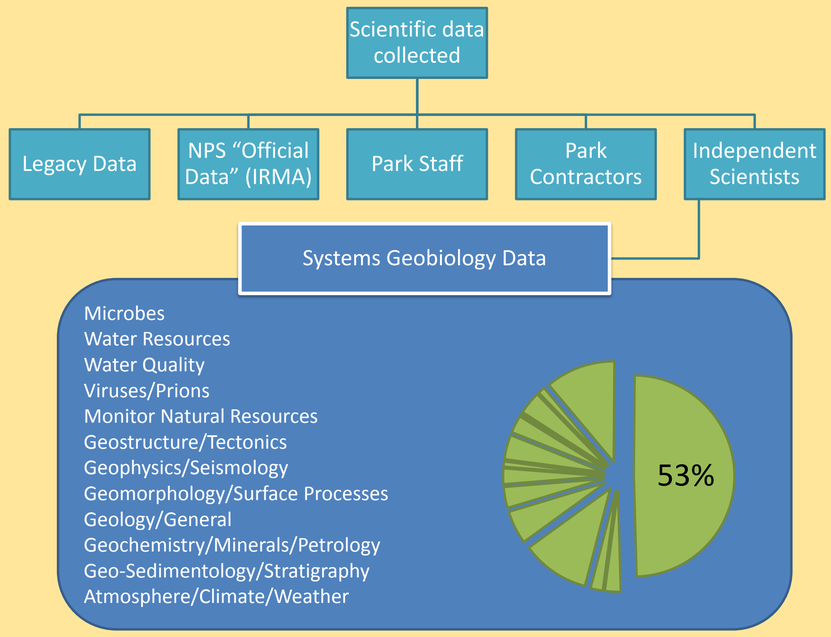
Site-Based Data Curation at Yellowstone National Park

Carole Palmer[1], Virgil Varvel[1], Andrea K. Thomer[1], Karen S. Baker[1], Karen M. Wickett[1], Bruce Fouke[2], Abigail Asangba[2], Ann Rodman[3], Sayeed Choudhury[4], and Tim DiLauro[4]

[1] *Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign*
[2] *Institute for Genomic Biology, University of Illinois at Urbana-Champaign*
[3] *Resource Information Management, Geographic Information Systems, Yellowstone National Park*
[4] *Digital Research and Curation Center at the Sheridan Libraries, Johns Hopkins University*

Center for Informatics Research in Science & Scholarship CIRSS

GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE
The iSchool at Illinois



Systems Geobiology within the context of YNP data

Stakeholder Workshop

A formative workshop with YNP data stakeholders will generate benchmarks for policy and curation guidelines for the collection, representation, sharing and quality assessment of datasets.

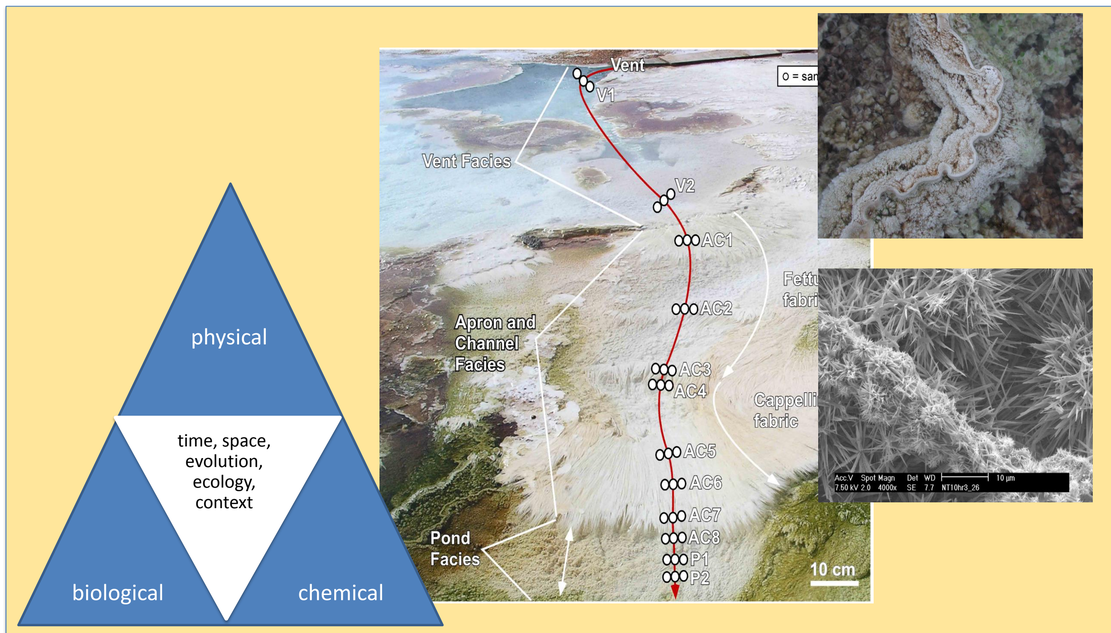
Workshop Activities:

- Pre-workshop questionnaire for participants
- Integrative science activity (shown above)
- Resource manager and scientist focus groups for informing data sharing priorities
- Break-out sessions to discuss guidelines for data reporting in YNP
- Post-workshop interviews

The Site-Based Data Curation (SBDC) project at the University of Illinois at Urbana-Champaign is developing a framework of policies and processes for the curation of research data generated at scientifically significant sites. The model will be built around the case of systems geobiology at Yellowstone National Park.

Project goals include:

- Articulating skills and principles for site-based curation
- Development of effective processes for transferring curated data into repositories for long-term preservation and access
- Exploring the inter-institutional relationships essential to site-based data curation



Researchers at Yellowstone collect data across multiple disciplines (e.g. geology, microbiology, chemistry) and multiple scales (microscopic to macroscopic). Teaching materials from PI Fouke's classes have been adapted for the stakeholder workshop. In the Integrative Science Exercise above, participants will be asked to map their data to different points of the triangle. This will help us understand how researchers studying different but related aspects of geobiology can better share and integrate their data.

Rock Deposits

Photography
modern and ancient

Macro-Scale (1-100+m)
geomorph and stratigraphy
Meso-Scale (1cm – 1m)
geomorph and stratigraphy
Micro-Scale (<1nm-1cm)
crystal size, shape
distribution, minerology
(aragonite, calcite)

Modeling
cellular automata
stochastic physics
time-lapse
particle velocity
crystallization rate
ecological

Water

Photography

Macro-,Meso-Scale
Temperature
alkalinity
pH
flow rate
water depth
major-minor elements
isotopes
gases

Modeling
saturation state
thermodynamic
mass balance
water/rock interaction
thermodynamic
mass balance

Microbe

Photography

Macro-, Meso-Scale
microbial mats
size, shape
pigment colors
gross morphology
DNA, RNA
Proteins
cell physiology
Viruses

Modeling
metagenomics
phylogenetic trees
metabolisms
pathway models

Climate

Photography

Macro-, Meso-Scale
temperature
tumidity
solar radiation
windspeed
rainfall

Modeling
GCM
predictions
reconstructions

Example of one scientist's range of data from a single kind of site.

Project Research Questions

Research questions addressing curation "upstream" in the research process:

- What data, and series of data, are most valuable to scientists and resource managers for long-term access?
- What parameters are most important for defining data? What aspects beyond geo-temporal are essential, such as sub-sites and their relationships?
- How should continuing series be curated and managed?

Research questions examining inter-institutional dynamics:

- What principles should underpin policies and processes of site-based curation, from various stakeholder perspectives?
- What repository expertise should inform site-based curation?
- What site expertise should inform repository operations?
- How can site and repository policies and processes be aligned?

INSTITUTE of Museum and Library SERVICES
IMLS National Leadership Grant LG-06-12-0706-12

DataConservancy