Developing an integrated agriculture data management system

Engaged research teams often times generate large amounts of diverse, heterogeneous data. Yet the integration and use of such data in a collaborative manner can be difficult. The areas of data management and how best to envision integrated, yet heterogeneous, research collaboration have spawned several models for data storage and analysis (Chernenak 2000, Papazoglou, 2004). **REACCH PNA** – a USDA funded coordinated agriculture project to explore climate impacts on cereal production systems in the pacific northwest, has developed a data management system for these areas of heterogeneous science data integration.

REACCH PNA DM System: Integrated Data Science Architecture

The REACCH data management effort has a focus to develop modular, sustainable, and extensible systems/processes that would allow for the collection, storing, and analyzing of REACCH-related data and content in a transdisciplinary manner. In support of this strategy –we have built out five core integrated systems to implement this approach:

1. Our http://www.reacchpna.org portal;
2. the REACCH Data Library for data uploading/searching (http://data.reacchpna.org) ;
3. the REACCH Analysis Library for analytical tools (http://analysis.reacchpna.org);
4. a THREDDS Data Catalog for array based meteorological and climate scenario data(http://thredds.reacchpna..org); and
5. an Interactive Python Notebook Server (http://IPython.reacchpna.org).

Integrating data using RESTful web services

As mentioned above, the data management methodology is based on the development of systems that are modular, sustainable after the life of the project, and extensible – with regards to interacting with other systems and processes, as well as usable by researchers and the public at large.
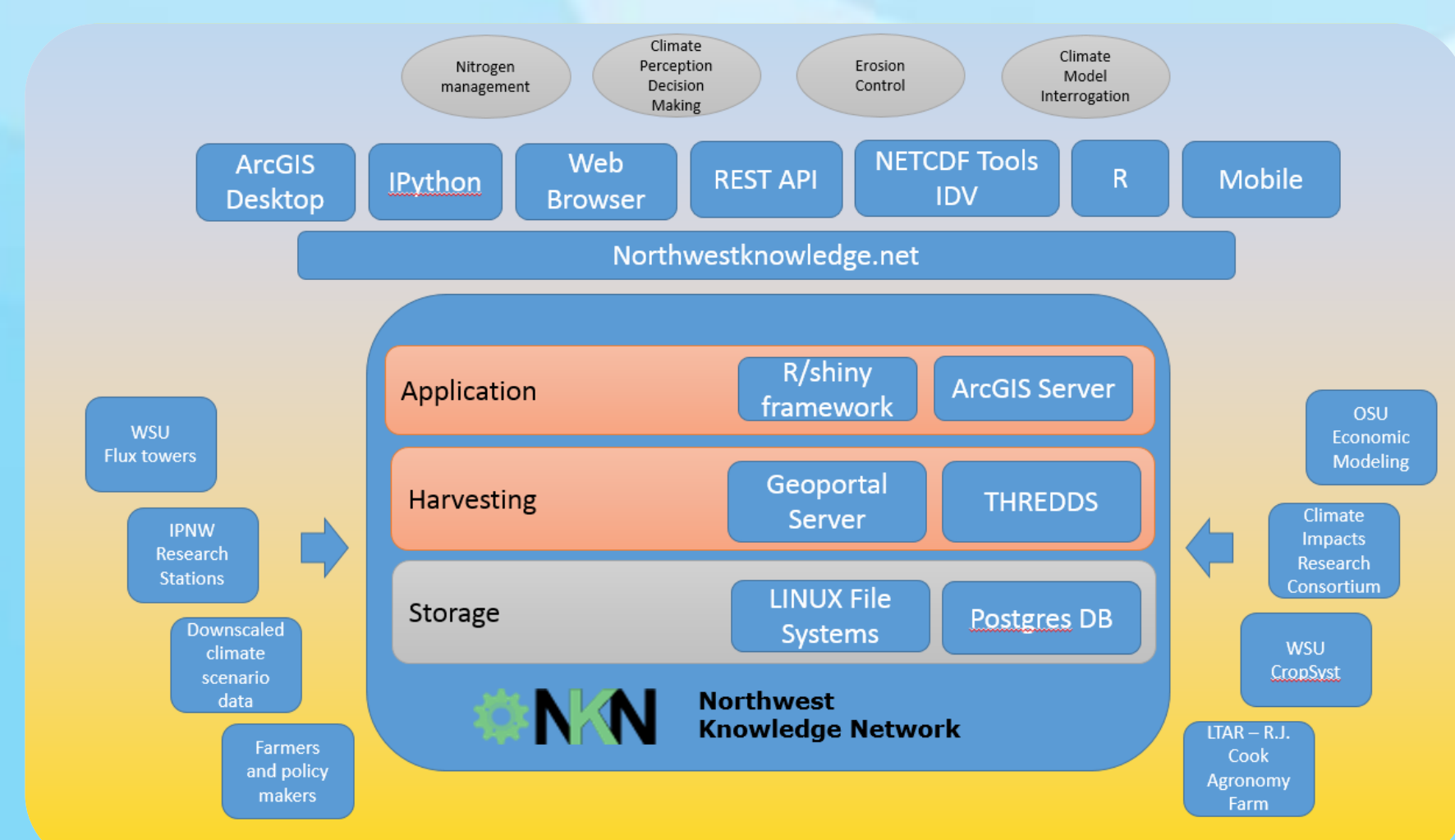
In addition, we use Representational State Transfer Protocol (REST) to display and provide query functionality to data stored in our REACCH Data Library, as well as our THREDDS server, which contains array-based historical and future climate scenario datasets. As an analytic ingestion tool – IPython (interactive Python) serves as a mechanism to access and combine datasets.

Interactive Python Notebook Server. Interactive Python, or Ipython, is a somewhat new development over the last two years, is to enable the compilation of Python within a web browser. This shared notebook model provides a new way for scientific researchers to collaborate, in real-time, on data analysis and interrogation using Python (https://ipython.reacchpna.org).

THREDDS. Thematic Realtime Environmental Data Distribution Services (THREDDS), is a data cataloging approach that has been developed UNIDATA, a group that is part of the University Consortium of Atmospheric Research (UCAR). THREDDS is a java-based server technology that is used for dissemination, aggregation, and sub-setting of multi-variable data, such as NetCDF formatted datasets.

Figure 1. REACCH Data Interaction Diagram (2013)

Implemented Solution Case Study: REACCHPNA

REACCH Data Library

The REACCH Data Library, accessed from the REACCH web portal - is the core location to access REACCH data. Implemented using ESRI's Geoportal Server, running on Linux and using a PostgreSQL geospatial database, the Data Library is a common location for heterogeneous data access.
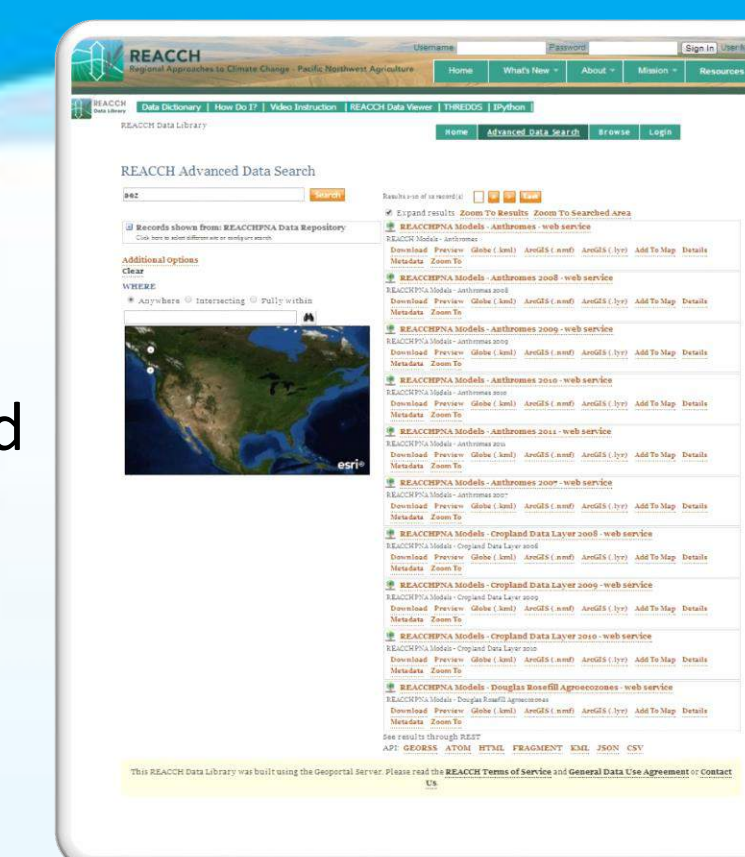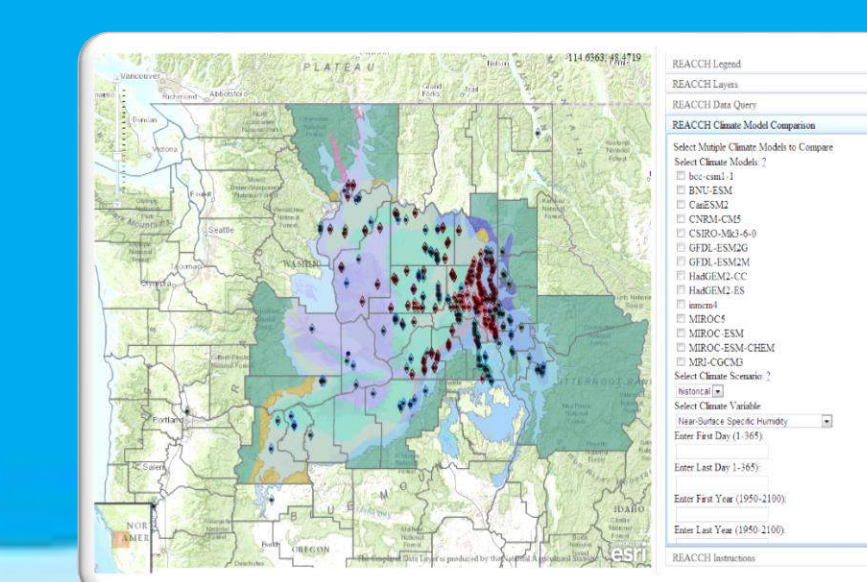

Figure 3. REACCH Data Library

REACCHPNA.ORG Web Portal

The REACCH Web Portal (https://www.reacchpna.org) is the central entry point for all public and secure information. REACCH members access the portal using a secure login and password, which in turn provides varying access to data uploading, searching, and analysis tools.
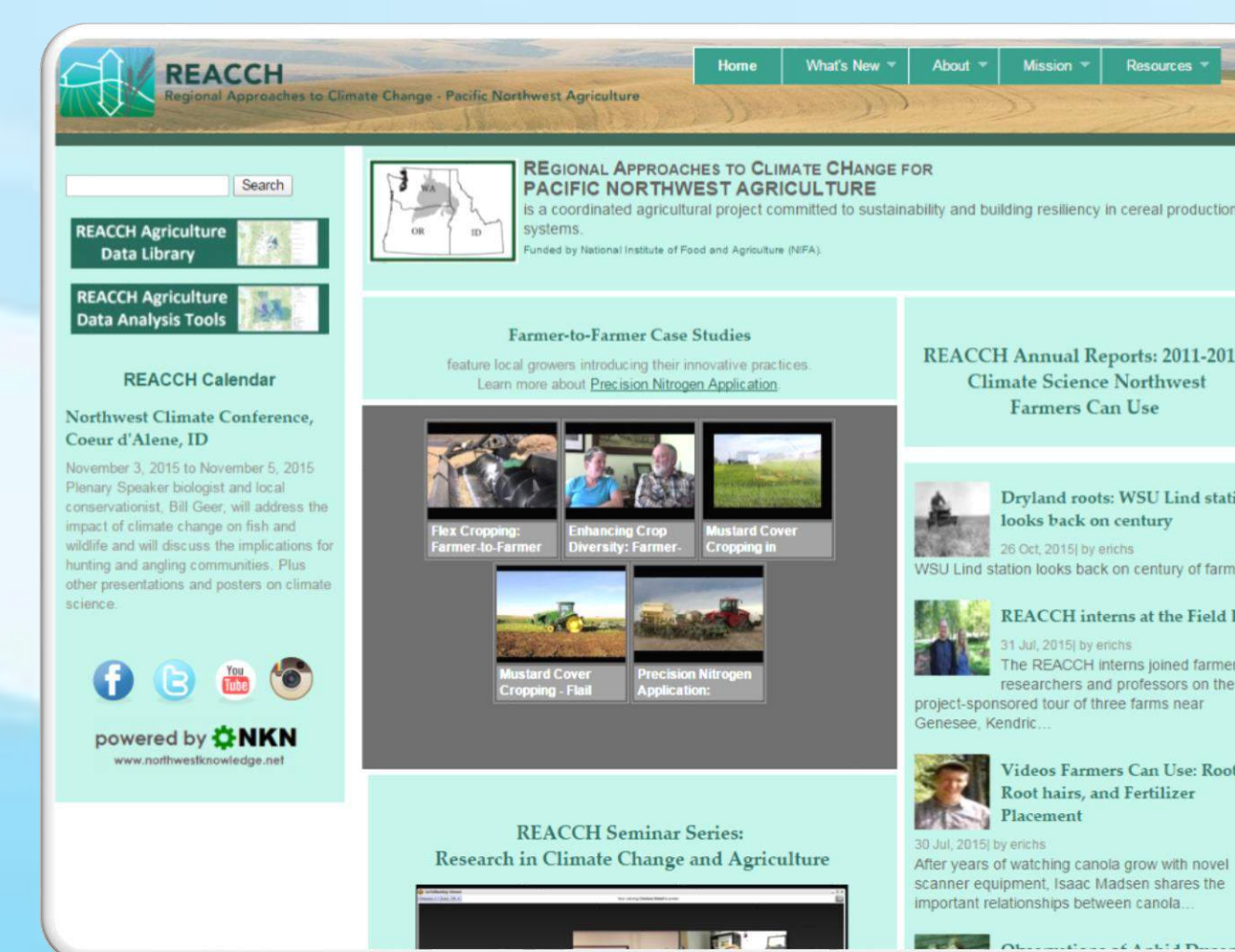

Figure 2. https://reacchpna.org web portal

REACCH Data Policy Agreements

The REACCH Data Policy and Data Agreement is an important aspect of the overall project. Defining and describing the policies that regulate data contribution, data management, as well as the protocols and procedures that researchers will abide by regarding data collaboration – is extremely important.

REACCH Analysis Library

Figure 4. REACCH Analysis Library

The REACCH Analysis Library , also accessed from the REACCH web portal, is a central location where we provide access to all of our technology-based analysis tools. Using ArcGIS Server, Javascript, and Python.

Transformation steps:
1. Data is uploaded to our data library, in varying formats;
2. That data is transformed, when possible, into geographic datasets and stored in our geospatially-enabled PostgreSQL database;
3. These data layers are then exposed via a web/map service using ArcGIS server that is then consumed and manipulated using Javascript and Python.

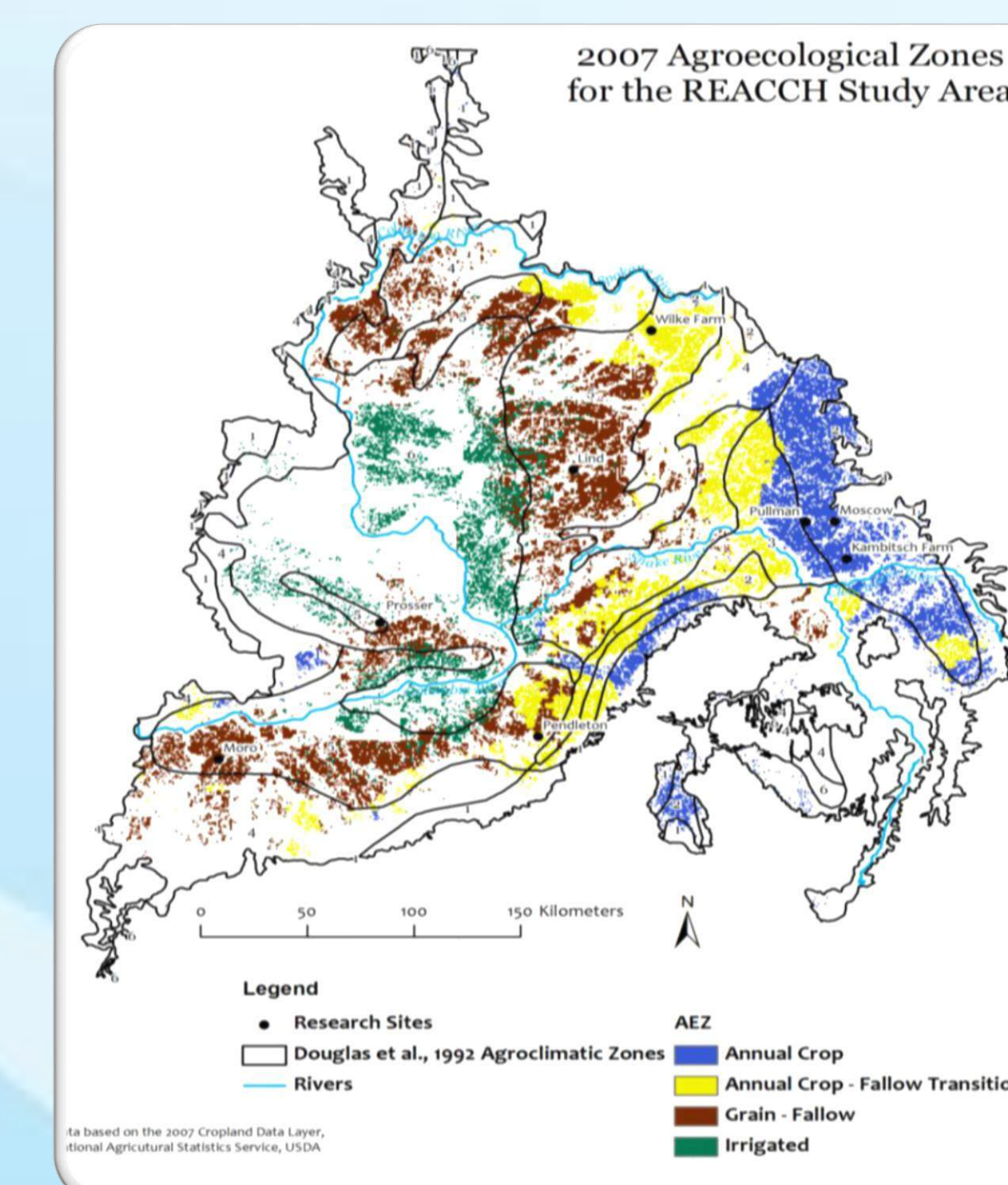### REACCHPNA Example of Integrated Data Analysis

Figure 7. Estimating Agroecozone classes for 2007 (Huggins, 2013)
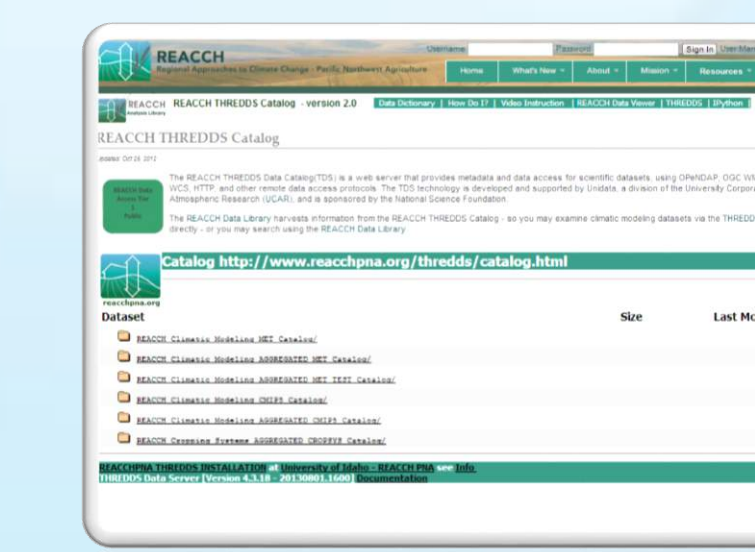
REACCH Interactive Python and THREDDS Server
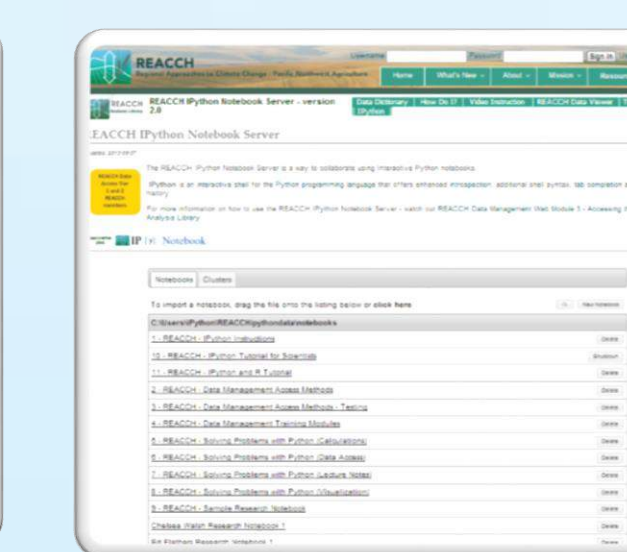

Figure 5. REACCH THREDDS Server

Figure 6. REACCH Interactive Python Server

The REACCH Interactive Python Server (Figure 7) is a server-side use of Interactive Python, exposed to REACCH members for collaboration and programming purposes. The REACCH THREDDS Data Catalog (Figure 6) is a server-side software technology that aggregates large datasets that cannot be typically stored in a database. THREDDS is a Java-based technology that is developed by UNIDATA, a technology wing of the University Corporation forAtmospheric Research (UCAR).
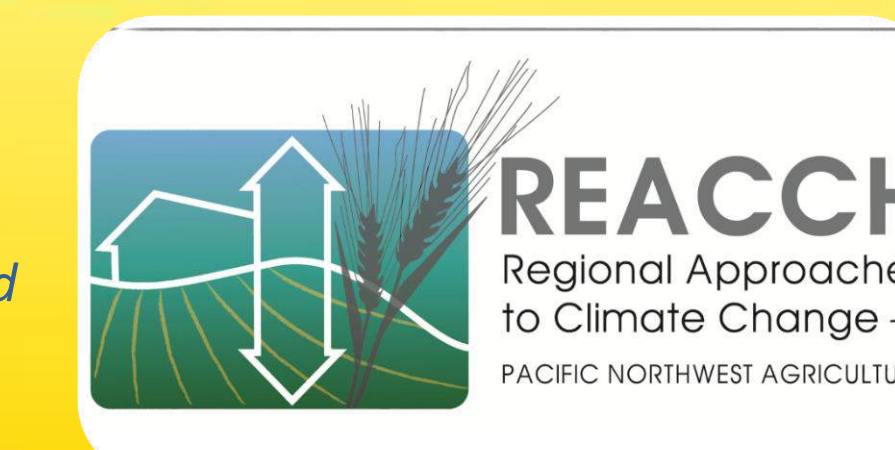
Leveraging the solution at a national level: National Ag Data Network Harmonization
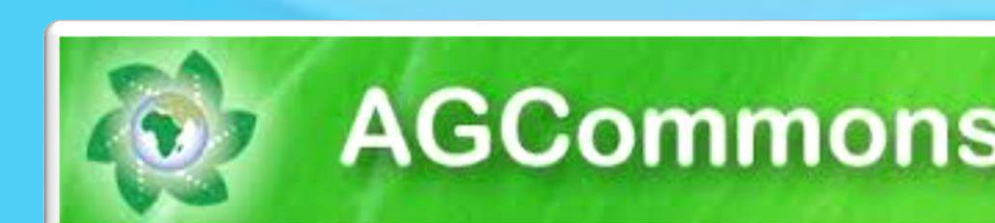
The REACCH team is engaged with a number of collaborators at the University of Florida, and other institutions, to establish a USDA based national data network for agriculture.
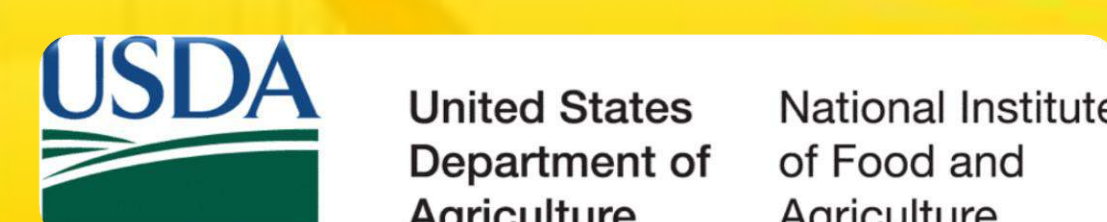
"Developing an agriculturally-focused data management system for climate assessment, adaptation, and mitigation: Regional Approaches to Climate Change for Pacific Northwest Agriculture (REACCHPNA)"

Erich Seamon[1] , Paul Gessler[1], Edward Flathers[1], Luke Sheneman[3], Sanford Eigenbrode[2], Richard Rupp[4]
University of Idaho
[2]College of Agricultural Sciences, Department of Plant, Soil, and Entomological Sciences
[1]College of Natural Resources, Department of Forest, Rangeland, and Fire Sciences
[3]Office of Research and Economic Development
[4]Washington State University

University of Idaho