Workflow Project Overview

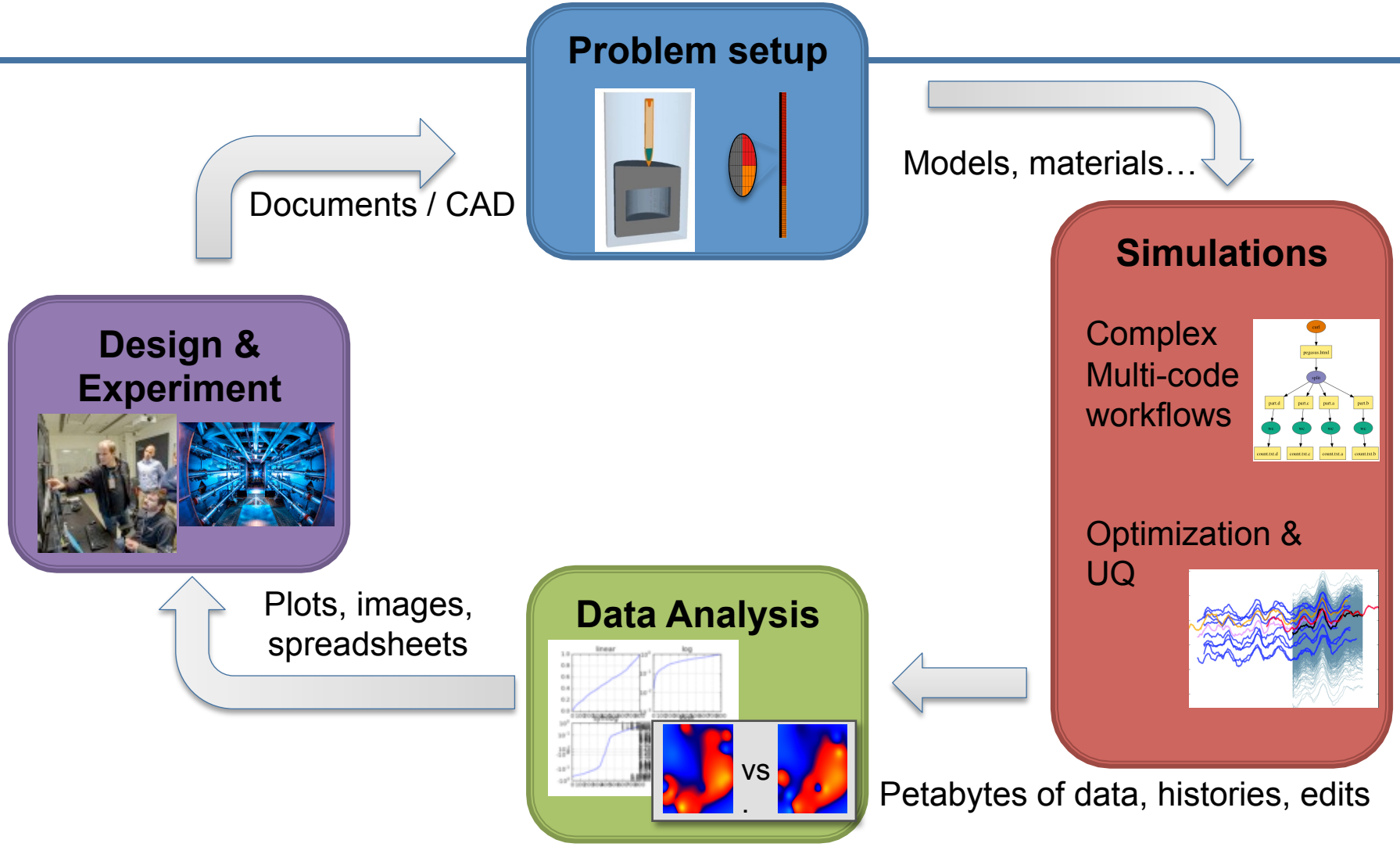
Daniel Laney
Workflow Project Lead

February 25, 2016

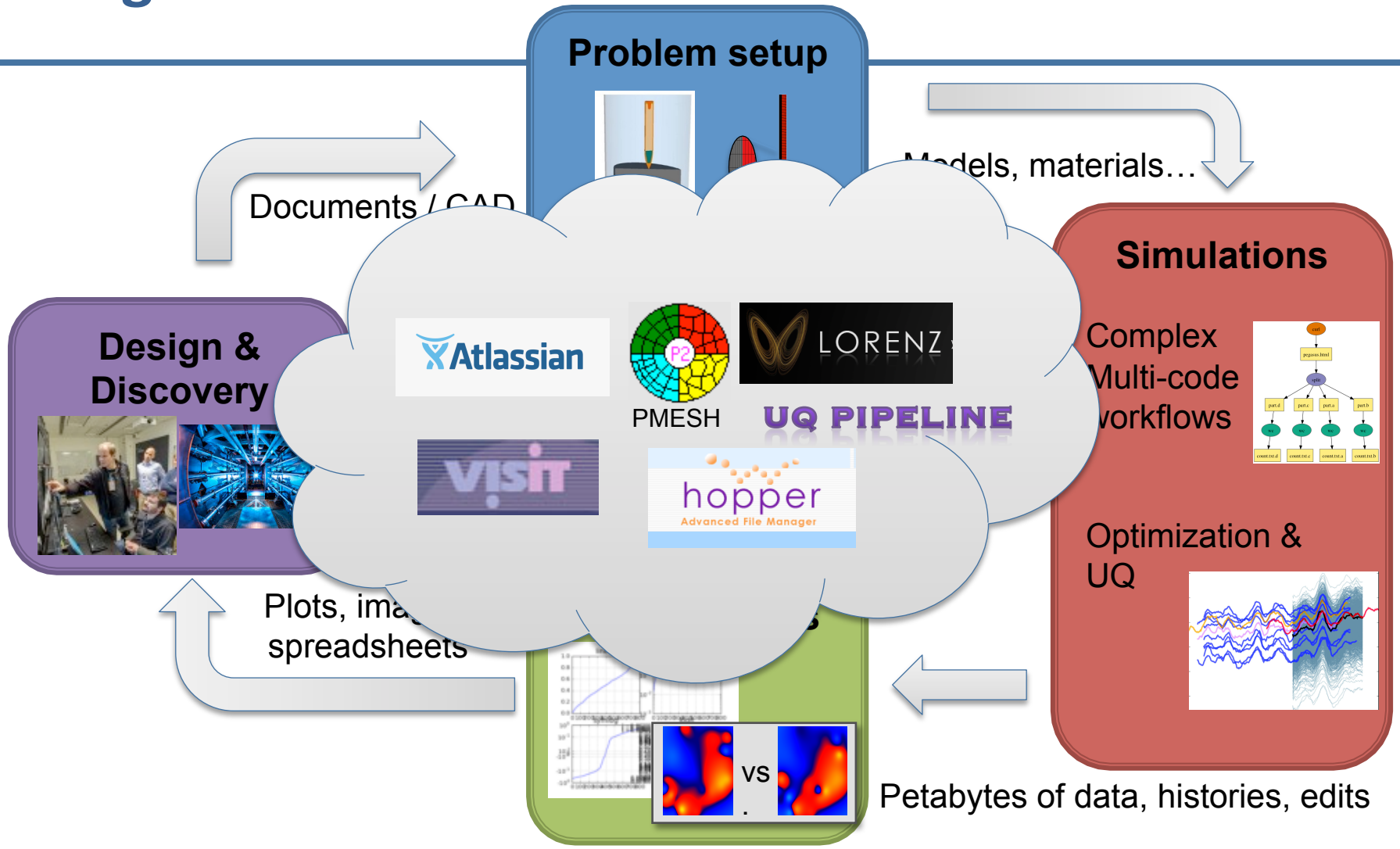
LLNL-PRES-684861 This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

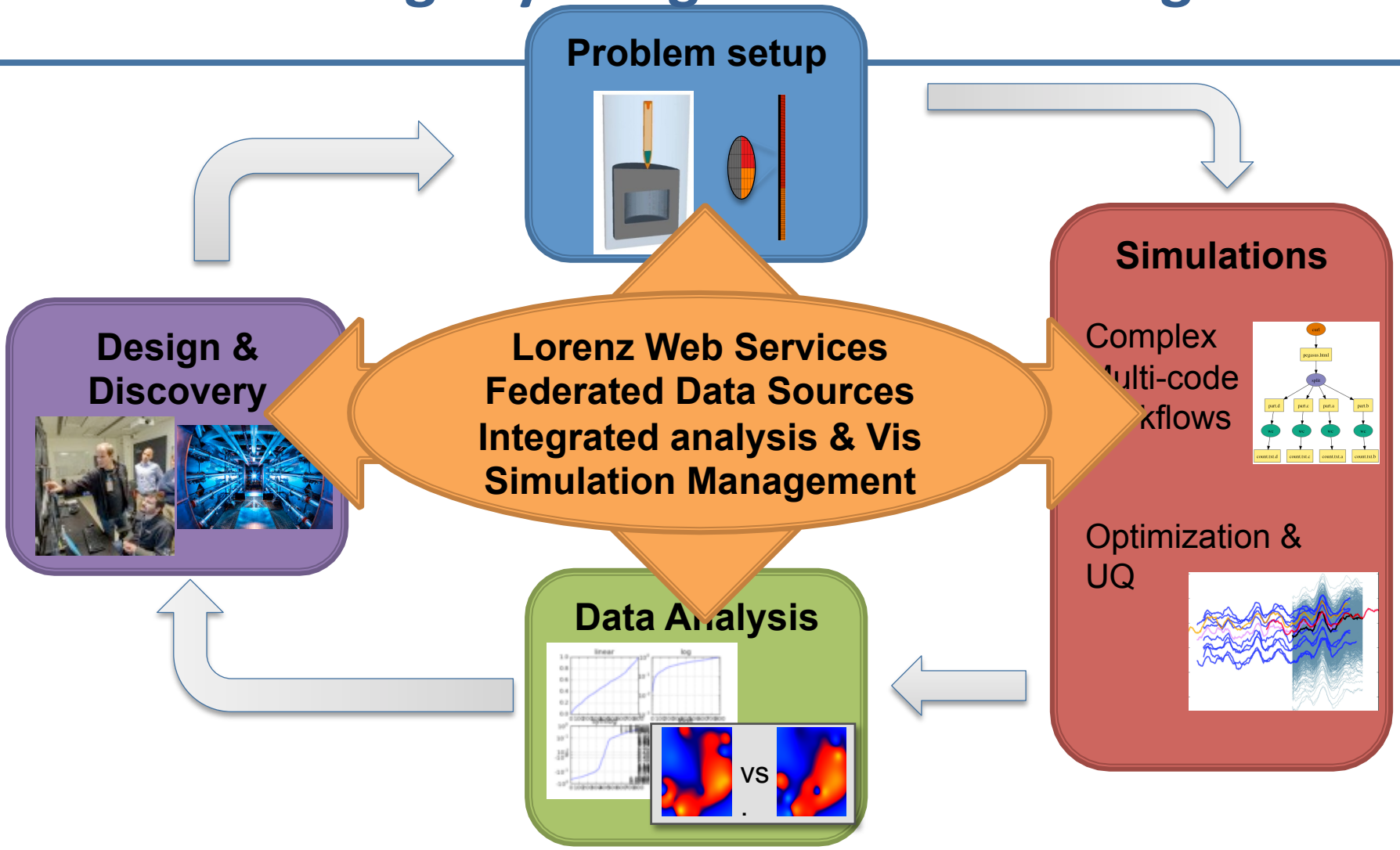
Our users have a diverse set of complex, iterative workflows



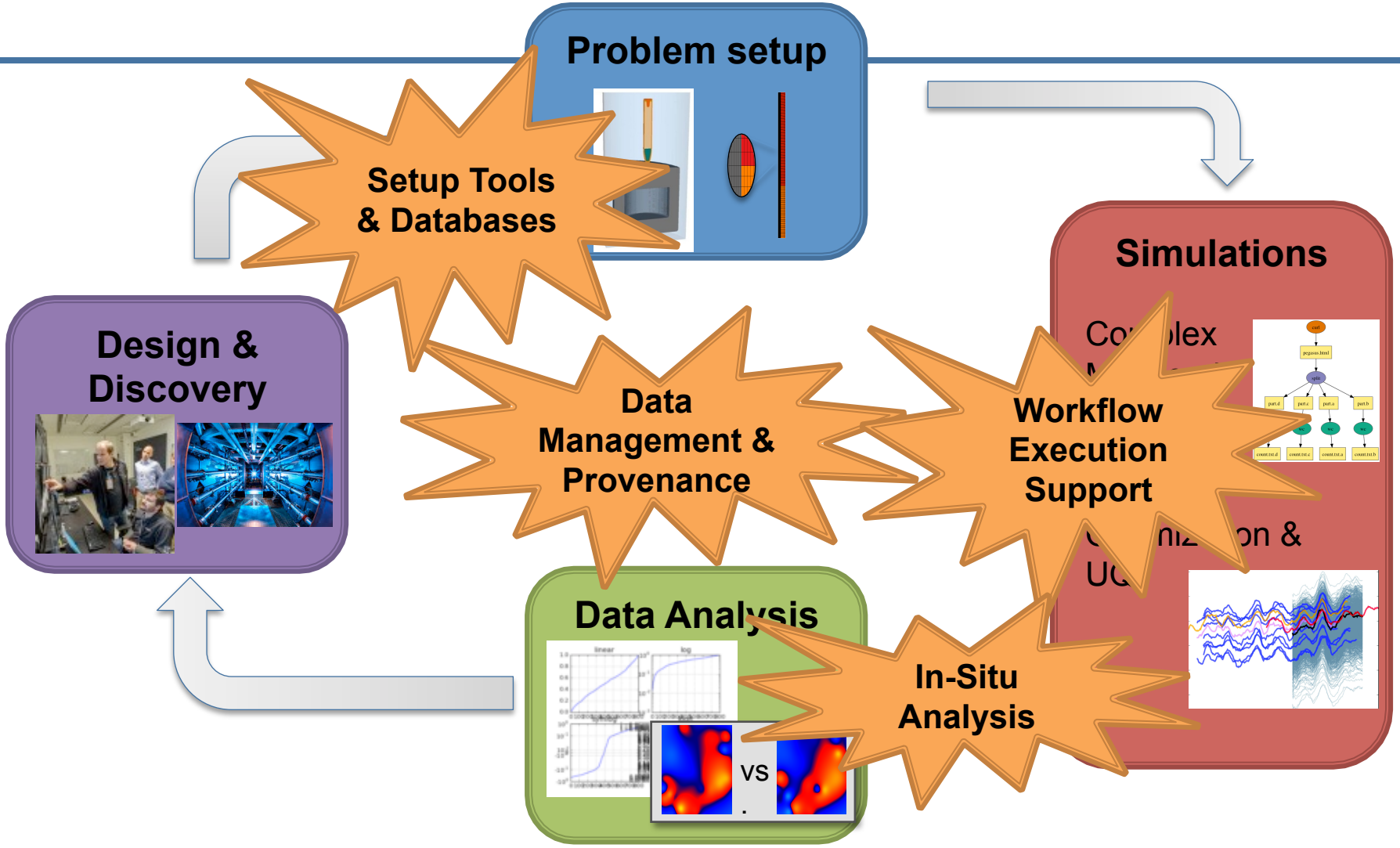
We provide a rich set of tools, but lack an integrated workflow environment for our users



Our vision is to unify our tools around Lorenz services and tightly integrated data management



ATDM must close several gaps to realize this vision



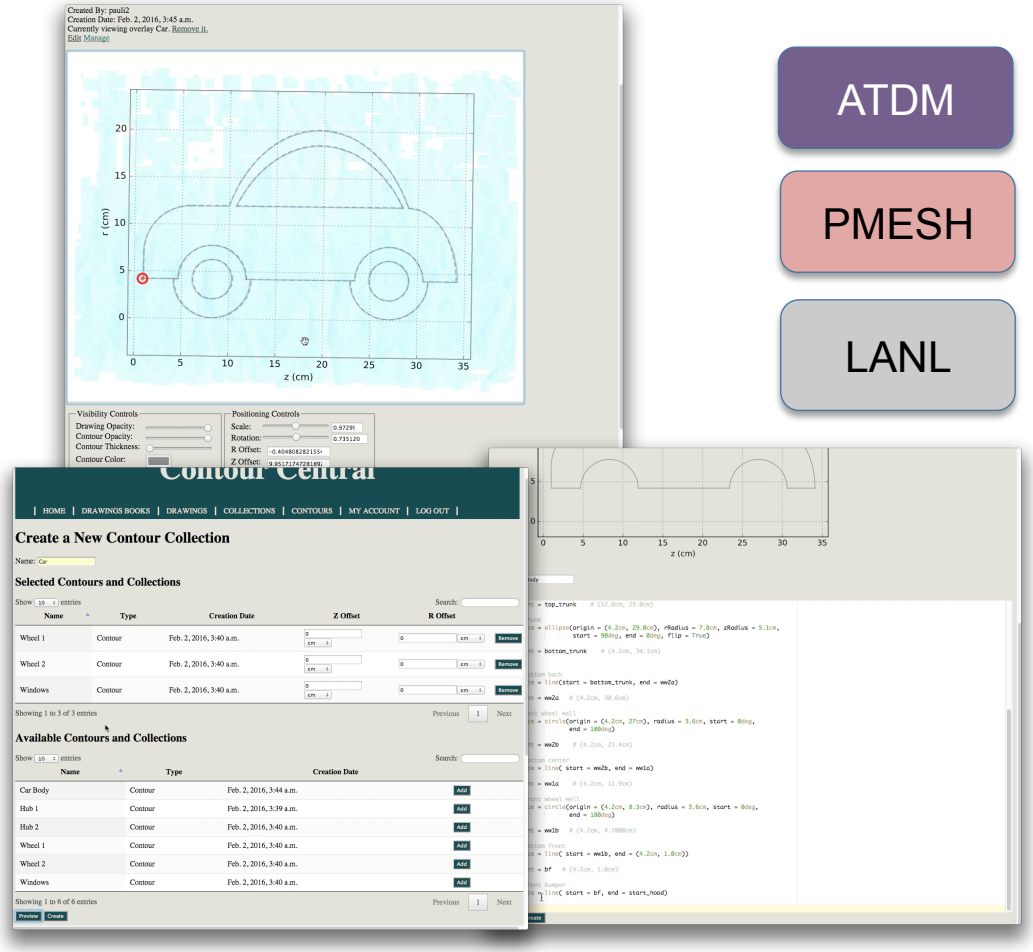
ATDM efforts in these areas are leveraging new collaborations and partnerships

Setup Tools & Databases	Integrated Data Management	Workflow Execution Support	In-situ Analysis
LANL - *Geometry Data format* ***SNL –*** *Sandia Analysis Workbench (SAW)* ***University of Illinois –*** *enhancements to OpenLorenz portal*	***SNL -*** SAW team scientific data management service ***Earth System Grid Federation (LLNL)***	***University of Southern California -*** *Pegasus Workflow Management system* ***LDRD –*** *Alkemi machine learning project at LLNL*	**Atomic Weapons Establishment (UK) -** Strawman batch in-situ infrastructure (LLNL led project)

LLNL – LANL – SNL Collaboration on shared nomenclature and workflow taxonomies

Geometry database and Contour Central demonstrator

- **Goal:** Common input geometry specification across multiple validated codes and labs
- Simple versioning and powerful editing of files
- Associating hierarchies of documents and geometry components



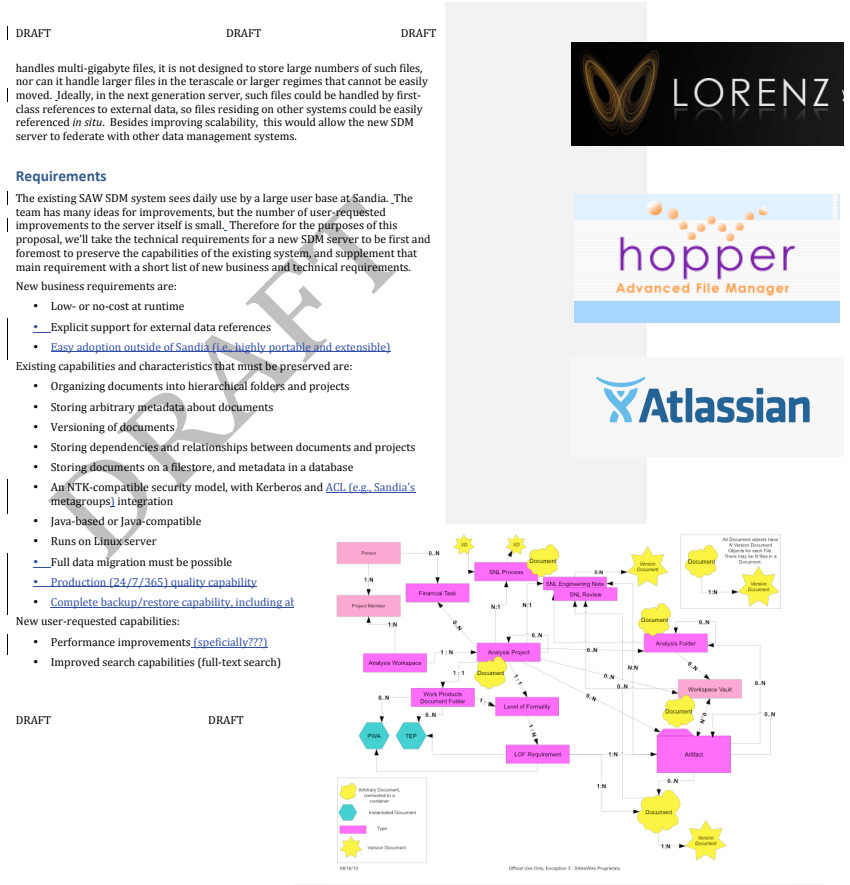
ATDM

PMESH

LANL

We are planning to co-develop a scientific data management infrastructure with Sandia

- **Goal:** to develop light weight components that can be utilized by both labs
- Share costs of surveying prior-art and new software development
- Status:
  - We have combined the initial set of requirements of both labs
  - Now working on logistics



DRAFT DRAFT DRAFT

handles multi-gigabyte files, it is not designed to store large numbers of such files, nor can it handle larger files in the terascale or larger regimes that cannot be easily moved. Ideally, in the next generation server, such files could be handled by first-class references to external data, so files residing on other systems could be easily referenced *in situ*. Besides improving scalability, this would allow the new SDM server to federate with other data management systems.

Requirements

The existing SAW SDM system sees daily use by a large user base at Sandia. The team has many ideas for improvements, but the number of user-requested improvements to the server itself is small. Therefore for the purposes of this proposal, we'll take the technical requirements for a new SDM server to be first and foremost to preserve the capabilities of the existing system, and supplement that main requirement with a short list of new business and technical requirements.

New business requirements are:

- Low- or no-cost at runtime
- Explicit support for external data references
- Easy adoption outside of Sandia (i.e. highly portable and extensible)

Existing capabilities and characteristics that must be preserved are:

- Organizing documents into hierarchical folders and projects
- Storing arbitrary metadata about documents
- Versioning of documents
- Storing dependencies and relationships between documents and projects
- Storing documents on a filestore, and metadata in a database
- An NTK-compatible security model, with Kerberos and ACL (e.g., Sandia's metagroups) integration
- Java-based or Java-compatible
- Runs on Linux server
- Full data migration must be possible
- Production (24/7/365) quality capability
- Complete backup/restore capability, including al

New user-requested capabilities:

- Performance improvements (speficially???)
- Improved search capabilities (full-text search)

DRAFT DRAFT

VV4ALE3D: a validation suite for ALE3D based on a new framework for managing simulations

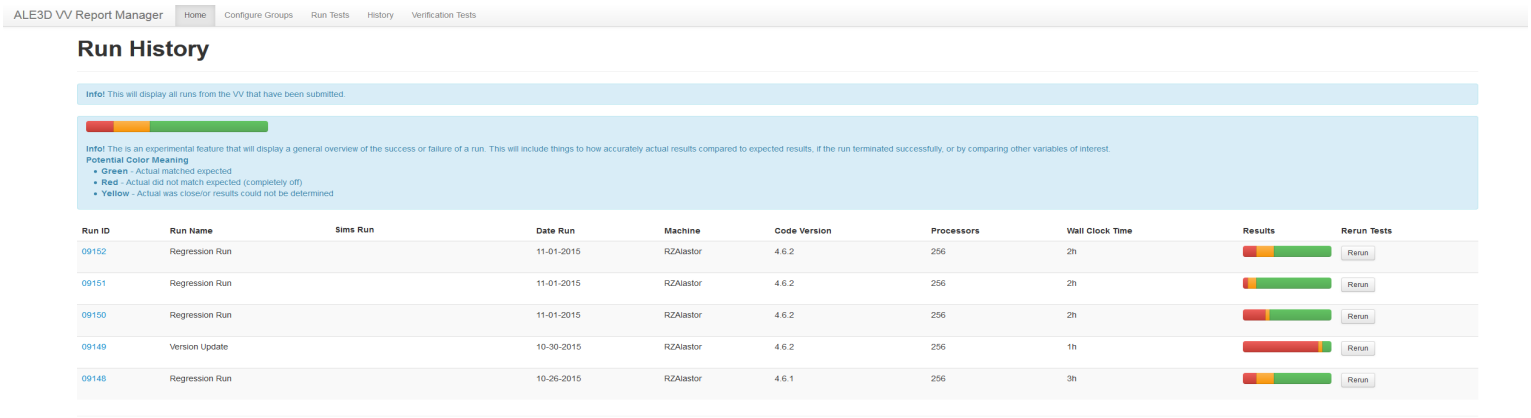
- **Goal:** to build extensible and plug-in oriented tools for managing and running suites of applications across WCI
- Integrating Web Technologies
  - Lorenz
  - Visualization and Analytics
- Helping to define data management requirements (SNL)
- Early target for Pegasus workflow engine evaluation (USC)

ATDM

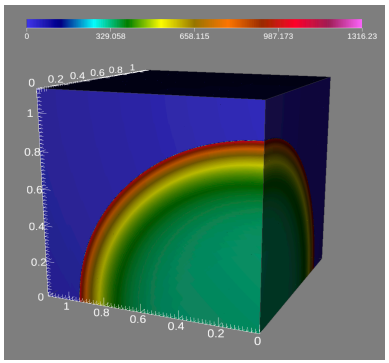
Lorenz (LC)

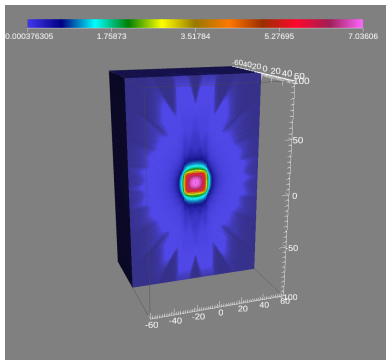
University of Illinois

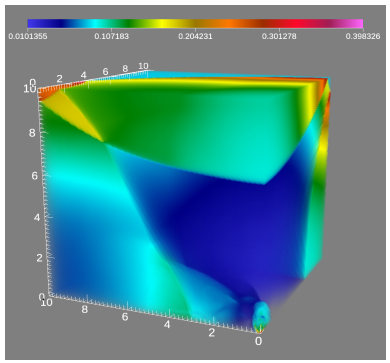
Pegasus (USC)



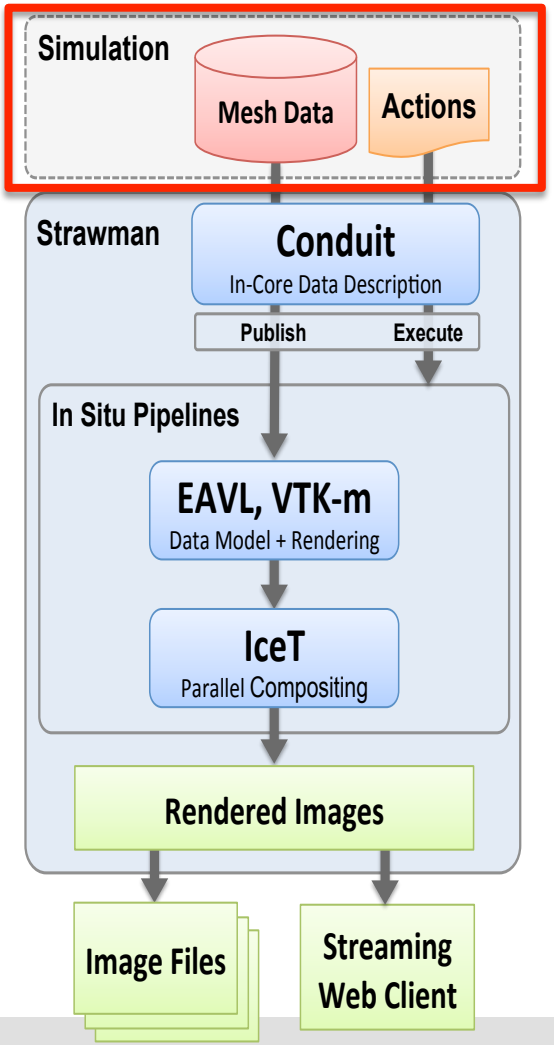
Strawman: a batch in-situ infrastructure for GPU-enabled analysis and visualization (with AWE in UK)



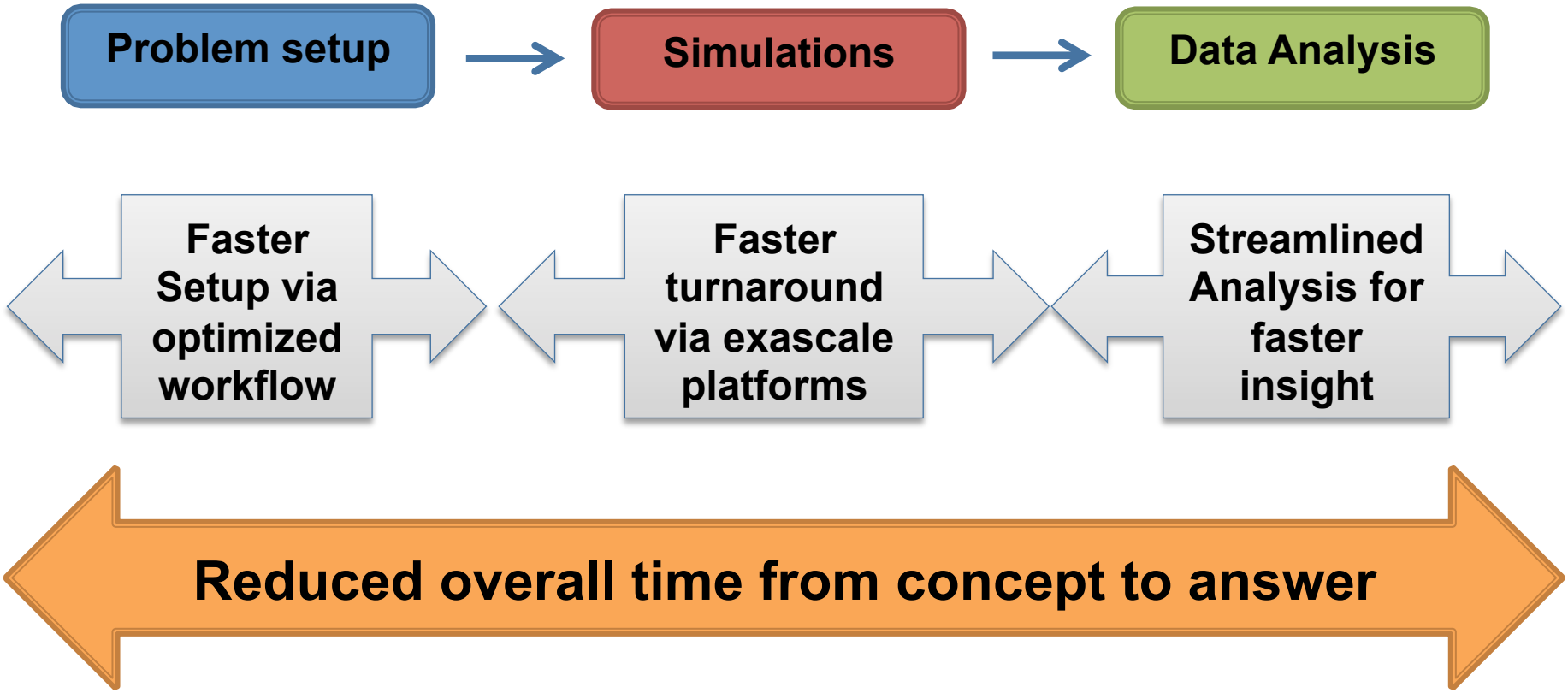




LULESH	Kripke	CloverLeaf3D
Hydrodynamics	Neutron Transport	Hydrodynamics
Unstructured	Uniform	Rectilinear
C++	C++	FORTRAN



Our goal is a cohesive workflow environment that enables our users to harness exascale computing



Lawrence Livermore
National Laboratory