Data Under Construction: Data Profiling for a Water Quality Lab

Jake R. Carlson  
*Purdue University*

Marianne S. Bracke  
*Purdue University*, mbracke@purdue.edu

Follow this and additional works at: [http://docs.lib.purdue.edu/lib_fsdocs](http://docs.lib.purdue.edu/lib_fsdocs)  
Part of the [Library and Information Science Commons](http://docs.lib.purdue.edu/lib_fsdocs)

Recommended Citation  
http://docs.lib.purdue.edu/lib_fsdocs/8

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.
Data Under Construction: Data Profiling for a Water Quality Lab

Jake Carlson
Associate Professor &
Data Services Specialist

Marianne Stowell Bracke
Associate Professor &
Agricultural Sciences Information Specialist
Data Sharing

Tables

<table>
<thead>
<tr>
<th>Position</th>
<th>Soil</th>
<th>50% Stand at Height</th>
<th>8 Wt.</th>
<th>Kanekal Plant.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-n</td>
<td>in</td>
<td>42.5</td>
<td>21</td>
<td>21,500</td>
</tr>
<tr>
<td>b-n</td>
<td>bet.</td>
<td>61.6</td>
<td>19</td>
<td>34,685</td>
</tr>
<tr>
<td>c-n</td>
<td>bet.</td>
<td>62.2</td>
<td>19</td>
<td>22,958</td>
</tr>
<tr>
<td>d-n</td>
<td>bet.</td>
<td>63.4</td>
<td>17</td>
<td>25,708</td>
</tr>
</tbody>
</table>

- Strip-preparation included row cleaners plus 3 fluted coulters per row.
- a-n or between last year's corn rows.
- Average daily soil temperature for first 5 weeks.

Analysis

Results

Pre-analysis

Samples

Readout

Master spreadsheet

All photos: Purdue School of Agricultural
The Water Quality Field Station

- On a 991 acre farm facility northwest of Purdue opened in 1992.
- Used to identify agricultural practices that minimize movement of AG chemicals into water supplies.
- Informs the development of new and more ecologically-balanced technologies for crop production.
Data gathered by one graduate student may be useful to others, but there are few common practices for documenting, managing or sharing data in the lab.
Project Steps

Utilize Data Curation Profiles to collect information about current data gathering, workflow and documentation.

Identify common issues and needs as observed in the Data Curation Profiles.

Produce a report with recommendations and possible approaches to addressing issues and needs.
The Data Curation Profile is a means to determine:

- Information about a particular data set.
- What a researcher is doing to manage / curate the data set.
- What a researcher would like to do with the data.
DCPs and the WQFS

Modified for this project:

- The Data Set
- The Data Lifecycle
- Sharing
- Organization and Description
- Tools
- Data Management
Identify

- 6 interviews with Graduate Students conducted in summer of 2011
- Developed Data Curation Profiles from these interviews: http://datacurationprofiles.org
- Reviewed DCPs for needs
Data are not being generated or processed in ways that could facilitate sharing locally or externally.

- All graduate students identified other data sets in the lab that would be potentially useful to them. Most graduate students who received data reported problems understanding or making use of the data.
Graduate Students expressed a lack of knowledge as to how they could / should share their data with others.

- Most students were willing to share their data, but expressed concerns over timing and intended use.
- Graduate students are not always aware of expectations for managing, sharing, or preserving their data from funding agencies or other external sources.
- Incentives for Graduate Students to share their data needs further exploration.
Graduate Students stated that they lack knowledge and skills of how they should document, describe, organize and manage their data.

- These activities tend to be done in relative isolation from the lab, or even the advisor.
- Physical lab notebooks are still the primary means of documentation / provenance.
It is unclear how the data are being managed long-term, especially once the graduate student moves on and the data are no longer actively used.

- Inheriting data from previous graduate students was common and potentially problematic.
- Some of the graduate students keep elements of their data for historical record purposes.
- Data management and security issues are not given a great deal of attention by graduate students.
Customer Requirements Document (REQ)

CRIS - Computational Research Infrastructure for the Sciences
Brouder Field Plot Data Management
Document ID: REQ000006
Thank you!

Jake Carlson
jrcarlso@purdue.edu

Marianne Stowell Bracke
mbracke@purdue.edu