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 <u>does not</u> 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!

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

ENTER VOEIS

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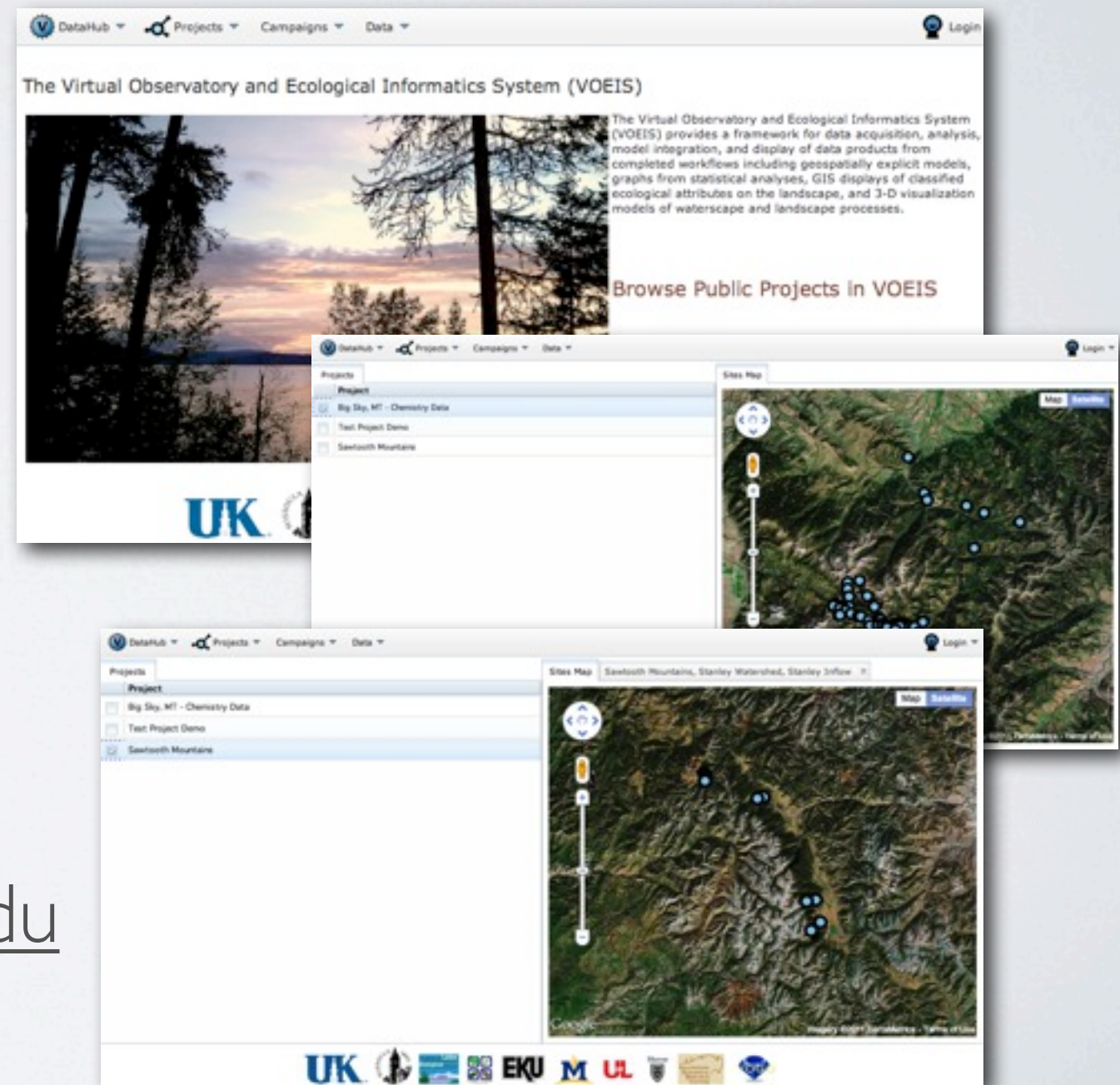
Open source data management and publication software
Designed to store and organize a variety of ecological data
The system supports and interfaces with CUAHSI HIS

PROJECT BASED DATA

- Organized into Sites

- Versioned data

- QA/QC
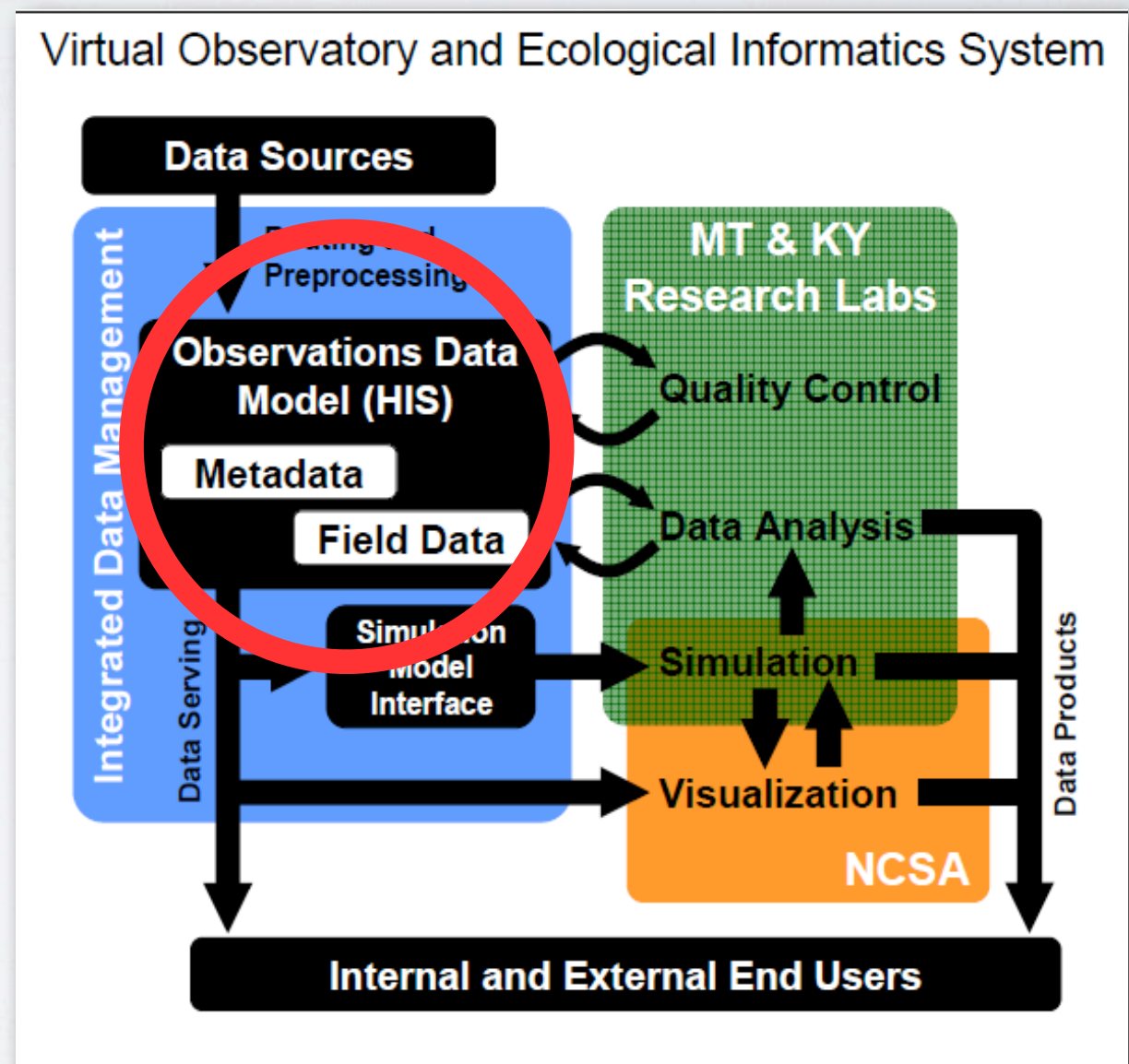
- Data is public/private until...

VOEIS ARCHITECTURE

- Cross Platform

- Yogo Framework Application

  - Database Agnostic

  - Versioned Data

- All Open Source

VOIES USES HIS

- Push data to HIS for:

  - Discovery
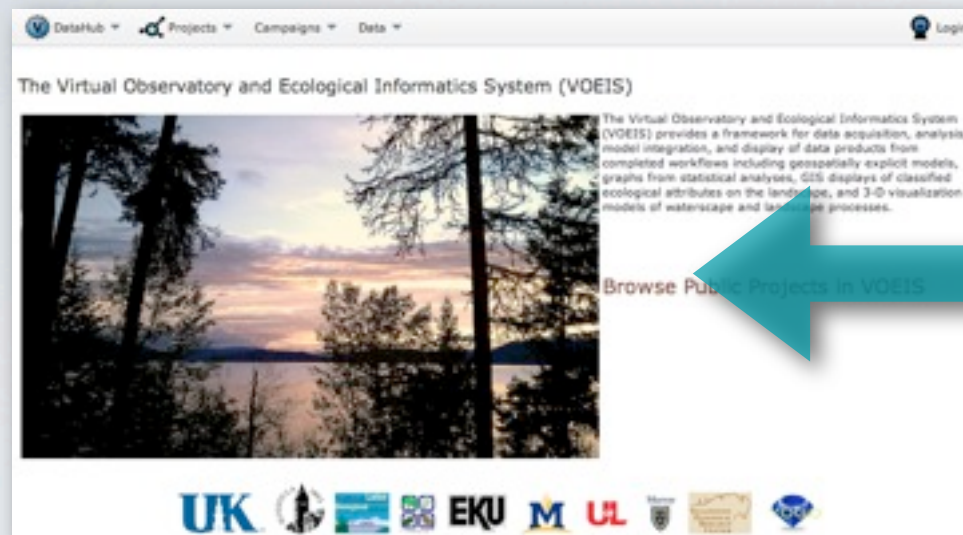
  - Publishing

  - Collaboration with the community

Virtual Observatory and Ecological Informatics System

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...

ADDITIONAL FORMATS

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?

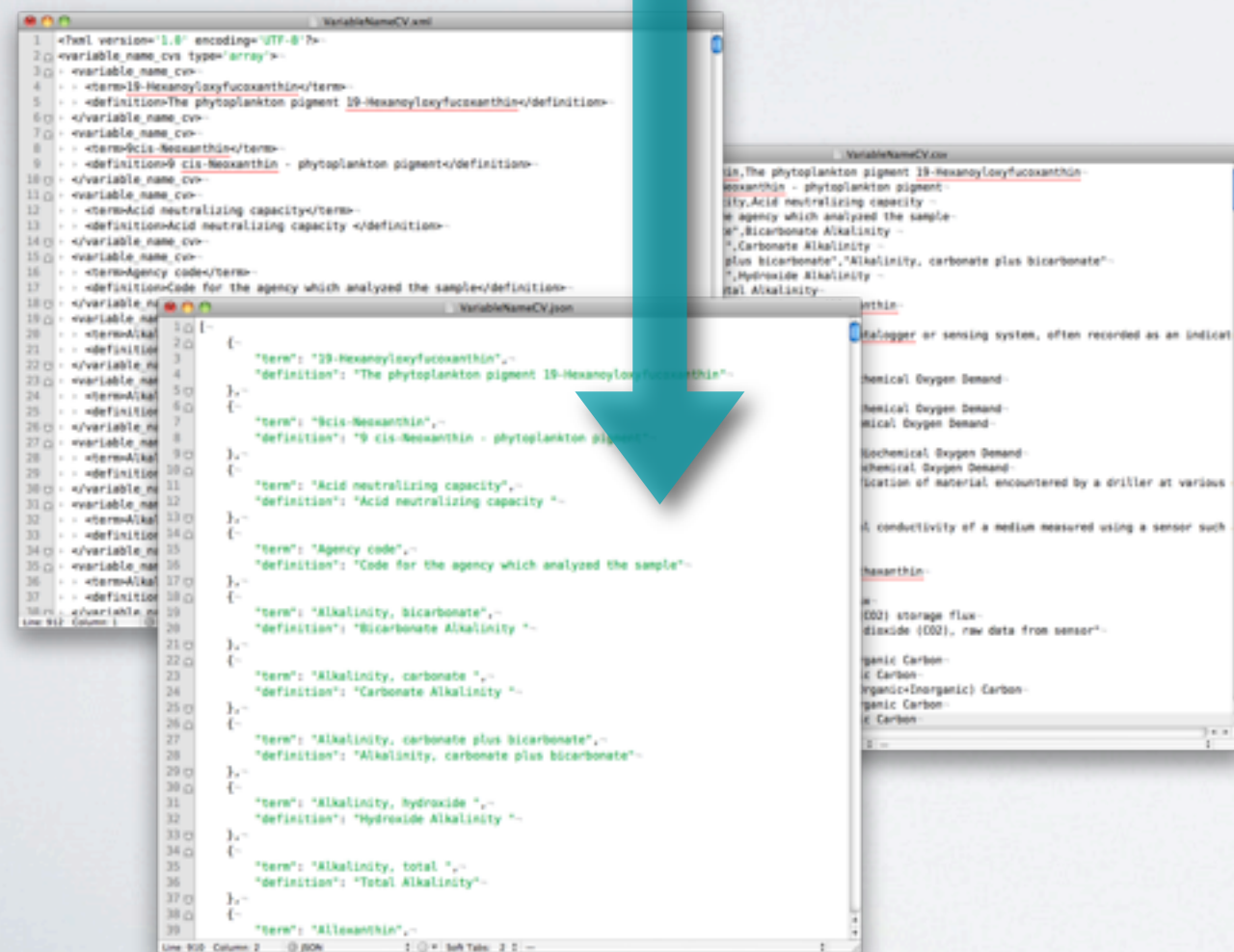- A service which provides REST access to a HIS server.

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?

https://hisgw.local/data_item

- Representational State Transfer

  - Map HTTP URLs to data

  - Very popular web API

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
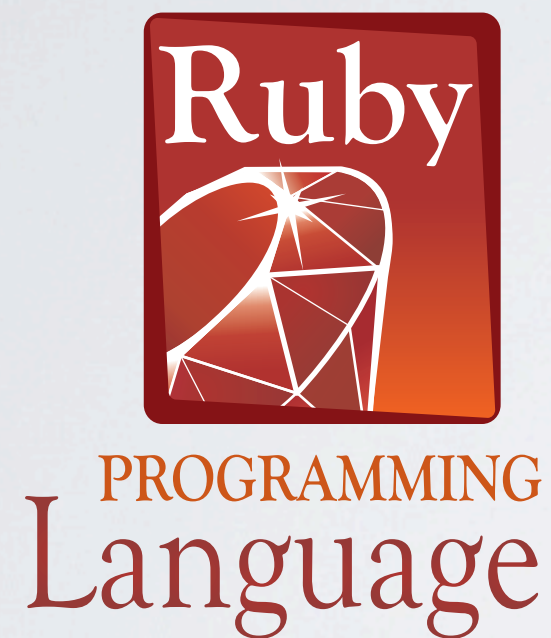
RUBY

- 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

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_categ
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
```

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

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

REST

- 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.

SUPPORTED DATA FORMATS
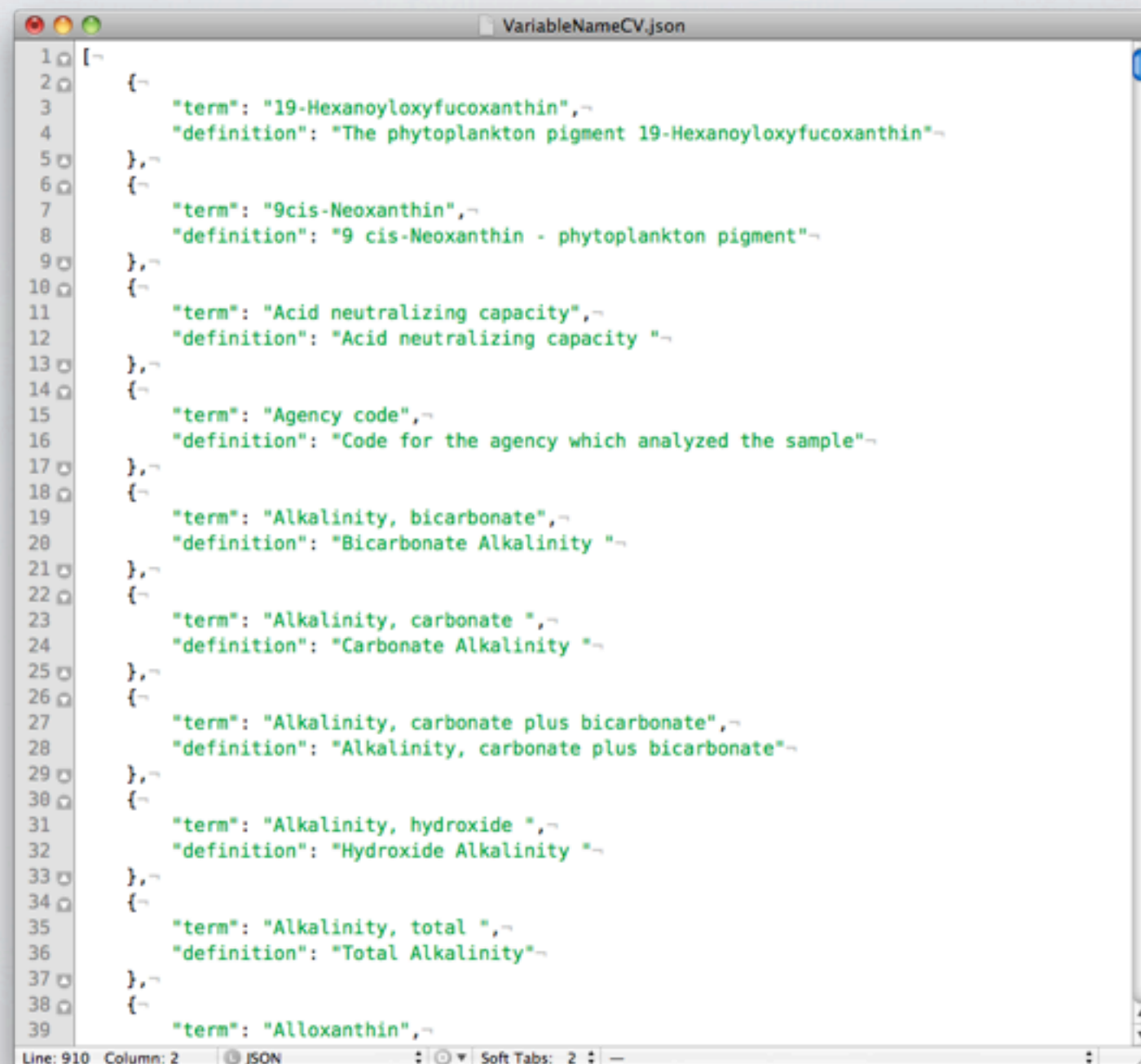
- JSON
- XML
- CSV
- YAML

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

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

VariableNameCV (3).json

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

Line: 6   Column: 2   JSON   Soft Tabs: 2

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

THANK YOU

Any questions?