The REACCH data management system

Erich Seamon (erichs@uidaho.edu) UI, Paul Gessler UI, and Sanford Eigenbrode UI

The REACCH Data Management System is a "behind-the-scenes" collection of information technology tools (e.g., servers, networking hardware, database software, web portal and interface software, and people) organized to help researchers and stakeholders store and archive data, explore and discover data, and integrate data sets for collaborative research. It is designed to be secure, expandable, and flexible. In partnership with the Northwest Knowledge Network (NKN) at UI we are working to preserve preserve the data and analysis applications for the life of the REACCH project and beyond. The data and developed applications will be an important legacy of the REACCH project that will be accessible via our collaborating institutions and national repositories.

IMPACT

The REACCH data management system is a "behind-the-scenes" collection of information technology tools (e.g., servers, networking hardware, database software, web portal and interface software, and people) organized to help researchers and stakeholders store and archive data, explore and discover data, and integrate data sets for collaborative research.

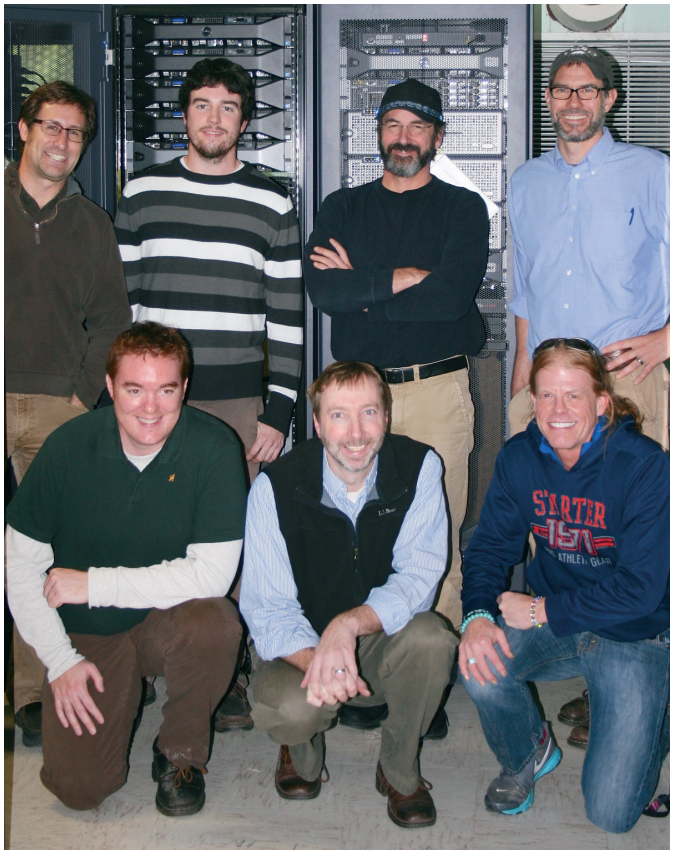

Figure 1. *Members of the REACCH cyberinfrastructure and NKN personnel in front of enterprise servers: back row from left: Paul Gessler, Stephen Fricke, David Vollmer, Luke Sheneman. Front row from left: Ed Flathers, Bruce Godfrey, Erich Seamon. Photo by Brad Stokes.*

The integration of these tools and technologies is critical to our project success. Most of the REACCH research teams are collecting data using a wide variety of methods and formats. With a diverse number of research locations and types of data, the integration of such information to facilitate useful decision making is an important challenge.

To address this challenge, the REACCH Data Management and Cyberinfrastructure team (Figure 1) has developed and implemented a technology strategy focuses on three core areas:

- Data management, harvesting, and ingestion through a central web portal. Our www.reacchpna.org web portal is a central point for both public and secure information access, where users can search for and analyze data. Our REACCH Data Library, accessible via www.reacchpna.org, uses technology that allows us to upload or harvest data from a variety of sources and then allow those data to be searchable.
- Data meta-tagging and transformation. Our data management efforts have focused strongly on ensuring that the description of data is complete. This helps ensure that the data are easily discoverable using web search tools and that the data can be used indefinitely.
- Data exposure and consumption. With data stored and exposed using standard web protocols, we can make the data available in a variety of ways so that users can download or link to the data for use in a wide variety of applications and modes. Advantages include the ability to dynamically link to the data from within specific applications without having to download the data (Figure 2).

Access to REACCH data can be grouped into two areas: the REACCH Data Library and the REACCH Analysis Library. Accessible from the www.reacchpna.org portal, the Data Library and the Analysis Library provide clear and straightforward mechanisms to upload, search for, and analyze REACCH-based data.

The REACCH Data Library provides access to raw data, publications, presentations, images, and other content that may be REACCH project related. By meta-tagging each data set and research product, we can see relationships between data and products and allow users to explore how data and publications are interconnected.

The REACCH Analysis Library allows users to examine data within the REACCH Data Library for research-related functions, agriculture-based decision making, education, and other stakeholder needs. Some of the tools currently being developed for the

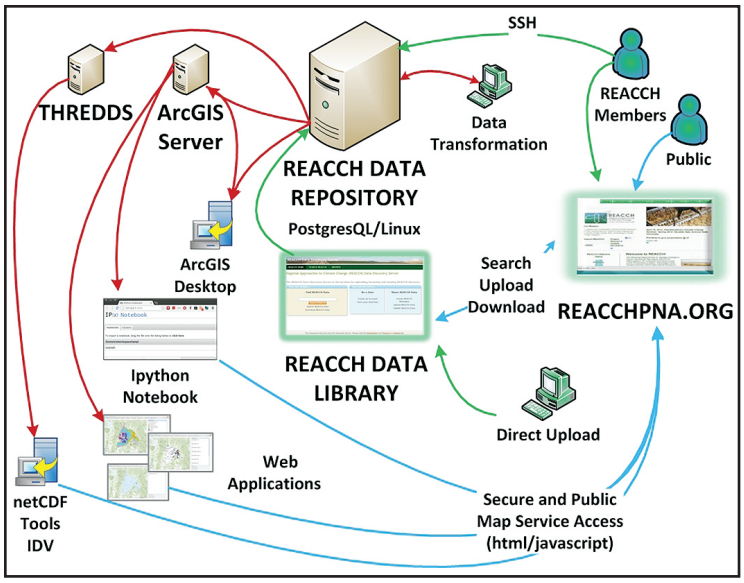

Figure 2. *The REACCH technology architecture focuses on harvesting, meta-tagging, and then integrating data through the use of geographically based web service protocols.*

REACCH Analysis Library includes the following:

- Inland Northwest growing degree calculator
- Climatic model data aggregation and filtering
- Inland Pacific Northwest biotic data examination
- Agroecozone geospatial model development
- Interrogation of data using interactive programming tools such as Python™

Figure 3. *REACCH students participating in a data management course in August 2013. Photo by Brad Stokes.*