OUTCOMES OF THE “DATA CURATION FOR GEOBIOLOGY AT YELLOWSTONE NATIONAL PARK” WORKSHOP

Andrea K. Thomer¹, Carole L. Palmer², Bruce W. Fouke³, Ann Rodman³, G. Sayeed Choudhury, Karen A. Baker¹, Abigail E. Asangha⁴, Karen M. Wickett¹, Timothy DiLauro¹

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

Aim: Framework of policies and processes for curating of research data generated at scientifically significant sites

Site: Yellowstone National Park
- range of research, from origin of life on Earth to life on other planets
- unique environmental conditions
- permitting & reporting for points intervention for curation
- potential for collections of geo-located longitudinal data
- multiple data sources valuable for triangulation and context

Research Questions:
- What are optimal criteria for constructing & describing reusable sets of data?
- How should “series” of data be curated and managed?
- How can site and repository policies and processes be aligned?

Participants:
9 geologists, geochemists, and microbiologists
7 park personnel, representing research permitting and reporting, YNP research library and archive.

Activities:
- focus groups
- integrative science exercise
- roundtable discussions

Outcomes:
Field sampling processes: data reuse dependent on detailed, consistent records
Photos:
Key to organization of physical, biological, & chemical data from a sampling event
- reliable capture of sampling context & conditions
- ability to monitor site changes over time
Data sharing:
Should be understood as broader impact; moves focus from short-term impediments to long-term benefits.

Site

Stakeholder Group
Data archive developers and managers
YNP Resource Management
YNP Library & Archive professionals
Independent scientists from multiple disciplines

Interests
Data curation; repository policies and procedures
Protect park resources; oversee research access
Record park activities
Data collection; grand challenge research questions

Priorities
Improve data readiness for deposit
Adhere to NPS and federal policies; facilitate research in park
Preserve scholarly record
Data sharing and aggregation, tools to document data and context; stronger collaboration

Core metadata:
- Investigator name
- GPS location
- Altitude
- Date and time
- Site name
- Event ID

Other specialized metadata (i.e. pH, temperature, etc.)

Field photos become valuable organizational tools when linked to...

Outcomes:

YPN data in the aggregate:
Value for scientists
- Advance “big picture” research questions
- Identify next area of study, even beyond YNP
- Longitudinal analysis within & across features

Value for YNP resource managers:
- Coordination & transparency of data collection
- Visibility of trends & connections across projects

Next Steps:
- Vet and finalize minimal & optimal parameters for documenting sampling events
- Promotional materials for YNP researchers on data curation & broader impacts of data sharing
- Assess potential for points of intervention with current reporting and permitting processes
- Align with internal YNP processes for data management & coordination