Argonne NATIONAL LABORATORY

Planning for 2019 In Situ Data Management (ISDM) Workshop

February 2019

Washington DC area

Disclaimer: Tentative pending final approval

ASCR SSIO Workshop
September 20, 2018

Tom Peterka
tpeterka@mcs.anl.gov
Mathematics and Computer Science Division

Workshop Planning

Planning for 2019 In Situ Data Management (ISDM) Workshop

Organizing Committee

Name	Affiliation	Role
Tom Peterka	ANL	Chair
Debbie Bard	NERSC	Organizer
Janine Bennett	SNL	Organizer
Wes Bethel	LBNL	Organizer
Ron Oldfield	SNL	Organizer
Line Pouchard	BNL	Organizer
Christine Sweeney	LANL	Organizer
Matthew Wolf	ORNL	Organizer
Laura Biven	DOE-ASCR	Program Manager

Logistics

Date	Location
Feb. 2019	D.C. area

Abstract

This workshop seeks community input on the development of in situ capabilities for managing the execution and data flow among a wide variety of coordinated tasks for scientific computing.

The workshop considers ISDM beyond the traditional roles of accelerating simulation I/O and visualizing simulation results, to more broadly support future scientific computing needs. In particular, the convergence of simulation, data analysis, and artificial intelligence will require machine learning, data manipulation, creation of data products, assimilation of experimental and observational data, analysis across ensemble members, and, eventually the incorporation of tasks on non-von Neumann architecture.

Definition of In Situ

"In situ data management (ISDM) is the coordination of heterogeneous applications, executing simultaneously in an HPC system, cooperating toward a common objective." – ISDM planning document

Coordination:

Scheduling, launching, controlling, monitoring, communicating

Heterogeneous applications:

Parallel executables (large tasks), different authors, languages, programming/data models, runtimes

Simultaneous execution:

Part of the same "job" or scheduled concurrently, as opposed to post hoc (sometime later)

HPC system:

Co-located machines (supercomputer + vis cluster), accelerators, coprocessors in the same facility

Common objective:

Intentional, not coincidental coordination

Strict in situ ("in place"), in transit, online, run-time, co-processing, time partitioning, space partitioning $\subseteq$ in situ.
Postprocessing, post hoc, distributed area $\not\subseteq$ in situ.

Crosscutting Themes

- Uses beyond visualization and I/O
  - E.g., multiphysics, ensembles, EOD streaming data, AI, ML, Big Data, workflows
- Unconventional tasks, data, and hardware
  - E.g., finding correlations between simulations and experiments using deep neural networks on classical CPUs/GPUs accelerated by neuromorphic processors
- Modularity and interoperability
  - Within and outside of ISDM
- Software productivity and sustainability
  - Why isn't anyone using my software? Why do users roll their own code?
- Assumptions and dependencies (A&D) on other parts of the system
  - E.g., distributed workflow management, system software and services

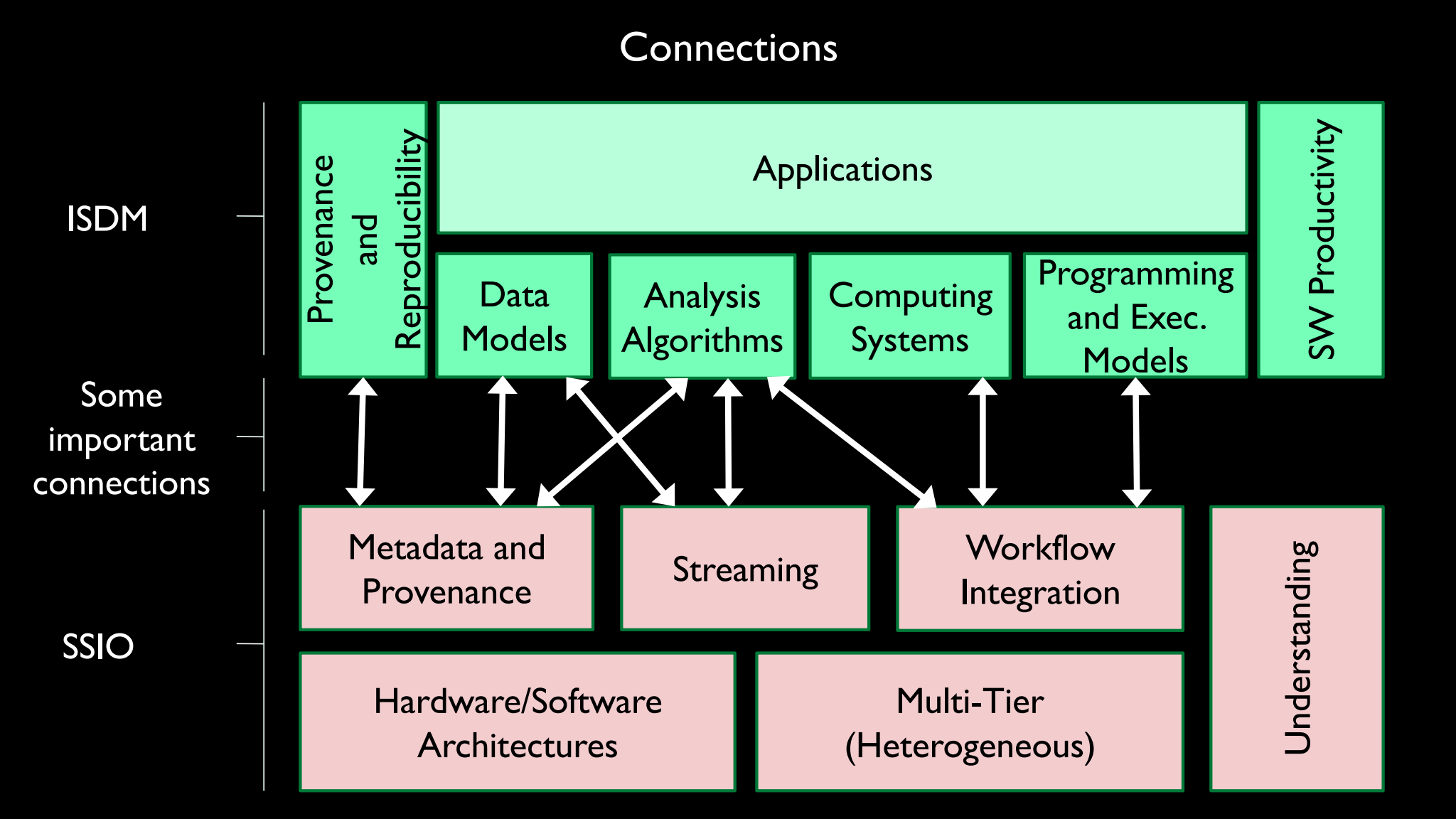
Topic Areas (= Workshop Sessions)

- Science applications
  - Interface to applications and science workflows given user needs and current solutions
- Computing systems
  - Interface to hardware and system software stacks and design of ISDM for future platforms
- Data models
  - Managing the structure and semantics of in situ data given disparate models of constituent data
- Programming and execution models
  - Programming and executing an ISDM framework given disparate models of constituent tasks
- Provenance and reproducibility
  - Collecting performance data and validating scientific data in order to achieve reproducibility
- Analysis algorithms
  - Developing portable, high-performance algorithms that can be used in situ and elsewhere
- Software productivity
  - Designing software that can be built, deployed, sustained, and used to support DOE science

SSIO Coordination

Observations

- Common drivers
  - DOE mission needs
  - Emerging systems
- SSIO
  - More system-focused
  - More about mechanisms (less about policy)
- ISDM
  - More application-focused
  - Influences policies that drive SSIO mechanisms
- Examples
  - SSIO provenance support might focus on system- and service-level information, while ISDM provenance support focuses on workflow and application data
  - ISDM data models might build on SSIO metadata capabilities (e.g., indexing, distributed data services)


Connections
ISDM
Some important connections
SSIO
Provenance and Reproducibility
Applications
SW Productivity
Data Models
Analysis Algorithms
Computing Systems
Programming and Exec. Models
Metadata and Provenance
Streaming
Workflow Integration
Understanding
Hardware/Software Architectures
Multi-Tier (Heterogeneous)

Acknowledgments

Organizing Committee

Debbie Bard	(NERSC)
Janine Bennett	(SNL)
Wes Bethel	(LBNL)
Ron Oldfield	(SNL)
Line Pouchard	(BNL)
Christine Sweeney	(LANL)
Matthew Wolf	(ORNL)

SSIO Coordination

Rob Ross

DOE Program Managers

Laura Biven

Tom Peterka
tpeterka@mcs.anl.gov
Mathematics and Computer Science Division