Designing a Queriable Community Data System

Step 1: Data Capturing
- Simple data output (flat files)
- No imposed data structure
- Handles “bad” data okay

Low queriability

Step 2: Data Pre-Processing
- Flexibly for system - can handle any data, little overhead
- Inflexible for user - need to download multiple data files and process locally before performing analysis

Complex, lots of data pre-processing

Step 3: Data Sharing
Low vs. High queriability

Far left is easy for system development and maintenance, but awkward for end usability.

Far right is easy for end usability, but awkward for system design/maintenance and data pre-processing.

We need the best of both worlds. A system that requires as little data pre-processing as possible, yet has the ability to provide powerful queries and automated data visualizations such as plots and graphs.

High queriability

- Advanced, dynamic output w/ metadata
- Requires structured data
- Requires clean data

Low queriability

- Simple data output (flat files)
- No imposed data structure
- Handles “bad” data okay

A low queriable data system requires little to no data pre-processing before the data is input into the system.

All queriable data is accessible.

Not all accessible data is queriable.

A community data system is a self-organizing scientific and technical community with a common interest for information management, systems, workflow, and science. We need the best of both worlds. A system that requires as little data pre-processing as possible, yet has the ability to provide powerful queries and automated data visualizations such as plots and graphs.

We are a self-organizing scientific and technical community with a common interest for information management, systems, workflow, and science.