

Developed by Team EcoData: Michael Solomon and Colby Blair

Sponsored by Alex Premier, College of Natural Resources

Problem

Large amounts of data are often collected for modern scientific projects. Once this data is collected, it must be stored in an accessible location with easy access to tools for information management.

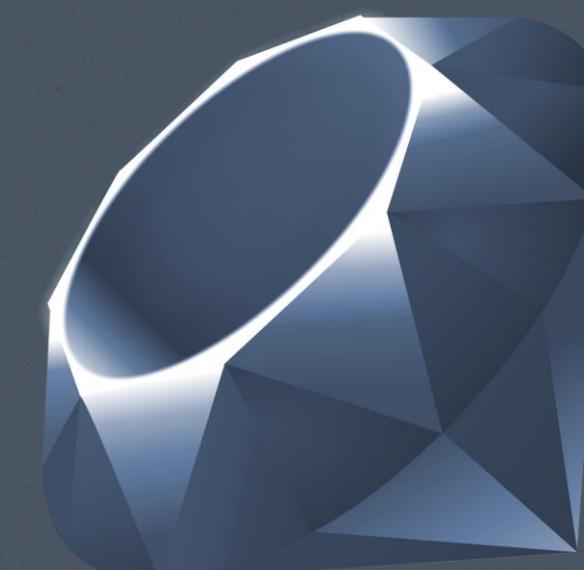
Centralized databases often provide storage and access, but do not provide easy-to-use data management tools or basic data exploration such as graphing and visualization.

Our primary target is easing the management of large sets of ecological data from projects like PTAGIS, which collects information on tagged fish in the Columbia Basin.

A solution must meet the following requirements:

- Provide:
 - Reliable data storage
 - Basic data manipulations
 - Simple visualizations
- Have a gentle learning curve
- Allow remote access
- Allow local access
- Support many data sets (i.e., not just PTAGIS data)

Employed Technologies



Ruby
1.9.3



Ruby on Rails
3.2.0

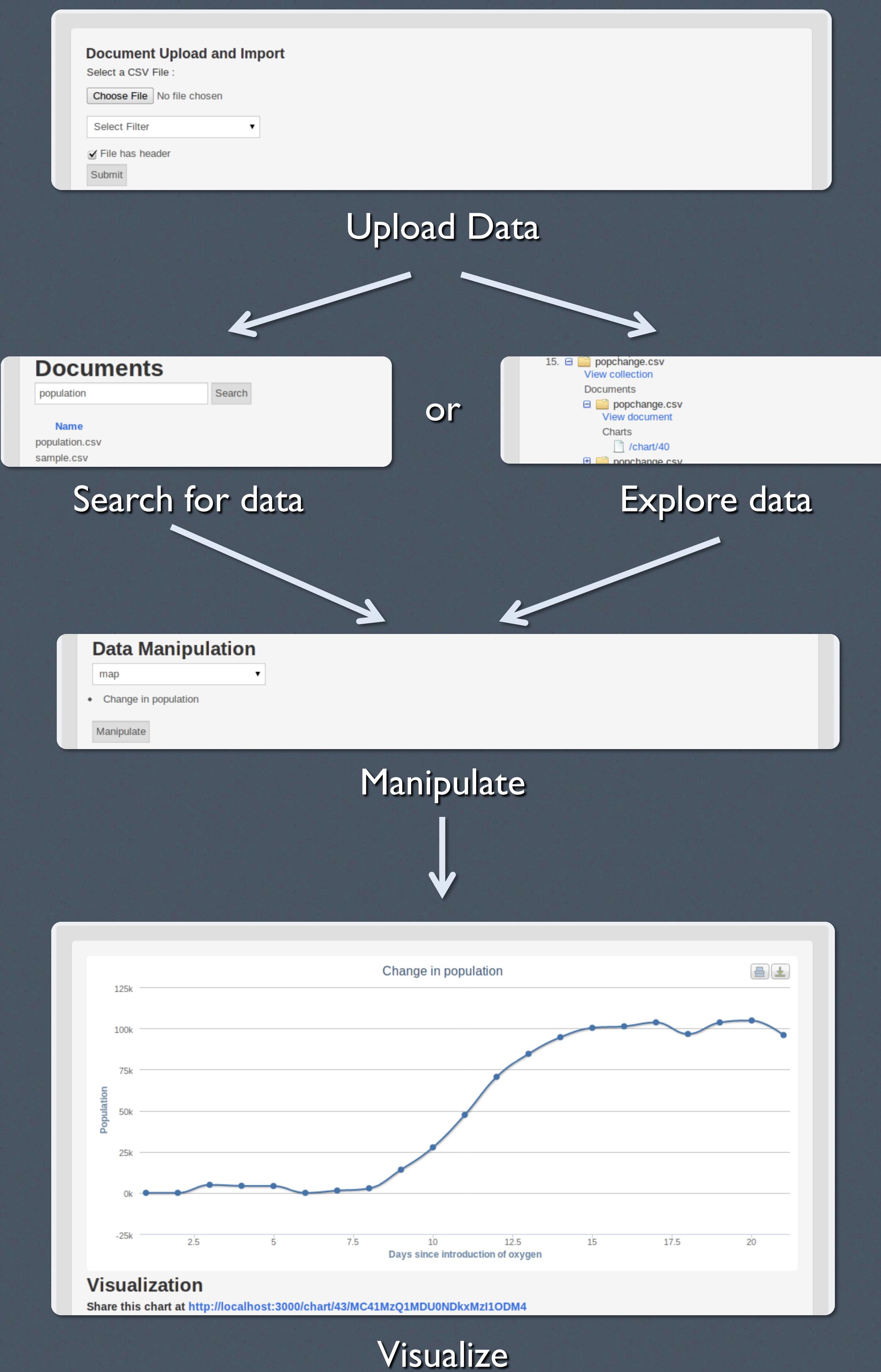


CouchDB
1.2.0

Project Goal

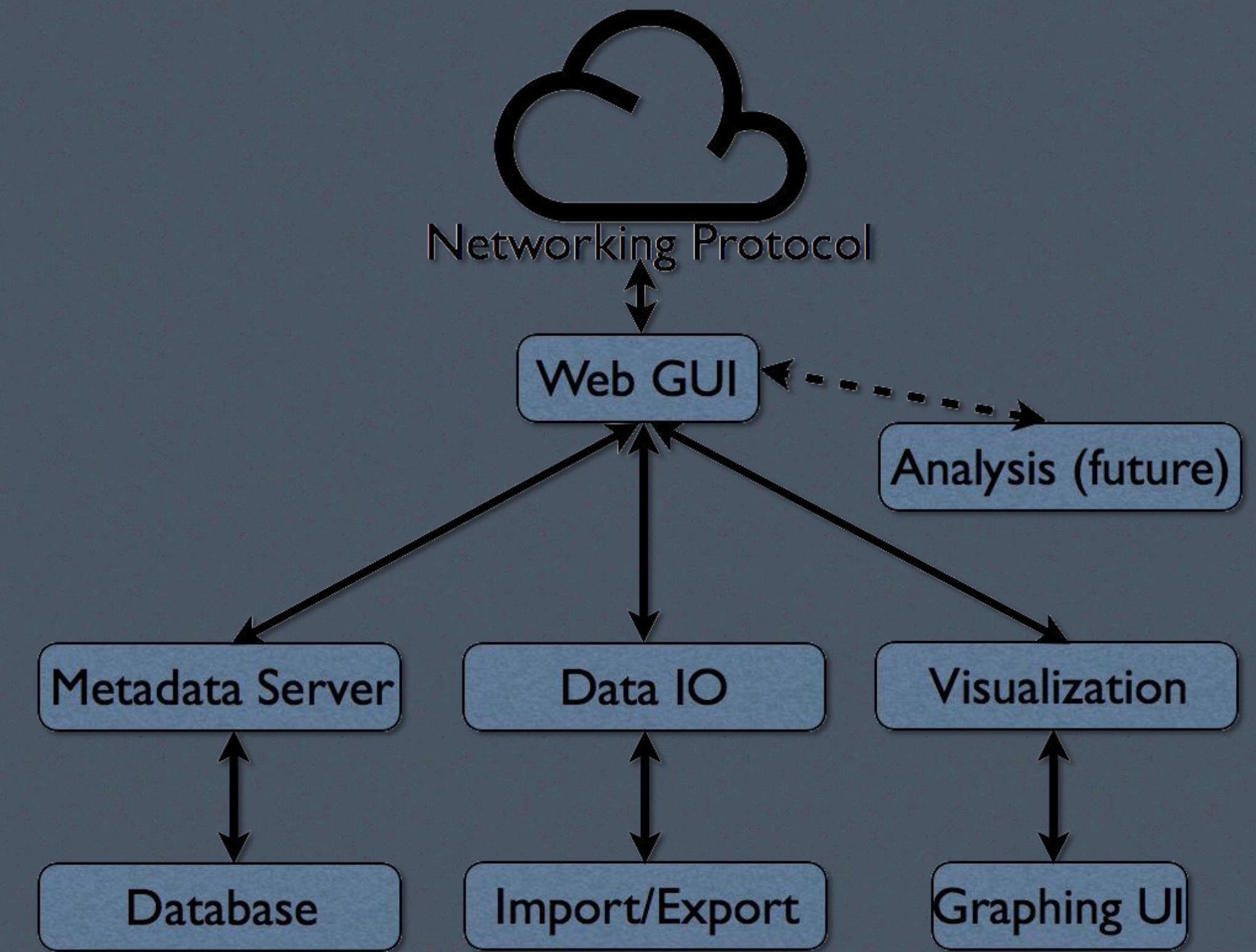
Design and implement a web-based application to store, manipulate, and visualize scientific data.

Typical Data Workflow



Design

Hatch is a web application that can be accessed remotely or run on a local machine and accessed through a web browser.



Data Import

Documents are uploaded to the system and stored efficiently in Apache CouchDB, a scalable document storage database that can automatically replicate valuable data.

Data Manipulations

Basic manipulations can then be performed on the data. Search and data merging uses map-reduce written in JavaScript running inside CouchDB. Manipulations are designed to be extensible for future integration with a sophisticated analysis pipeline.

Data Visualization

Visualization is accomplished through the HighCharts JavaScript library. It enables interactive charts and graphs, including live charts that update at a default of once per second.