

4-29-2008

Partnering with Agronomy Faculty to Create a Framework of Data Sharing

Marianne S. Bracke

Purdue University, mbracke@purdue.edu

Michael Witt

Purdue University, mwitt@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_research

Bracke, Marianne S. and Witt, Michael, "Partnering with Agronomy Faculty to Create a Framework of Data Sharing" (2008). *Libraries Research Publications*. Paper 86.

http://docs.lib.purdue.edu/lib_research/86

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

PARTNERSHIP

PARTICIPANTS

Sylvie Brouder, Agronomy professor and soil fertility researcher
Michael Witt, Interdisciplinary Research Librarian
Marianne Stowell Bracke, Agricultural Sciences Information Specialist

To be successful in this endeavor, three areas must be addressed:

INTERDISCIPLINARITY in SCIENCE

- Team work based on contributing expertise (familiar model to many science researchers)
- Faculty acknowledge the need for long-term data management plans and often a willingness to see data shared or re-purposed
- Opportunities exist if you can articulate value of library science to disciplinary researchers and faculty

ADMINISTRATIVE SUPPORT

- Support from the Dean of Libraries
 - Recognized need to engage faculty in new ways
 - Convinced faculty and department heads of the applicability of librarian skills to research problems
- Associate Dean for Research was appointed
 - Gave Libraries a seat at the campus table
 - Gave Libraries a better understanding of the processes behind sponsored research

GROUND LEVEL PARTNERSHIP

- Ability to articulate what the library can do
 - What are our skills unique to library science
 - What librarians can deliver
 - These are new approaches – willingness to learn as you go
- Willingness to approach faculty as peers
- Time
 - Developing relationships takes time
 - The reality is that numerous proposals/overtures may not be accepted - this takes time and requires that the library is ok with occasional rejection
- Develop a Network to draw upon
 - No one has all the requisite talents or resources for a project
 - Develop relationships within the libraries and other campus units (e.g., computing) as well

PROJECT

This is a collaboration between the Department of Agronomy and librarians at Purdue University. The goal of this project is to preserve and increase the accessibility of water quality datasets being collected by researchers at the Water Quality Field Station (WQFS) at Purdue's Agronomy Center for Research and Education (ACRE) and to create a functional proof-of-concept system to demonstrate a workflow for automating the generation of descriptive metadata and the ingestion of these datasets into an institutional repository. This work is supported by a seed grant from the Center for the Environment at Purdue's Discovery Park.



THE PROBLEM

In the test fields of the WQFS, sensors regularly sampled the water flow. Data had been generated and some of it analyzed, but there was no data management plan in place for preservation, description or long-term storage of this data. This data, if organized and made more accessible, would be of use to numerous other experiments being performed on the same area of land. The librarians gathered the old data from the WQFS that was scattered across various files and began to develop a plan to organize, describe, and preserve the data, as well as anticipating the ingestion of future data. The librarians chose an existing metadata schema, Ecological Metadata Language (EML), to describe the data.

ACKNOWLEDGING SUPPORT FROM
Discovery Park
Center for the Environment

INFRASTRUCTURE

AN INSTITUTIONAL DATA REPOSITORY

While institutional repositories typically contain electronic documents such as journal article pre-prints, technical and working papers, and other grey literature, some libraries are beginning to use them to curate other scholarly materials such as research datasets. The Purdue University Libraries are developing an institutional repository that is based on the FEDORA web services framework that is dedicated to preserving and providing access to research data. Custom metadata schemas and disseminators can be created to support different applications and digital objects, such as water quality data.

INGESTING DATA

In reviewing five year's worth of past WQFS data, two general categories of data emerged: unstructured and structured. Unstructured data were in a variety of file formats and generally lacked any consistent organization. These were, for example, nutrient analysis of water samples that were performed in the lab and manually inputted into a spreadsheet and field observations that were typed into a word processor document.

A field technician from the facility is working with the librarians to accession these data into the repository using a web-based form and http file upload to individual describe and ingest these unstructured data.

Structured data include the "tip counts" from a sensor that measure the flow of water as it drains out of the test plots. The output from this device, a Campbell CR-10 data logger, is comma-delimited text file. These are periodically downloaded by a graduate student onto a laptop in the field. Back at the dry lab, the files are downloaded to computer, and a batch script is run that converts the files into XML, using a schema developed by the librarians. After the files have been converted, they can be validated and ingested into the FEDORA repository as a batch process.



FUTURE WORK

The workflow and automation have been established for both past and future flow data. We would like to identify other, complementary structured data streams from the WQFS and perhaps other facilities at ACRE to add to the collection in the data repository. Cataloging and accessioning the unstructured data remains a work-in-progress. Integrating the data repository into the future scientific workflow may influence more regimented practices for recording observations and other unstructured data in a more consistent manner.

Partnering with Agronomy Faculty to Create a Framework of Data Sharing

Marianne S. Bracke (mbracke@purdue.edu) & Michael Witt (mwitt@purdue.edu), Purdue University Libraries, West Lafayette, Indiana, USA