment. Simple modules such as *Blog* could be a great collaborative tool that can be used to communicate your insight on any topic. *Web File Manager* (WebFM) could be used to manage files in Drupal. Users can upload files by simple drag and drop based setup. Other functionalities such as file preview, description and search are also provided. *Change Management* of another module to collaborate for common interests, using which users can perform fundamental analysis for improved peer-based planning and change review. *Class* module could be another collaborative tool for teachers, which provide options to add properties such as grades, classes they are taking, etc. Other popular modules such as *Gene*, *RSS*, *Biblio*, *Gmap*, *Faceted Search*, *Tracker, Trigger, and Workflow* can also provide the potential scientific collaboration environment.

III.   USE CASES: SCIENCE WEBSITE

At Oak Ridge National Laboratory (ORNL) many recent federally funded U.S. DOE & NASA projects such as Wind Energy Data & Information Gateway (WENDI), Next-Generation Ecosystem Experiments (NGEE)[4], Spruce and Peatland Responses Under Climatic and Environmental Change (SPRUCE) [5], Earth Science Data in Digital Object Repository Architecture (ESDORA), and Hiaper Pole-to-Pole Observations (HIPPO)[6] are being implemented in Drupal for its easy implementation and robust collaborative framework [15]. The SPRUCE project provides a good example of how Drupal can be used to quickly and efficiently deploy a website for scientific collaboration. The SPRUCE website brings together a large group of collaborators from the ORNL and the U.S. Forest Service to assess the response of northern peatland ecosystems to increases in temperature and exposures to elevated atmospheric CO_2 concentrations. The website was quickly deployed in Drupal with the goals of allowing collaborators to communicate policies, deadlines and announcements, and share data as a repository. The site uses modules like Google Map (*GMap*) to provide detailed map of SPRUCE bog locations (Fig. 1) and *Web File Manager* to upload and download data files.

In ESDORA project, we are integrating NASA ORNL Data center's Fedora digital object repository with Drupal CMS using an open source framework, ISLANDORA, developed by University of Price Edward Island's Roberson Library (Figure 2).

Scientists from anywhere can view and manage the digital objects stored in our Fedora Repository via Drupal's graphic user interface (Figure 3). This includes ingest, purge, add data stream, searching and browsing by collection.

A complete scientific dataset contains several data streams, such as its metadata, actual data file, guide, docs, etc. Fedora stores its relationships, location of the data streams as a data object. By exposing the Fedora to ISLANDORA (Figure 4), it creates the necessary mappings between the Drupal User requests and Fedora's digital objects.

This also has the capability to export existing data streams as structured information to the Web as Resource
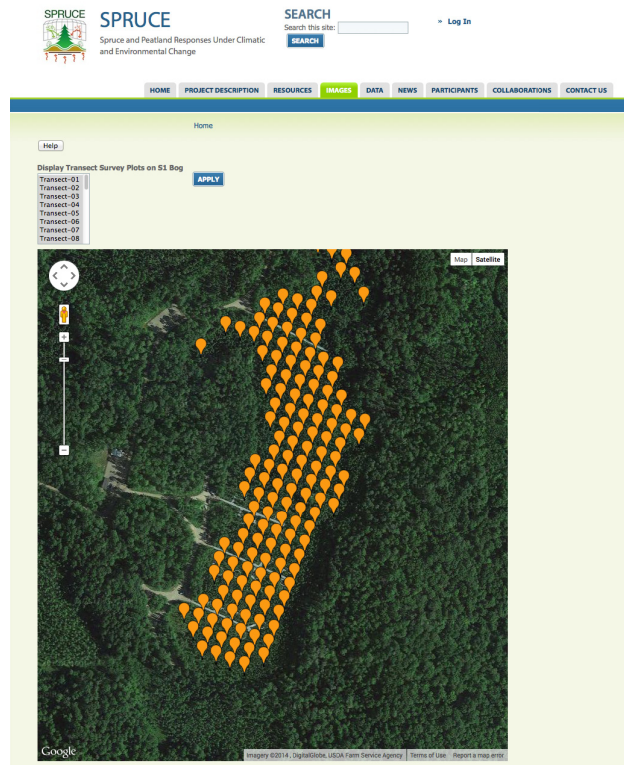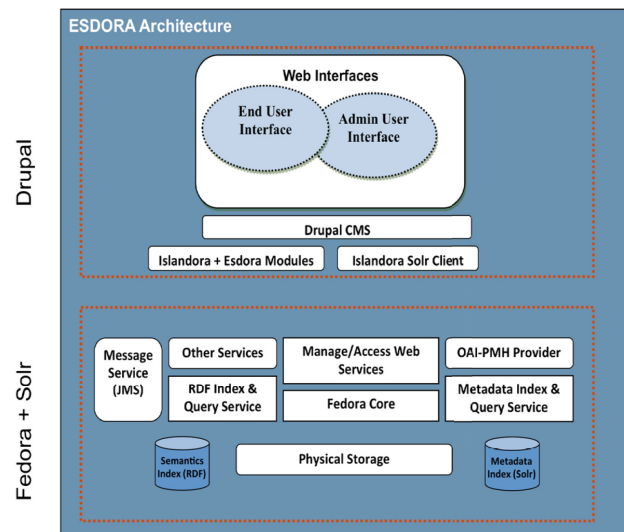

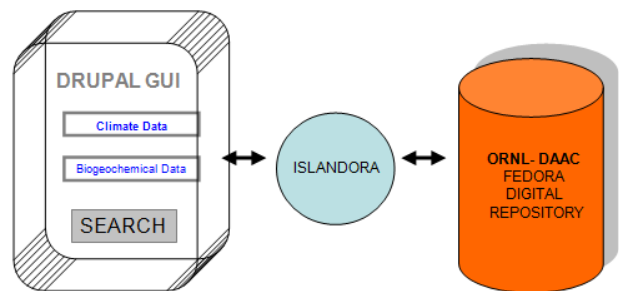Figure 1.   .SPRUCE project S1 Bog

Figure 2.   ESDORA Architecture

Figure 3.   Oak Ridge National Laboratory - Distributed Active Archive Center's Fedora - Drupal Repository

---

4 http://ngee.ornl.gov/
5 http://mnspruce.ornl.gov/
6 http://hippo.ornl.gov/

Description Framework (RDF). By exposing the data as RDF's, data objects could be reused or imported as external data resources.

WENDI is another DOE project for the archival, discovery, access, integration, and delivery of wind energy-related data and information. This project is intended to serve the information needs of a broad range of stakeholders in our nation's efforts to increase wind energy's contribution to U.S. electricity demand. To facilitate data search and retrieval functionality to this site, we used *iFrame* module to integrate an external metadata search tool called Mercury[7] [16]. Mercury uses WENDI's continually growing metadata holdings to enable users' discovery of and access to wind energy-related databases, publications, applications and web sites (Figure 5).

In NBII project, six of the twenty four Long Term Ecological Research sites in the USA have adopted Drupal to manage most of their information holdings. A core of seven custom content types stores personnel directory information, project, publications, research locations, datasets, data entities and measurement information, including details on methodologies, units of measure, frequency and date spans. All these interconnected informational categories expose their content via common XML specifications such as the Ecological Metadata Language or the Biological Data Profile (BDP), in addition to RSS feeds and more traditional user oriented formats like PDF or word. The Michigan University Biological Station is also collaborating with the group of LTER sites that co-develop this system

The USA National Phenology Network manages part of its metadata records using Drupal. The NBII is also developing a BDP based Drupal form to capture metadata through online services.

IV. CONCLUSION

In this article we discussed the need for collaboration in a research setting and how the new content management system Drupal is solving the problems of data interoperability, data longevity. The semantic Web is continuing to gain importance for its machine-machine interacting capabilities. Drupal is effectively employing the Semantic Web methodologies yet hiding their complexity of elements from the end users. We also discussed data sharing and where volume of scientific data could become a factor for choosing a software solution like Drupal and overall how Drupal is emerging as a collaborative framework in science research.

ACKNOWLEDGMENT

REFERENCES

[1] P.R., Orszag. (2010), Memorandum for the Heads of Executive Departments and Agencies. Available: http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-21.pdf [Accessed: January 15, 2011].
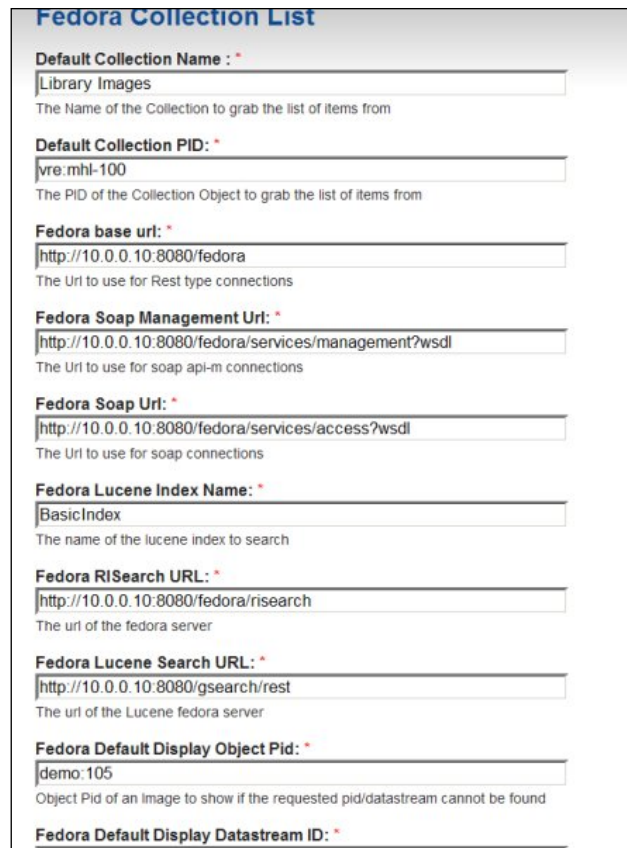
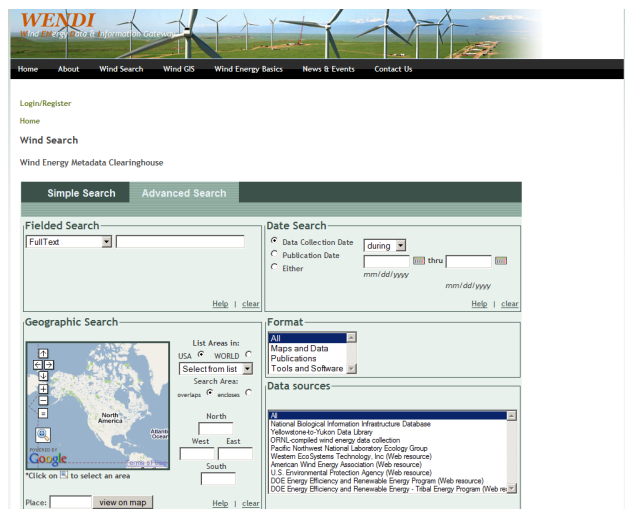Figure 4. Exposing Fedora to ISLANDORA

Figure 5. WENDI - Mercury metadata clearinghouse

[2] B. L. Clarke, Communication Patterns of Biomedical Scientists, *Federation Proceedings* 26 (1967) 1288-92.

[3] J. G. O'Connor, Growth of Multiple Authorship, DRTC Seminar 7 (1970) 463-83.

[4] T. D. Stokes and J. A. Hartley, Coauthorship, Social Structure and Influence within Specialties, Social Studies of Science 19 (1989) 101-25. http://dx.doi.org/10.1177/030631289019001003

[5] Pouchard, Line C., et al. "A Linked Science investigation: enhancing climate change data discovery with semantic technologies." Earth science informatics 6.3 (2013): 175-185.

[6] Devarakonda, R., & Shanafield, H. (2011, May). Drupal: Collaborative framework for science research. In Collaboration Technologies and Systems (CTS), 2011 International Conference on (pp. 643-643). IEEE.

[7] J. Gogolin, "The Web is Revolutionizing the Study of Climate Change", ON Magazine Issue 1. 2010. Available: http://www.emc.com/leadership/features/climate-change.htm [accessed January 11, 2011].

[8] Sheila A. McIlraith, Tran Cao Son, and Honglei Zeng. "Semantic Web Services", IEEE INTELLIGENT SYSTEMS. 2001. http://dx.doi.org/10.1109/5254.920599

[9] T. O'Reilly. What is Web 2.0 – design patterns and business models for the next generation of software (2005). Available:http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[10] Drupal – Open Source CMS | drupal.org. Available: http://drupal.org/about. [Accessed: January 04, 2011].

[11] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," Scientific American, May 2001, pp. 28-37.

[12] Das, S., Girard, L., Green, T., Weitzman, L., Lewis-Bowen, A., Clark, T.: Building biomedical Web communities using a semantically aware content management system. Briefings in Bioinformatics 10(2), 129–138 (2009). http://dx.doi.org/10.1093/bib/bbn052

[13] Bernstein, A., D. Karger, et al. (2009). Produce and Consume Linked Data with Drupal! The Semantic Web - ISWC 2009, Springer Berlin / Heidelberg. 5823: 763-778-778.

[14] Devarakonda, Ranjeet, et al. "Semantic search integration to climate data." Collaboration Technologies and Systems (CTS), 2014 International Conference on. IEEE, 2014.

[15] R. Devarakonda, G. Palanisamy, B.E. Wilson, and J.M. Green, Mercury: reusable metadata management data discovery and access system, Earth Science Informatics, vol. 3, 1: 87-94, May 2010. http://dx.doi.org/10.1007/s12145-010-0050-7

AUTHORS

Ranjeet Devarakonda and **Giriprakash Palanisamy** are with Oak Ridge National Laboratory – Climate Change Science Institute.

Inigo San Gil is with University of New Mexico.