

HOW WORKFLOW DOCUMENTATION FACILITATES CURATION PLANNING

Karen M. Wickett¹, Andrea K. Thomer¹, Karen Baker¹, Timothy DiLauro², Abigail Ezenam Asangba³
¹Center for Informatics Research in Science & Scholarship, Graduate School of Library & Information Science (GSLIS), University of Illinois at Urbana-Champaign
²Sheridan Library, Johns Hopkins University
³Department of Geology, University of Illinois at Urbana-Champaign

THE SBDC PROJECT

The Site-Based Data Curation (SBDC) at Yellowstone National Park project is developing a framework of guidelines and processes for the curation of research data generated at scientifically significant sites. The project is a collaboration between data curation researchers, data archiving and system development experts, geologists and microbiologists from the Institute for Genomic Biology, and resource managers from Yellowstone National Park.

CURATION PLANNING

The team has documented the key processes involved in data collection during a single field trip to Yellowstone National Park. The diagram shows the process simplified and divided into three stages: Planning, Fieldwork, and Processing & Analysis.

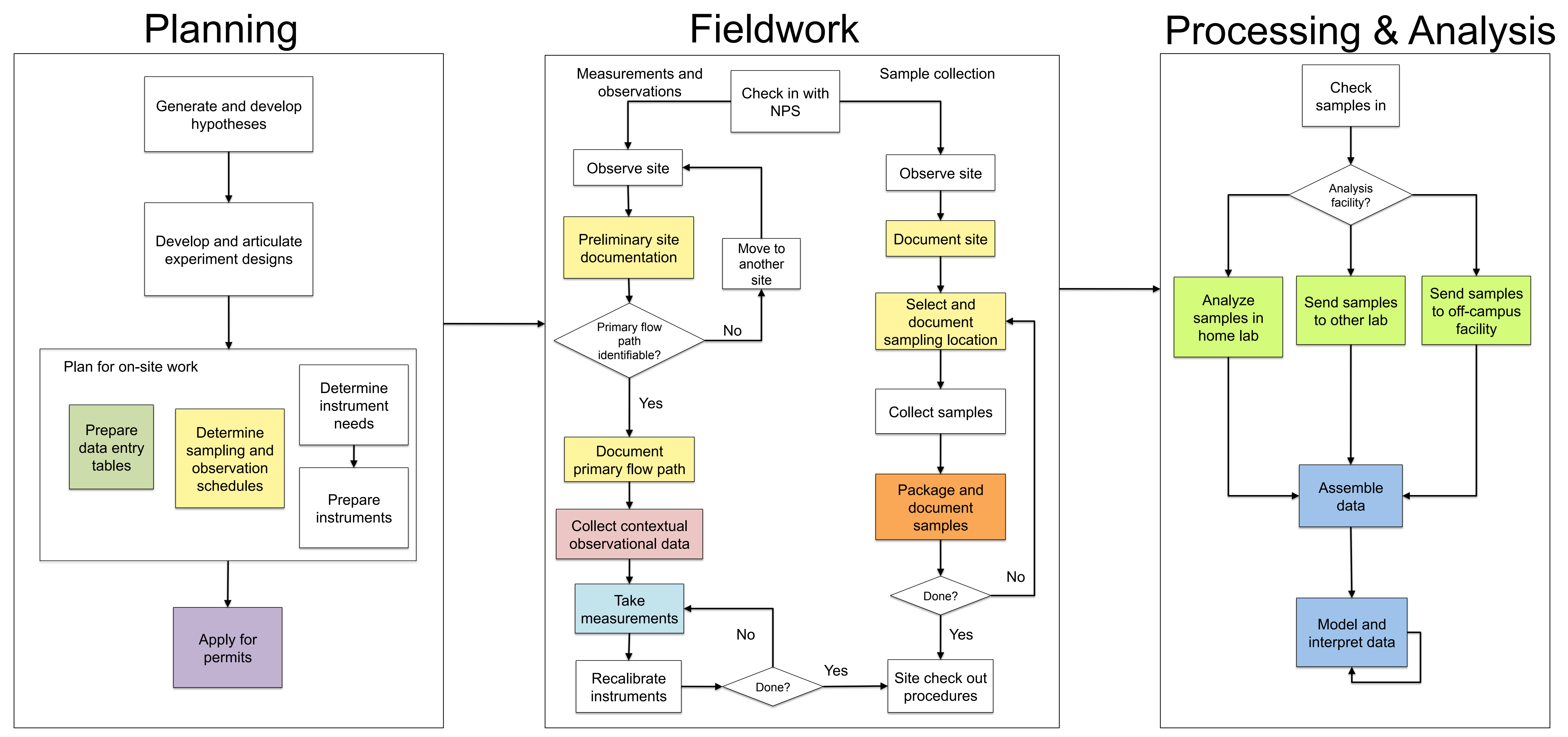
The workflow diagram can assist in curation planning by serving as a point of conversation between scientists and data managers to identify points of intervention for curation processes. For example, the transition points between each stage provide a minimal framework for data management interventions such as regularizing data formats, data description, or data quality and validation procedures.

The diagram shows key documents and data products that are generated during the scientific process. This approach draws attention to artifacts that are likely candidates for data release or deposit. Although some of these objects may be viewed by the researcher as intermediate or unfinished, with appropriate packaging and description they may serve as valuable observational data for other researchers.

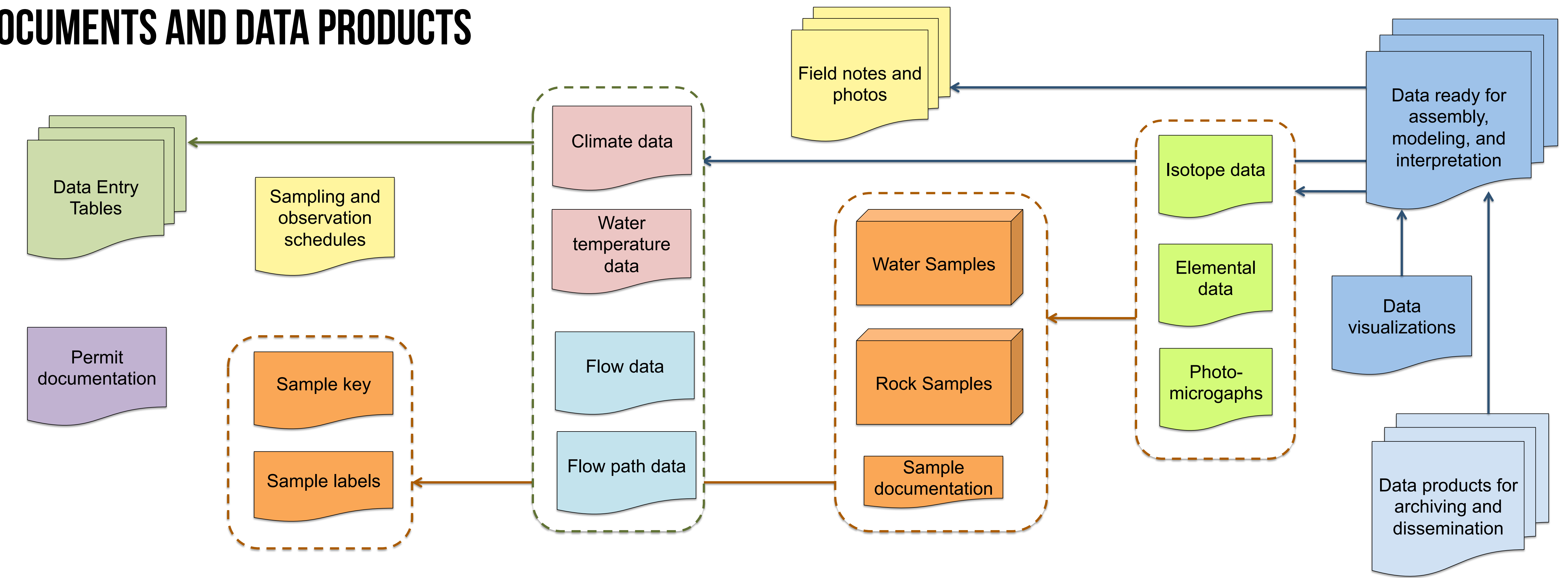
Documents and data products are linked to associated events in the process diagram (shown here with color coding). Detailed description of the methods and procedures used in the generation of data and collection of samples is essential for the re-use and interpretation of data. Workflow documentation supports a modular approach where commonly used protocols are described once and variations can be quickly noted during data generation.

Documents and data products are also linked back to objects that are significant in terms of the life-cycle of the resources. For example, the assembled data is linked to the component data. This allows for detailed tracking of the history of data products and support of validation and data quality procedures.

SIMPLIFIED WORKFLOW



KEY DOCUMENTS AND DATA PRODUCTS



Acknowledgments

The SBDC project is funded by Institute of Museum and Library Services National Leadership Grant LG-06-12-0706-12. Special thanks to Bruce Fouke and Gary Olsen.