

# 1. Intro to iSamples

ESIP Summer Meeting

7/14/2015

# iSamples Goals

An EarthCube RCN to...

“...dramatically improve the discovery, access, sharing, analysis, and curation **of physical samples** and the data generated by their study for the benefit of science and society.”

# More specifically...

It will build a new community of practice to advance digital access to samples through **socialization, knowledge creation, and best practices.**

- Socialization: **stakeholder alignment**
- knowledge creation: **iSamples Semantic Wiki & Registry of Earth Science Collections**
- best practices: **engagement of early career scientists** in order to transform sample and data practices by this community

# Working Groups

# iSamples WG1: Metadata & Identifiers

Reyna Jenkyns, David Arctur co-chairs

**Purpose:** This working group is concerned primarily with the scope, terminology, schema, and tools for **defining physical sample identifiers and metadata**, as well as addressing **semantics for interoperability** of data discovery, access & exchange.

# iSamples WG1: Metadata & Identifiers

## **Aims:**

1. Identify and evaluate relevant standards for identifiers & metadata
2. Identify, evaluate and associate relevant Identifiers
3. Identify and potentially rectify gaps with respect to physical samples.
4. Produce best practices and recommendations white paper based on findings from our evaluations above. We may also find redundancies and/or conflicts that need to be resolved.
5. Collect and share relevant publications using a reference management tool like Mendeley ([www.mendeley.com](http://www.mendeley.com)).
6. Engage communities at relevant conferences and workshops (ESIP, etc).

# WG 2

## User stories

**Chair Andrea Thomer**

**Co-Chair Sarah Ramdeen**

*We do not have a full picture of how people interact with samples, we don't have good life cycle based policies, and we don't know when/where different actors are needed (curators, investigators, etc.)*

# Goals

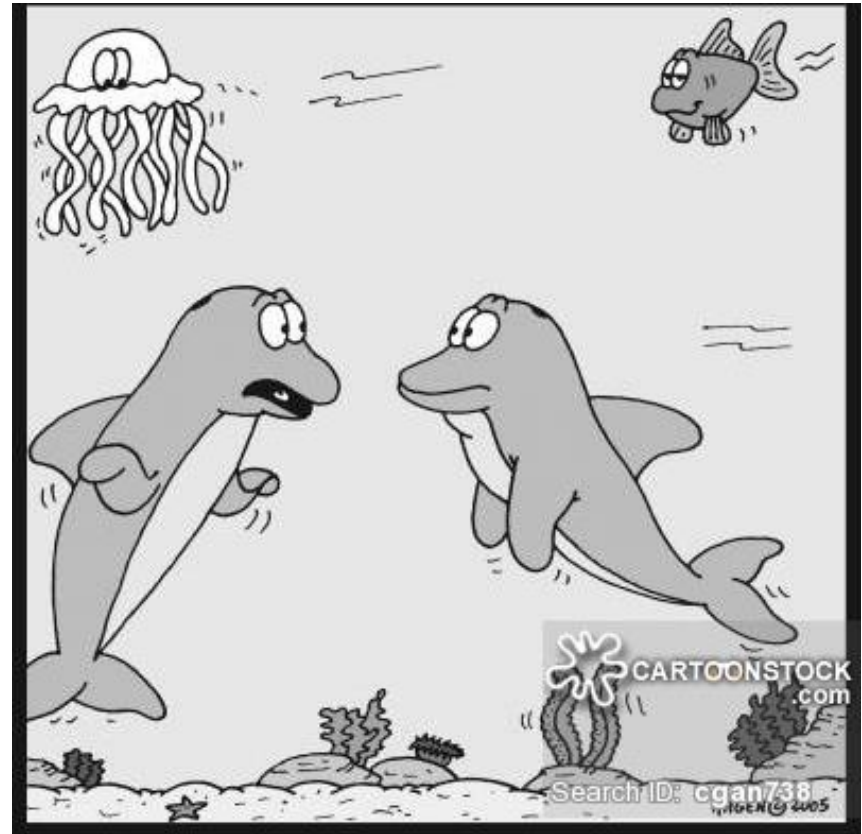
- Collect user stories
- Articulate different workflows and lifecycle practices for different users
- Identify potential points for curatorial intervention
- Develop policies/recommendations



# iSamples Working Group 3

## Communication, Education, Training

- **Co-chairs**
- Benjamin Hallett
- Ashlee Dere
  
- **Problem:** Insufficient appreciation of benefits of digital sample management



Frankly, I think we should stop trying to communicate with Humans: They're simply not smart enough to understand us

## WG3: Communication, Training, Education

### Goal:

- Develop communication plan to outline benefits and promulgate best practices in sample management

### Projects:

- Suite of 3 Pitch Slides
  - Draft slides
    - [https://docs.google.com/presentation/d/1hUgfgEGAkc9NbjtoIN55Inct6QjYEVkt3jW7YjzIH6Y/edit#slide=id.g984a80c0c\\_2\\_75](https://docs.google.com/presentation/d/1hUgfgEGAkc9NbjtoIN55Inct6QjYEVkt3jW7YjzIH6Y/edit#slide=id.g984a80c0c_2_75)
- Benefits of iSamples
- Ambassador Program

# Architecture, Workflows, and Citation of Samples

iSamples Working Group 4

# Objectives

## Workflows

- Describe workflows of scientists and curators
- Articulate Challenges faced in collecting, managing, studying, and citing earth samples

## Architecture

- Identify steps in earth sample workflows that will benefit from software support interventions

## Citation

- Develop mechanisms and best practices for the citation of earth samples

# WG5

## Physical Infrastructure, Digitization, Collections, Sample Rescue

*Co-Chairs* Ramona Walls and Leslie Hale

Problem: Loss of physical samples and the information (data/metadata) related to them.

# WG 5 Problems to Solve

## Objectives to Solve Problems:

- 1) Develop a set of best practices for digital archiving of sample data/metadata
- 2) Produce a system for national archive system for physical samples
- 3) Produce proposal for funding streams for a national archive system and systematic digitization of samples
- 4) Develop a set of criteria for what collections should be digitized first

## 2. Presentations

## 3. WG2 Use Case collection



# WG2

## Problem:

We don't have a complete understanding of how and when researchers, repository managers, and other stakeholders interact with earth samples.

Consequently, we don't have **lifecycle-based** policies or best practices for the curation of samples **and their associated data – particularly when it comes to metadata & identifier management**

**This is a huge obstacle to the integration and reuse of earth science data**

# Goals

- Collect use cases from broad range of stakeholders
  - E.g. field researchers, repository managers, **infrastructure or software developers**
  - Existing work:
    - Use scenarios from end-user workshop on geochem and petrology
    - Geobiology example from Site-Based Data Curation project

# Goals

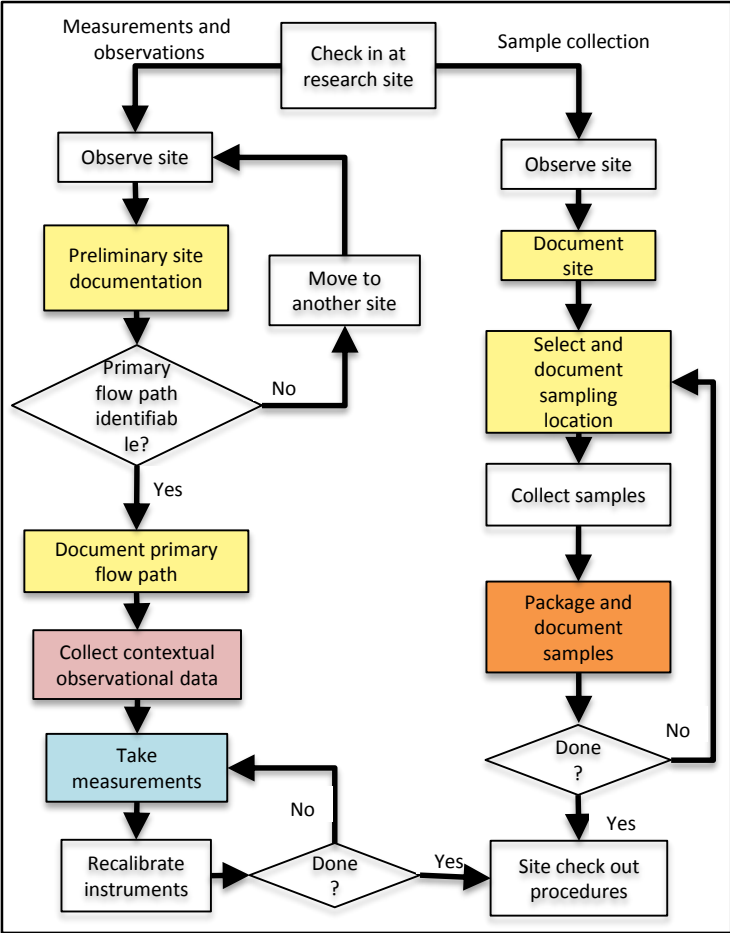
- Describe/diagram **common workflows** for different stakeholder groups
- Identify points for curatorial intervention
  - E.g. Where can IGSNs be assigned? When can we improve metadata quality?
- Develop recommendations for broader community

# Use case: **Geobiology research at Yellowstone National Park**

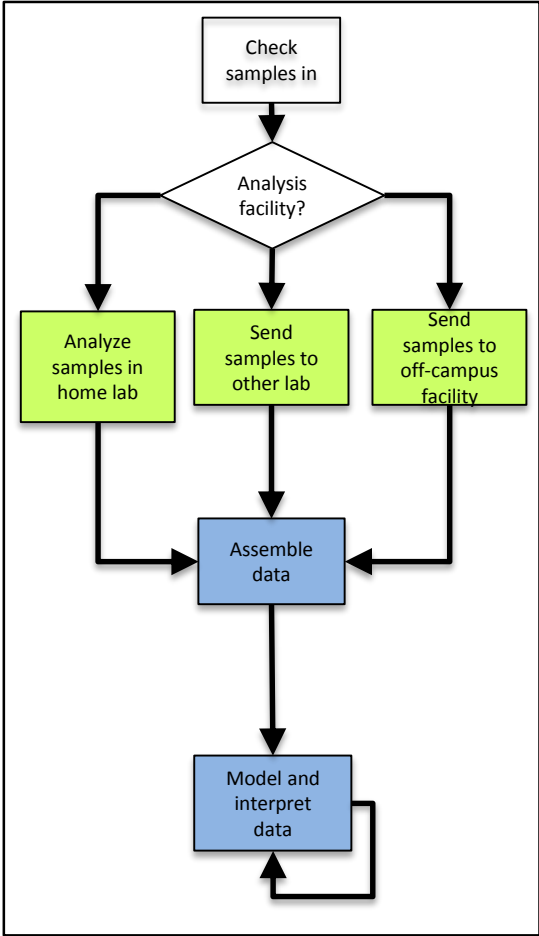
- Goals: Biological, physical and chemical data are extracted from samples out of the YNP hot springs – but disseminated through a broad range of publications and databases. We need to better link and contextualize these disparate data products.

# Use case: Geobiology research at Yellowstone National Park

## Fieldwork

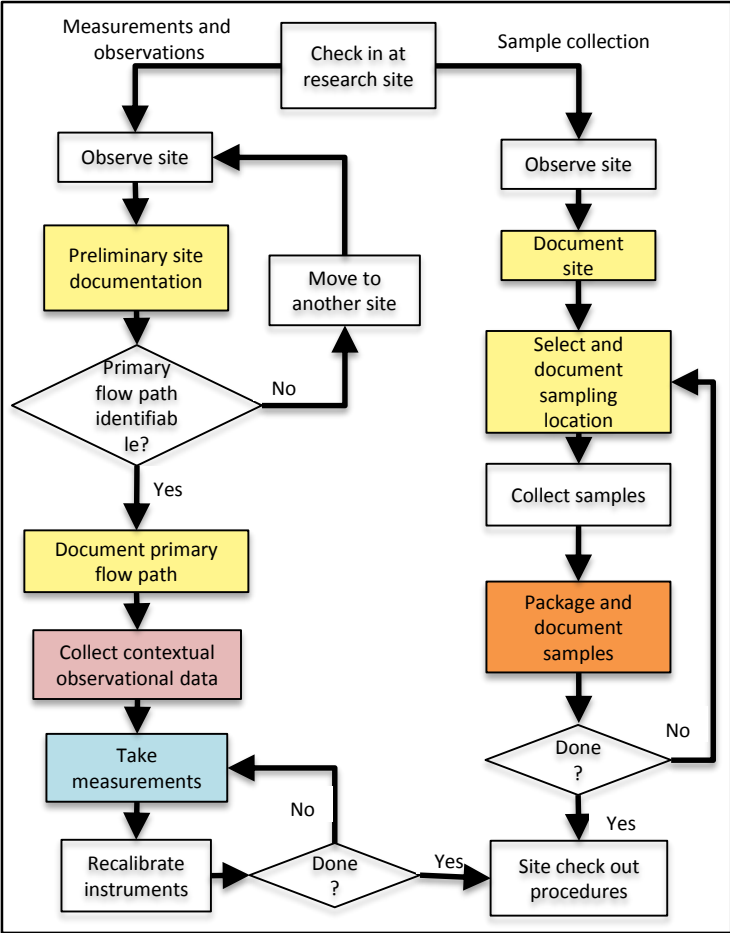


## Lab work

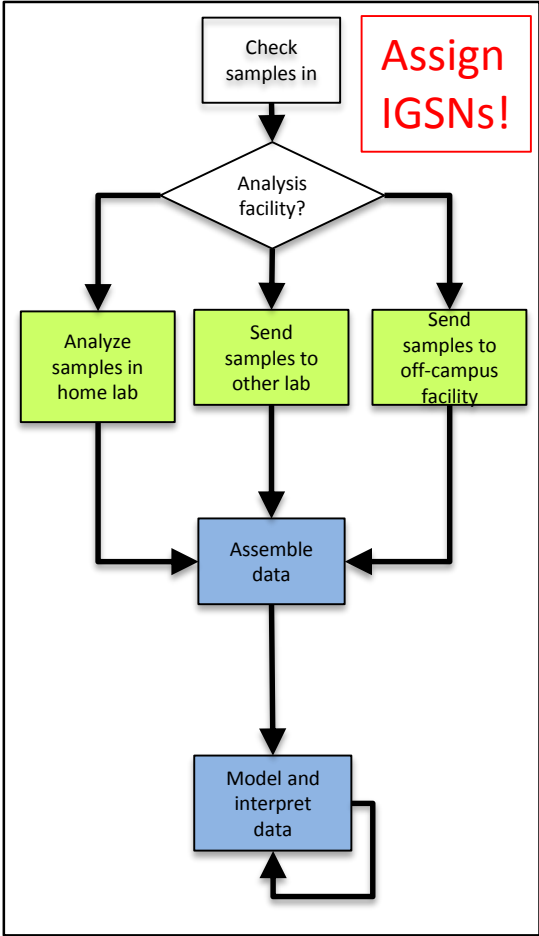


# Use case: Geobiology research at Yellowstone National Park

## Fieldwork



## Lab work



# Use case: **Geobiology research at Yellowstone National Park**

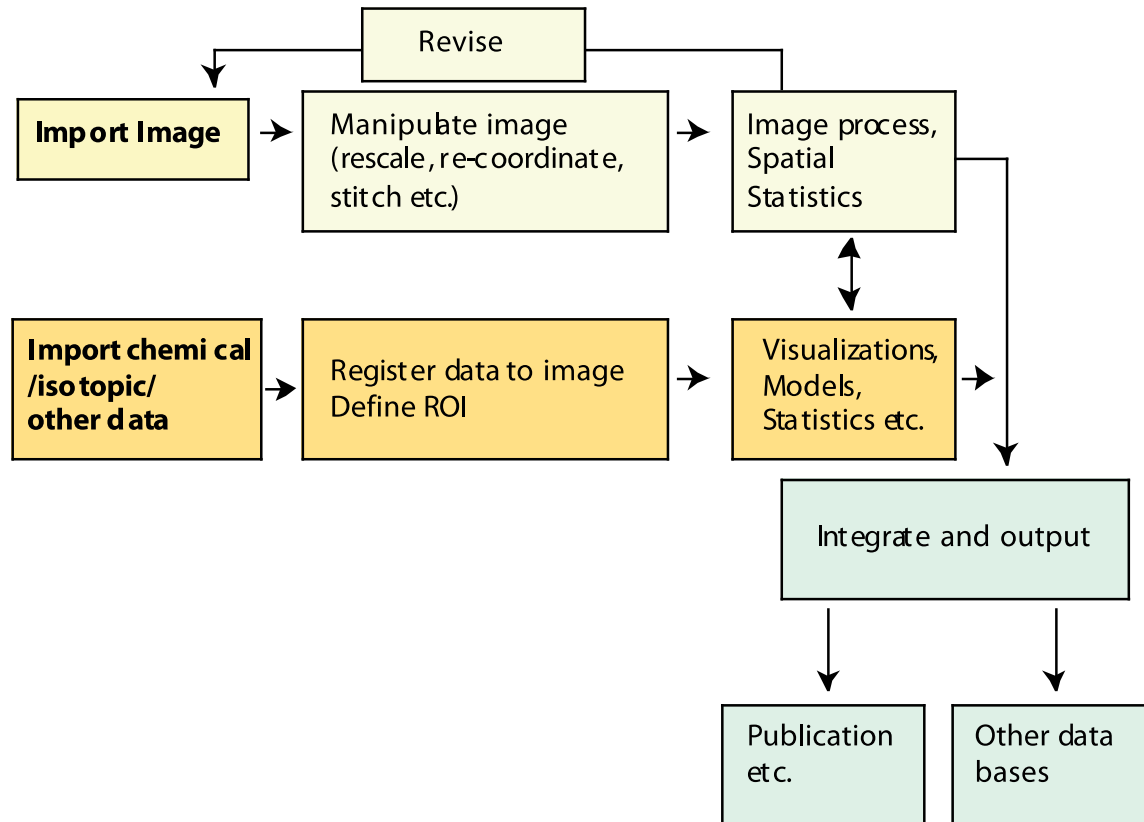
- Need to be able to “link” IGSNs with biological identifiers (GenBank accession numbers) – and vice versa
- IGSNs necessary in destructive sampling, gas and water samples

# Use case: “**Spatially resolved data at the scale of individual rock samples**”

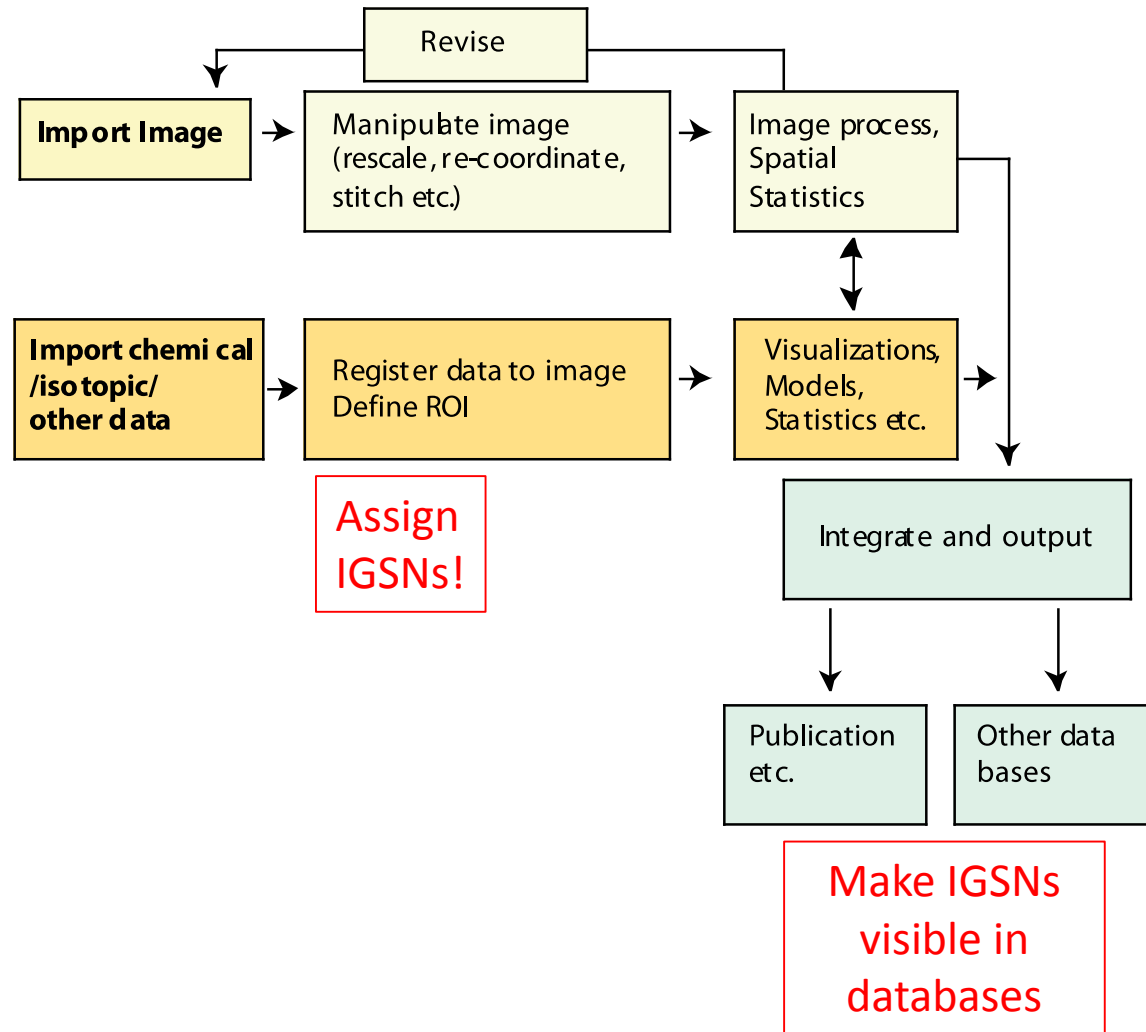
- Goal: To develop database and visualization resources to enhance analysis of spatially-resolved analysis of rock samples.
- Would require tools for
  - visualizing, processing and evaluating spatially-resolved data at the scale of individual samples,
  - image processing and spatial statistical analysis of images and associated data; and
  - integrating this information from other sources.



# Use case: “Spatially resolved data at the scale of individual rock samples”



# Use case: “Spatially resolved data at the scale of individual rock samples”



# Use case: “**Spatially resolved data at the scale of individual rock samples**”

- Sample imaging metadata needs to include IGSNs, spatial and scale information, additional image processing and “stitching” metadata
- Quantitative output from models must also be linked back to originating samples through IGSNs

# Use case: **New patterns in global volcanism from integration of physical eruptive data with the chemistry of erupted products**

- Goal: Mapping links btwn:
    - tectonic forcing functions (plate convergence rates, lithospheric structure, age) [**physical data**]
    - magma chemistry (volatile contents, major and trace element concentrations, isotopic ratios) [**chemical data**]
    - volcano behavior (explosive, effusive, emissions, deformation, repose intervals) [**historical data**]
- ...to improve hazards and risks models.

# Use case: New patterns in global volcanism from integration of physical eruptive data with the chemistry of erupted products

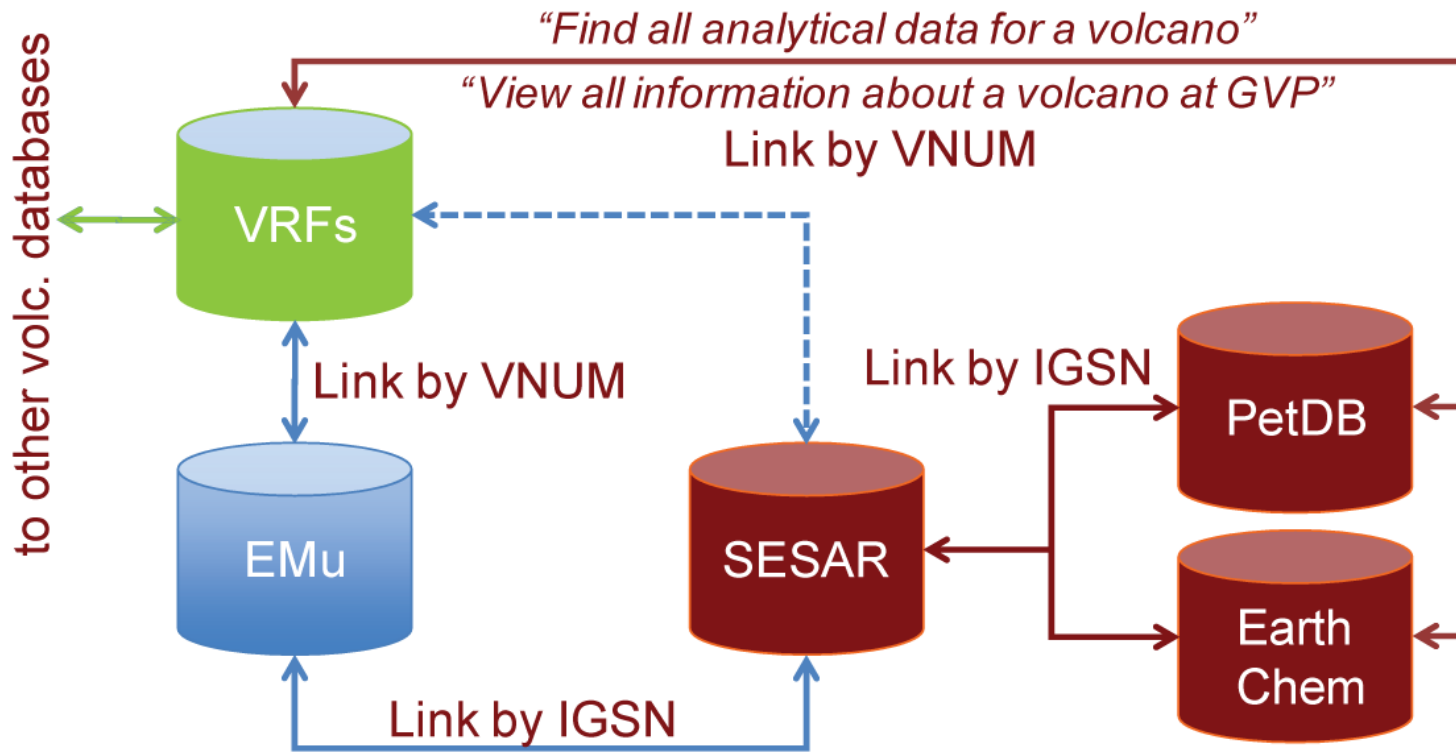


Figure 5. Proposed relationship between databases managed by NMNH (blue and green) and those managed by GfG (red).

## Use case: **New patterns in global volcanism from integration of physical eruptive data with the chemistry of erupted products**

- Can IGSNs be assigned through collections databases such as KE Emu? Specify? Others?
- What metadata about a sample do we need to support integration?
  - physical and chemical eruption-level information cannot be easily linked because the meta-data does not exist in IEDA databases (e.g. Earthchem).
  - “Time” could link physical and chemical data but timescales in Earthchem (e.g. C-14 date) are at a different scale than timescales in other DB’s (e.g. Global Volcanism Program DB = eruption date)

# iSamples + ESIP

- Are ESIP members' roles represented in use cases like these?
- If not: what other use cases?
  - Particularly, what other infrastructure use cases?

# Getting involved

- iSamples listserv:  
[http://earthcube.org/mailman/listinfo/isamples\\_earthcube.org](http://earthcube.org/mailman/listinfo/isamples_earthcube.org)