THE VOEIS HIS GATEWAY
A REST Interface for HydroServer using ODM 1.1
WHO AM I?

• Pol Llovet
  • Research Software Engineer
  • Research Computing Group
  • Montana State University
HIS IS ESSENTIAL

- Data publishing
- Data discovery

- HIS Gateway **does not** replicate the functionality of HIS
It is assumed that the data in HIS is accurate, that it is well curated. But, it is well known that there are no guarantees about the quality of the data management practices used prior to the data being published in HIS. It can be difficult to This is a known issue, and efforts are being by the community to pin down minimum curation levels, etc.
DATA MANAGEMENT

• NSF requires data management
• Increased investigator burden
• HIS is part of the data management solution

CUAHSI HIS software does fill this need to some extent, but our requirements were more complicated
  – there are point tools, but not end-to-end software solutions
  – collaboration and complex user management/access is not available
FULL DATA MANAGEMENT

- Data Versioning
- QA/QC
- Detailed project-based user access
  - Publishing and collaboration
- One-click publishing to HIS
VOEIS

*NSF Montana EPSCoR American Recovery and Reinvestment Act program with grant award M66012/66013

- Virtual Observatory and Ecological Informatics System
- Integrated sensor and ecological informatics
- NSF EPSCOR Track 2*
- https://voeis.msu.montana.edu

Open source data management and publication software
Designed to store and organize a variety of ecological data
The system supports and interfaces with CUAHSI HIS
PROJECT BASED DATA

- Organized into Sites
- Versioned data
- QA/QC
- Data is public/private until...
VOEIS ARCHITECTURE

- Cross Platform
- Yogo Framework Application
  - Database Agnostic
  - Versioned Data
- All Open Source
VOIES USES HIS

- Push data to HIS for:
  - Discovery
  - Publishing
  - Collaboration with the community
ENOUGH ABOUT VOEIS

• This presentation is not about VOEIS

BUT:

• VOEIS was the reason for the HIS Gateway
WHY BUILD A HIS GATEWAY?

the HIS Gateway was built specifically for the VOEIS project
NOT WINDOWS BASED

Nothing against windows, but Voeis is not a windows-based software stack, so does not natively connect to HIS
Native connection from VOEIS to SQLServer was tested but found lacking due to both performance and the increased software stack complexity
Plus: there are other benefits...
We are able to extend the HIS API with additional data formats (JSON, XML, CSV etc)
AUTHENTICATION

ability to define access without updating the sqlserver
The functionality is provided in the gateway
WHAT IS THE HIS GATEWAY?

- A service which provides REST access to a HIS server.

Ultimately, It is very simple
At its heart, the HIS gateway is just a REST API provided by a minimal standalone web service. This service grants REST API access to the HIS server.
WHAT IS REST?

REST is a way of making data available over the internet, mapping site URLs to data. REST is a very popular API for data sharing on the web.
FILE FORMATS

- Other file formats
  - URL/item.json
  - URL/item.xml

Append a file extension to the path to access other formats. It results in the same data, but the format that is desired.
ANATOMY OF THE HISGW

Ruby, Sinatra, DataMapper, Open Source
For those that don’t know, Ruby is...
Our group programs mainly in ruby, this is the primary reason for this choice
The java foundation of JRuby allows us to use the Java ODBC to connect to SQLServer
DATAMAPPER

- Ruby Object Relational Mapper
- Maps database elements to Ruby objects
- Define interaction with ODM on HIS
- http://datamapper.org

Side-effect of being a nice, readable definition of the ODM in Ruby
Example on the side of some (clipped) code for the “Variable” object
- there are defined properties and relationships with explicit fields
Sinatra is good for webservices in particular, since it is very lightweight. Rack is a popular (in the ruby community) middleware webserver api – there are many webservers that have rack-compatible plugins (nginx and apache) and can run any Rack application.
MAPS REST TO ODM

• Ultimately:
  • The HIS Gateway maps REST to ODM
  • Read/write operations are supported
  • Allows for authenticated access

Ultimately, the functionality
Database operations are Create, Read, Update, Delete
AUTHENTICATION

• Performed through API keys
• Using a key, a client can be allowed access to restricted operations
• The client sends new data in JSON

Currently authentication in production for writes is being tested and isn’t fully supported. We’re working on it, but don’t need that functionality immediately.
The current set of formats that are supported
This could be extended to include others.
EXAMPLE OF USAGE

• Go to a HIS Gateway URL
• Retrieve data values in JSON
• Retrieve one data element by query
Example of retrieving the whole variable name controlled vocabulary
ONE RECORD RETRIEVAL

http://voeis.msu.montana.edu:4000/variable_name_cvs.json?
term=Borehole%20log%20material%20classification

Get a particular term
FUTURE FUNCTIONALITY

- WaterML Compliance
- More robust authentication
  - possibly role-based authorization
- RQL querying
WHY USE HISGW?

• Software that can’t access SQLServer natively
  • Mobile apps
  • API Services for data retrieval/update
  • Automated systems
OPEN SOURCE SOFTWARE

Available at http://github.com/yogo/HISgw

Tuesday, June 21, 2011
THANK YOU

Any questions?