

Abstract Submitted
for the DPP14 Meeting of
The American Physical Society

Documenting scientific workflow: the metadata, provenance and ontology project¹ MARTIN GREENWALD, J. STILLERMAN, J. WRIGHT, MIT - PSFC, G. ABLA, R. CHANTHAVONG, D. SCHISSEL, General Atomics, A. ROMOSAN, A. SHOSHANI, LBNL — Careful management of data, its creation and transformation (provenance) and associated metadata is a critical part of any scientific enterprise. Traditionally this was the role of the lab notebook, but the digital era has resulted instead in the fragmentation of data, processing and annotation. This paper describes an ongoing multi-institutional project aimed at remedying this problem by developing tools to automate documentation of scientific workflows and associated information. Data and all processes that create or modify that data are represented mathematically as a directed acyclic graph, providing explicit information about the relationships between elements with all elements having globally unique and persistent IDs. The export of data, for publication, presentation or external databases would be recorded, allowing traceability in either direction – answering the questions “Where was this data used?” or “Where did the data in this figure come from.” Namespace management is provided through a well structured “ontology,” which can be customized for any particular community or application.

¹Supported by DOE contract DE-SC0008736.

Martin Greenwald
MIT - PSFC

Date submitted: 10 Jul 2014

Electronic form version 1.4