

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

DATA QUALITY?

- Data in a given HIS server is of unknown quality
- We presume that the data in HIS is accurate and well curated
 - this is not guaranteed!



Tuesday, June 21, 2011

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

Tuesday, June 21, 2011

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



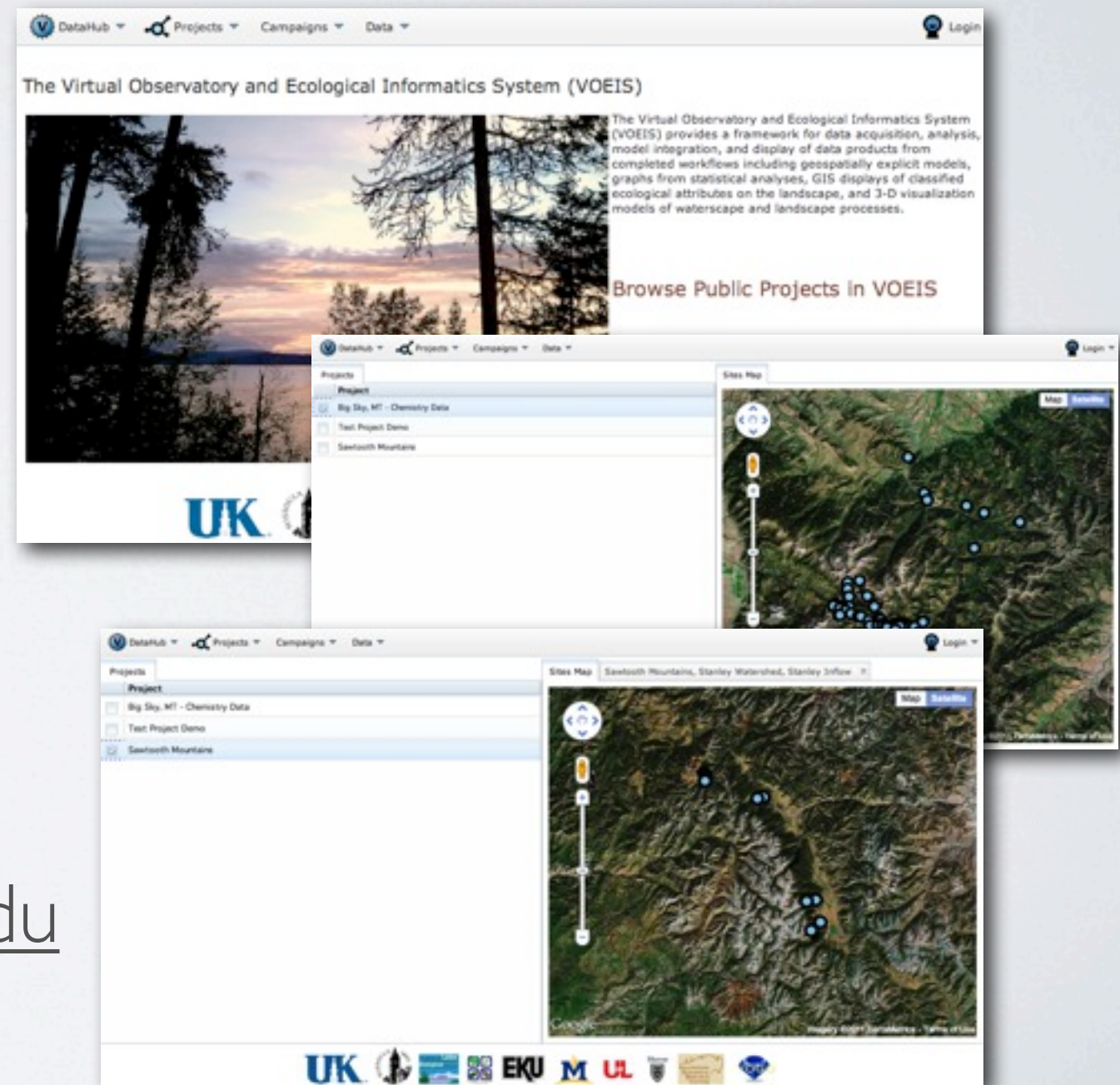


ENTERVOEIS

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>



Tuesday, June 21, 2011

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...

The screenshot displays the DataHub web application interface. The top navigation bar includes 'DataHub', 'Projects', 'Campaigns', and 'Data'. The 'Projects' tab is active, showing a list of projects: 'Big Sky, MT - Chemistry Data' (selected), 'Test Project Demo', and 'Sawtooth Mountains'. To the right, a 'Sites Map' shows a satellite view of a mountainous region with several blue location markers. Below the project list, the 'Big Sky, MT - Chemistry Data' project is expanded, revealing a table of sites. The table has columns for 'Site', 'State', 'Lat', and 'Lng'. The sites listed are all in Montana, with coordinates ranging from 45.2233 to 45.2381 latitude and -111.2374 to -111.349 longitude. A 'Sites Map' on the right shows the same satellite view with a cluster of blue markers. At the bottom of the interface, there is a row of logos for various institutions: UK, ECU, M, UL, and others.

Site	State	Lat	Lng
North Fork of the West Fork of the Gallatin River			
Rocky Mountains Southwestern Montana, Upper Gallatin River, Porcupine Creek	Montana	45.2233	-111.2374
Rocky Mountains Southwestern Montana, Upper Gallatin River, South Fork of the West Fork of the Gallatin River	Montana	45.2429	-111.3318
Rocky Mountains Southwestern Montana, Upper Gallatin River, South Fork of the West Fork of the Gallatin River	Montana	45.2384	-111.3455
Rocky Mountains Southwestern Montana, Upper Gallatin River, South Fork of the West Fork of the Gallatin River	Montana	45.2672	-111.2723
Rocky Mountains Southwestern Montana, Upper Gallatin River, South Fork of the West Fork of the West	Montana	45.2381	-111.349

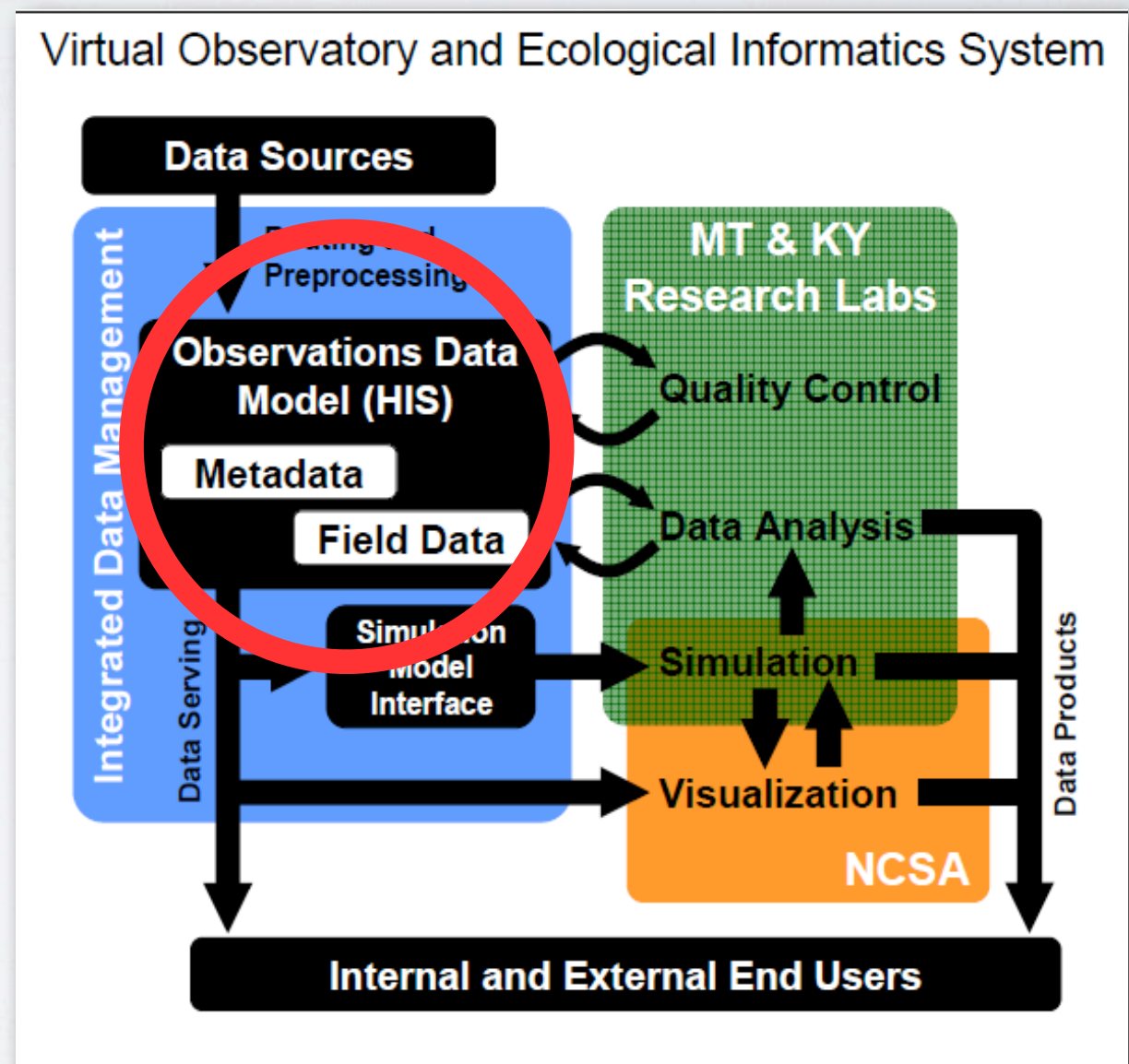
VOEIS ARCHITECTURE



- 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

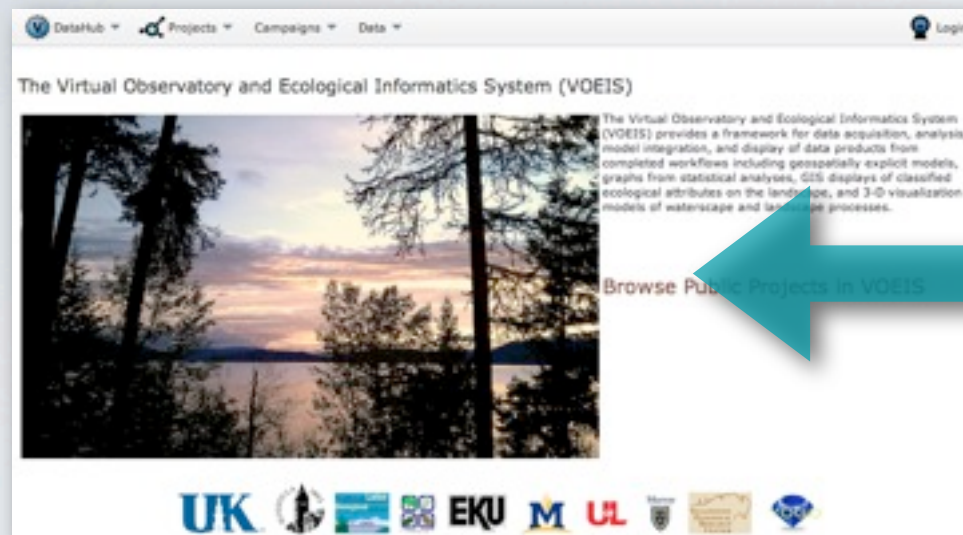


ENOUGH ABOUT VOEIS

- This presentation is not about VOEIS

BUT:

- VOEIS was the reason for the HIS Gateway



CUAHSI
HIS
Sharing hydrologic data

WHY BUILD A HIS GATEWAY?

Tuesday, June 21, 2011

the HIS Gateway was built specifically for the VOEIS project

NOT WINDOWS BASED



Tuesday, June 21, 2011

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...

ADDITIONAL FORMATS



<?xml?>

Tuesday, June 21, 2011

We are able to extend the HIS API with additional data formats (JSON, XML, CSV etc)

AUTHENTICATION



Tuesday, June 21, 2011

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.



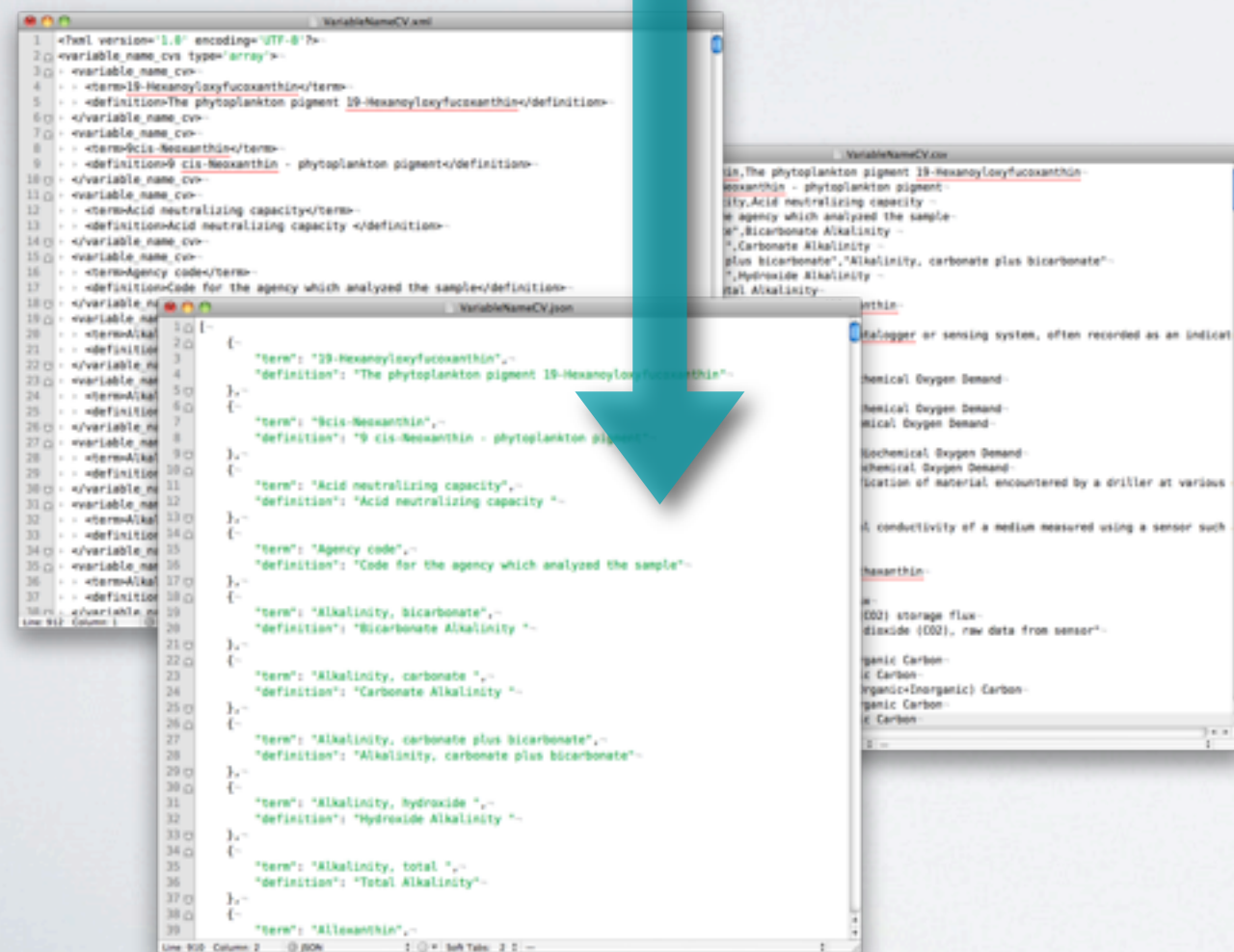
Tuesday, June 21, 2011

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?

https://hisgw.local/data_item



- Representational State Transfer
- Map HTTP URLs to data
- Very popular web API

Tuesday, June 21, 2011

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



```
1 {
2   {
3     "term": "19-Hexanoyloxyfucoxanthin",
4     "definition": "The phytoplankton pigment 19-Hexanoyloxyfucoxanthin"
5   },
6   {
7     "term": "9cis-Neoxanthin",
8     "definition": "9 cis-Neoxanthin - phytoplankton pigment"
9   },
10  {
11    "term": "Acid neutralizing capacity",
12    "definition": "Acid neutralizing capacity"
13  },
14  {
15    "term": "Agency code",
16    "definition": "Code for the agency which analyzed the sample"
17  },
18  {
19    "term": "Alkalinity, bicarbonate",
20    "definition": "Bicarbonate Alkalinity"
21  },
22  {
23    "term": "Alkalinity, carbonate",
24    "definition": "Carbonate Alkalinity"
25  },
26  {
27    "term": "Alkalinity, carbonate plus bicarbonate",
28    "definition": "Alkalinity, carbonate plus bicarbonate"
29  },
30  {
31    "term": "Alkalinity, hydroxide",
32    "definition": "Hydroxide Alkalinity"
33  },
34  {
35    "term": "Alkalinity, total",
36    "definition": "Total Alkalinity"
37  },
38  {
39    "term": "Alloxanthin",
40    "definition": "Alloxanthin"
41  }
42 }
```

<?xml?>

```
1 <?xml version='1.0' encoding='UTF-8'?>
2 <variable name_cvs type='array'>
3   <variable name_cvs>
4     <term>19-Hexanoyloxyfucoxanthin</term>
5     <definition>The phytoplankton pigment 19-Hexanoyloxyfucoxanthin</definition>
6   </variable name_cvs>
7   <variable name_cvs>
8     <term>9cis-Neoxanthin</term>
9     <definition>9 cis-Neoxanthin - phytoplankton pigment</definition>
10  </variable name_cvs>
11  <variable name_cvs>
12    <term>Acid neutralizing capacity</term>
13    <definition>Acid neutralizing capacity</definition>
14  </variable name_cvs>
15  <variable name_cvs>
16    <term>Agency code</term>
17    <definition>Code for the agency which analyzed the sample</definition>
18  </variable name_cvs>
19  <variable name_cvs>
20    <term>Alkalinity, bicarbonate</term>
21    <definition>Bicarbonate Alkalinity</definition>
22  </variable name_cvs>
23  <variable name_cvs>
24    <term>Alkalinity, carbonate</term>
25    <definition>Carbonate Alkalinity</definition>
26  </variable name_cvs>
27  <variable name_cvs>
28    <term>Alkalinity, carbonate plus bicarbonate</term>
29    <definition>Alkalinity, carbonate plus bicarbonate</definition>
30  </variable name_cvs>
31  <variable name_cvs>
32    <term>Alkalinity, hydroxide</term>
33    <definition>Hydroxide Alkalinity</definition>
34  </variable name_cvs>
35  <variable name_cvs>
36    <term>Alkalinity, total</term>
37    <definition>Total Alkalinity</definition>
38  </variable name_cvs>
39  <variable name_cvs>
40    <term>Alloxanthin</term>
41    <definition>Alloxanthin</definition>
42  </variable name_cvs>
43 </variable name_cvs>
```

Tuesday, June 21, 2011

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



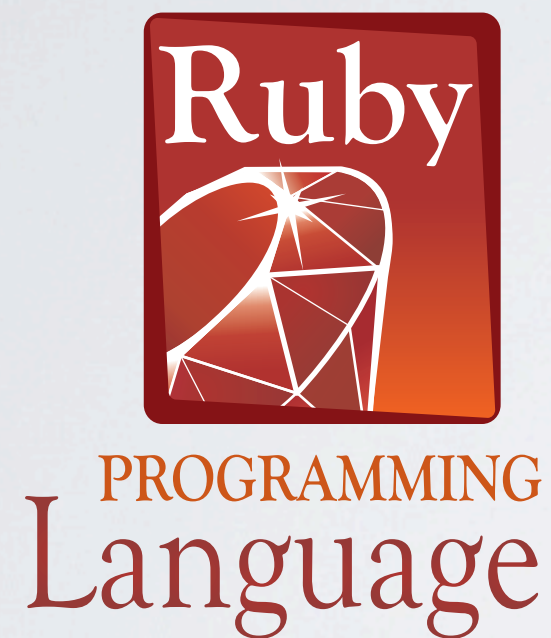
PROGRAMMING
Language



ANATOMY OF THE HISGW

Ruby, Sinatra, DataMapper, Open Source

RUBY



JRuby

- Ruby is a high level object-oriented programming language
- HIS Gateway uses JRuby
 - Java implementation of the Ruby VM
- <http://ruby-lang.org>
- <http://jruby.org>

Tuesday, June 21, 2011

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>

```
1 class Variable
2   include DataMapper::Resource
3
4   storage_names[:default] = "Variables"
5
6   property :id,          Serial, :required => true, :field => "VariableID",
7   property :variable_code, String, :required => true, :field => "VariableCode",
8   property :variable_name, String, :required => true, :field => "VariableName",
9   property :speciation,   String, :required => true, :field => "Speciation", :default =>
10  property :variable_units_id, Integer, :required => true, :field => "VariableUnitsID",
11  property :sample_medium, String, :required => true, :field => "SampleMedium", :default =>
12  property :value_type,    String, :required => true, :field => "ValueType", :default =>
13  property :is_regular,    Integer, :required => true, :field => "IsRegular", :default =>
14  property :time_support,  Float, :required => true, :field => "TimeSupport",
15  property :time_units_id, Integer, :required => true, :field => "TimeUnitsID", :default =>
16  property :data_type,     String, :required => true, :field => "DataType", :default =>
17  property :general_category, String, :required => true, :field => "GeneralCategory", :default =>
18  property :no_data_value, Float, :required => true, :field => "NoDataValue", :default =>
19
20  has n, :Categories, :model => "Categorie",
21  belongs_to :units, :model => "Unit", :child_key => [:time_units_id]
22  belongs_to :data_type_cv, :model => "DataTypeCV", :child_key => [:data_type]
23  belongs_to :general_category_cv, :model => "GeneralCategoryCV", :child_key => [:general_category]
24  belongs_to :sample_medium_cv, :model => "SampleMediumCV", :child_key => [:sample_medium]
25  belongs_to :value_type_cv, :model => "ValueTypeCV", :child_key => [:value_type]
26  belongs_to :variable_name_cv, :model => "VariableNameCV", :child_key => [:variable_name]
27  belongs_to :speciation_cv, :model => "SpeciationCV", :child_key => [:speciation]
28  belongs_to :data_values, :model => "DataValue", :child_key => [:variable_id]
29
30  validates_uniqueness_of :variable_code
31 end
```

Tuesday, June 21, 2011

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



- Model-View-Controller Micro-framework
 - Rack Compatible
 - Very simple to build and customize
 - <http://sinatrarb.com>

Tuesday, June 21, 2011

Sinatra is good for webservice in particular, since it is very lightweight.

Rack is a popular (in the ruby community) middleware webserver api

– there are many webserver that have rack-compatible plugins (nginx and apache) and can run any Rack application

MAPS REST TO ODM

REST



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

Tuesday, June 21, 2011

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



Tuesday, June 21, 2011

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.

SUPPORTED DATA FORMATS



<?xml?>



- JSON
- XML
- CSV
- YAML

Tuesday, June 21, 2011

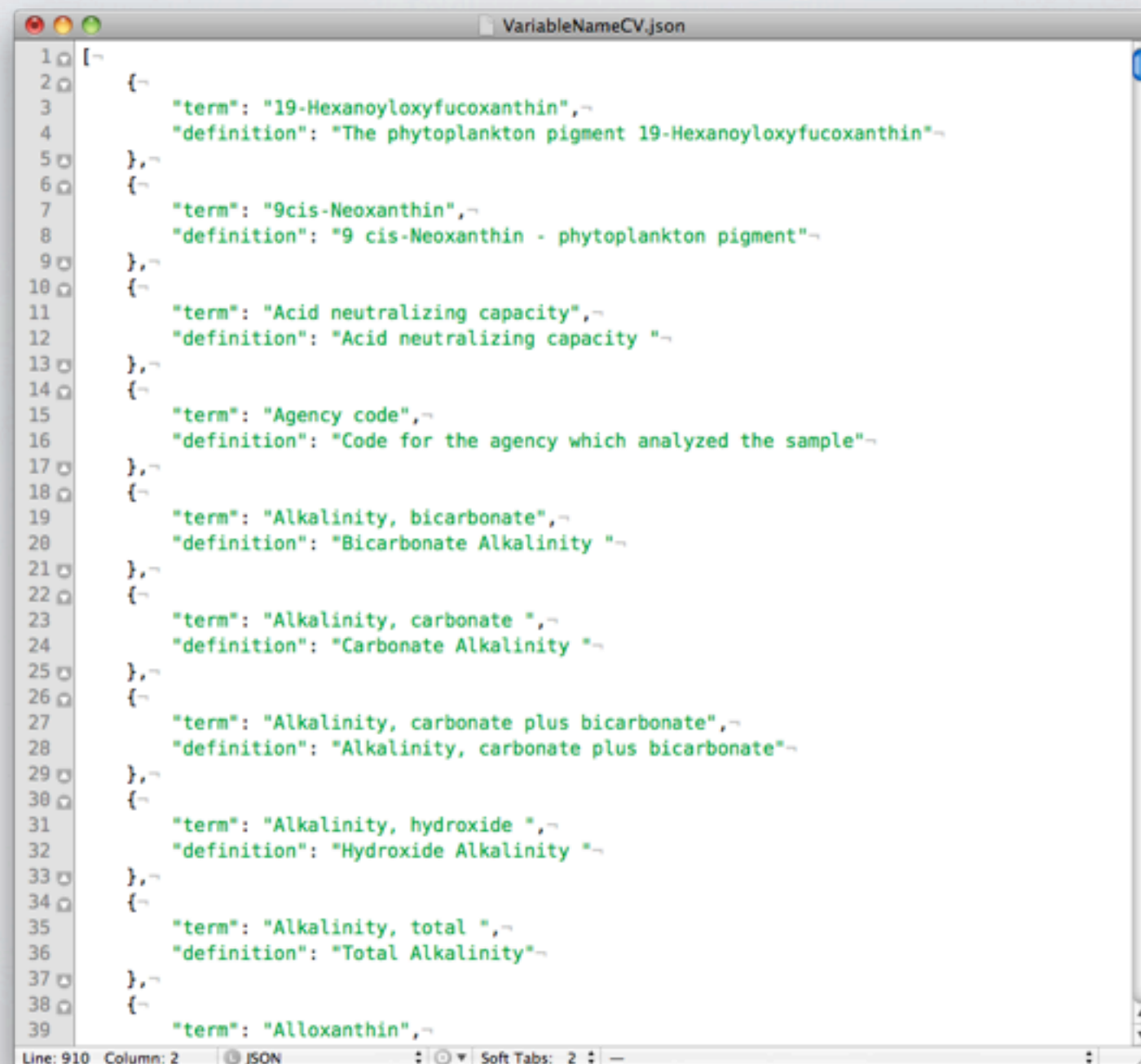
The current set of formats that are supported
This could be extended to include others.

EXAMPLE OF USAGE

- Go to a HIS Gateway URL
- Retrieve data values in JSON
- Retrieve one data element by query

FULL RETRIEVAL

https://hisgw.server/variable_name_cvs.json



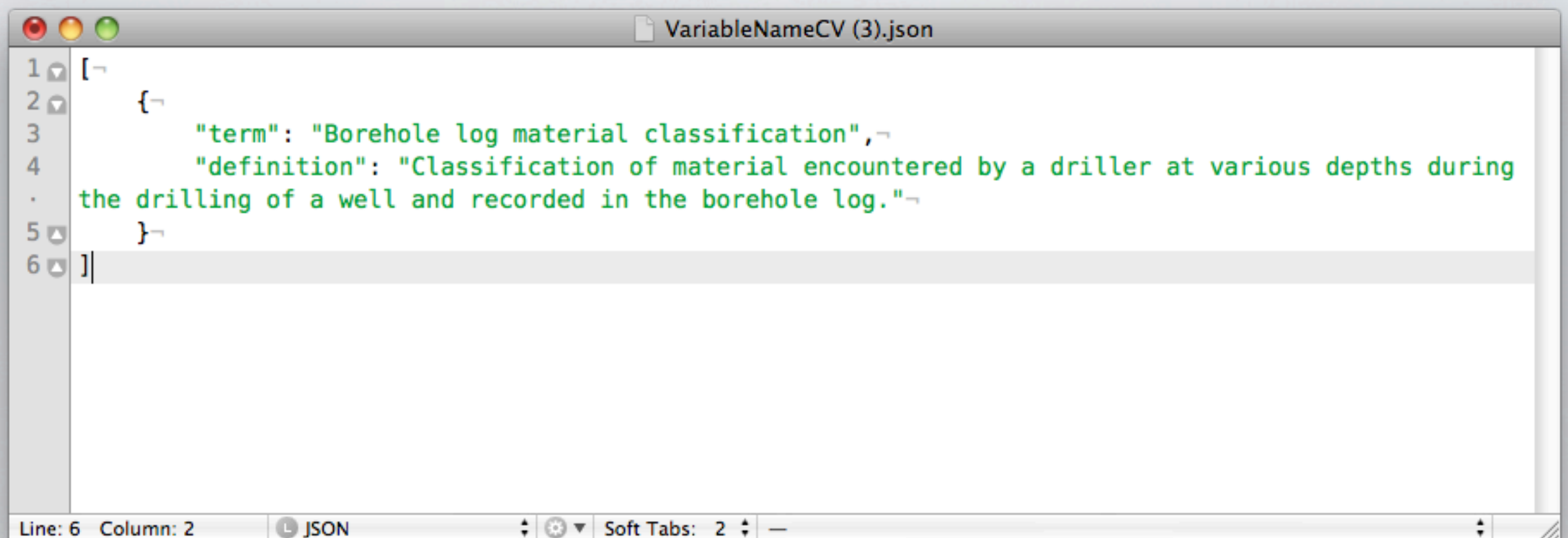
```
1  [~
2  {~
3    "term": "19-Hexanoyloxyfucoxanthin",~
4    "definition": "The phytoplankton pigment 19-Hexanoyloxyfucoxanthin"~
5  },~
6  {~
7    "term": "9cis-Neoxanthin",~
8    "definition": "9 cis-Neoxanthin - phytoplankton pigment"~
9  },~
10 {~
11   "term": "Acid neutralizing capacity",~
12   "definition": "Acid neutralizing capacity "~
13 },~
14 {~
15   "term": "Agency code",~
16   "definition": "Code for the agency which analyzed the sample"~
17 },~
18 {~
19   "term": "Alkalinity, bicarbonate",~
20   "definition": "Bicarbonate Alkalinity "~
21 },~
22 {~
23   "term": "Alkalinity, carbonate ",~
24   "definition": "Carbonate Alkalinity "~
25 },~
26 {~
27   "term": "Alkalinity, carbonate plus bicarbonate",~
28   "definition": "Alkalinity, carbonate plus bicarbonate"~
29 },~
30 {~
31   "term": "Alkalinity, hydroxide ",~
32   "definition": "Hydroxide Alkalinity "~
33 },~
34 {~
35   "term": "Alkalinity, total ",~
36   "definition": "Total Alkalinity"~
37 },~
38 {~
39   "term": "Alloxanthin",~
```

Tuesday, June 21, 2011

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



```
1 [
2   {
3     "term": "Borehole log material classification",
4     "definition": "Classification of material encountered by a driller at various depths during
· the drilling of a well and recorded in the borehole log."
5   }
6 ]
```

Line: 6 Column: 2 JSON Soft Tabs: 2

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

github
SOCIAL CODING

pol Dashboard Inbox 0 Account Settings Log Out

Explore GitHub Gist Blog Help Search...

yogo / HISgw

Admin Watch Fork Pull Request 1 1

Source Commits Network Pull Requests (0) Fork Queue Issues (0) Wiki (0) Graphs Branch: master

Switch Branches (1) Switch Tags (0) Branch List

click here to add a description
click here to add a homepage

Downloads

SSH HTTP Git Read-Only git@github.com:yogo/HISgw.git Read+Write access


Use the proper tags when returning XML

robbielamb (author)
April 08, 2011

commit 9f71cd38be03e13bd9ba
tree 9369ebb7f91643ab3723
parent 511b2cf2c642777f6691

HISgw /

name	age	message
app/	April 08, 2011	Use the proper tags when returning XML [robbielamb]
bin/	April 06, 2011	The app can now run in production. [robbielamb]
config/	April 06, 2011	Configuring datamapper logging. [robbielamb]
public/	September 21, 2010	Updated repo name, made favicon work. [irjudson]
gitignore	September 21, 2010	fixing requires to have the patch work [irjudson]


open source

OPEN SOURCE SOFTWARE

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



CUAHSI
HIS
Sharing hydrologic data



THANK YOU

Any questions?