

# The Materials Commons

A Novel Information Repository and Collaboration  
Platform for the Materials Community

Staff: Glenn Tarcea, Sravya Tamma,

Domain Scientists: Brian Puchala, Emmanuelle Marquis  
and John Allison

Information Scientists: Margaret Hedstrom (SI) and H.  
Jagadish (CSE)

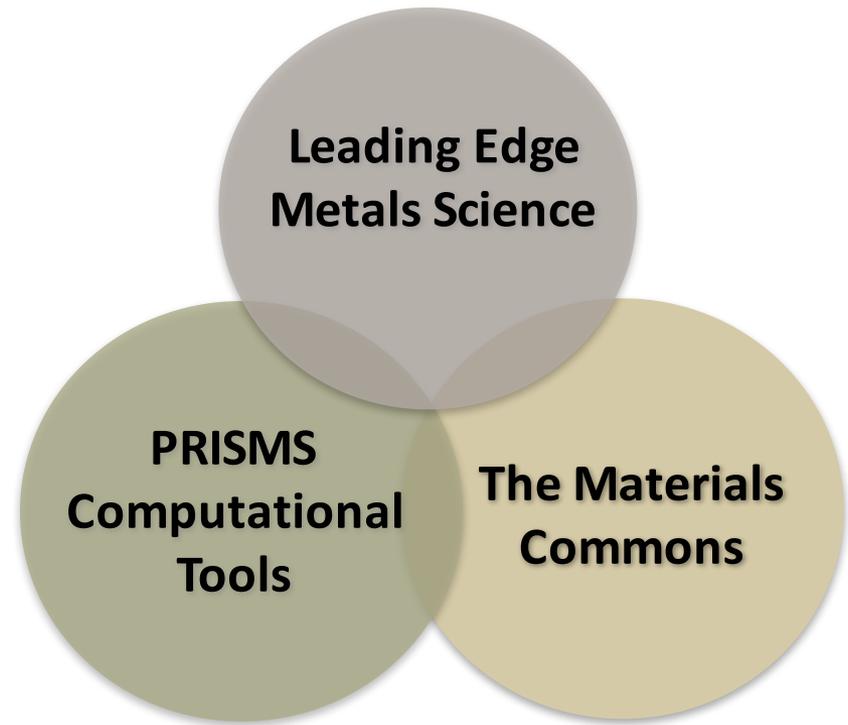
The University of Michigan

# PRISMS

DOE Software Innovation Center

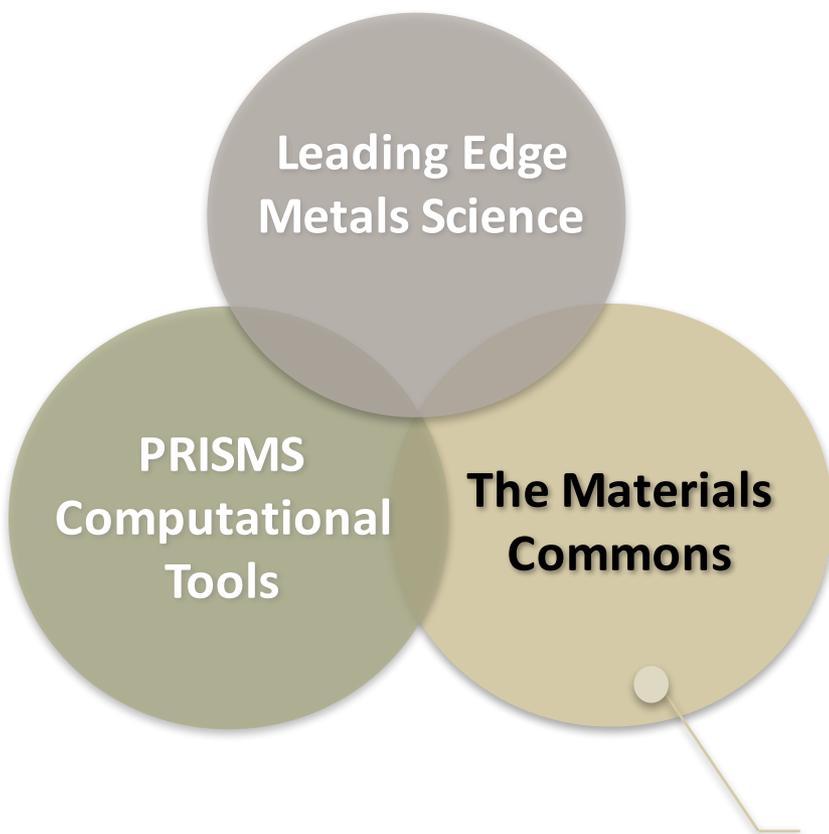
## Center for PRedictive Integrated Structural Materials Science (PRISMS)

- A 5-year grant **to accelerate the development of predictive materials science**
- Involves:
  - 11 faculty
  - 5 staff scientists
  - 16 students & postdocs
- DOE resources leveraged by significant UM cost share



# PRISMS Overarching Vision

Enable accelerated predictive materials science.



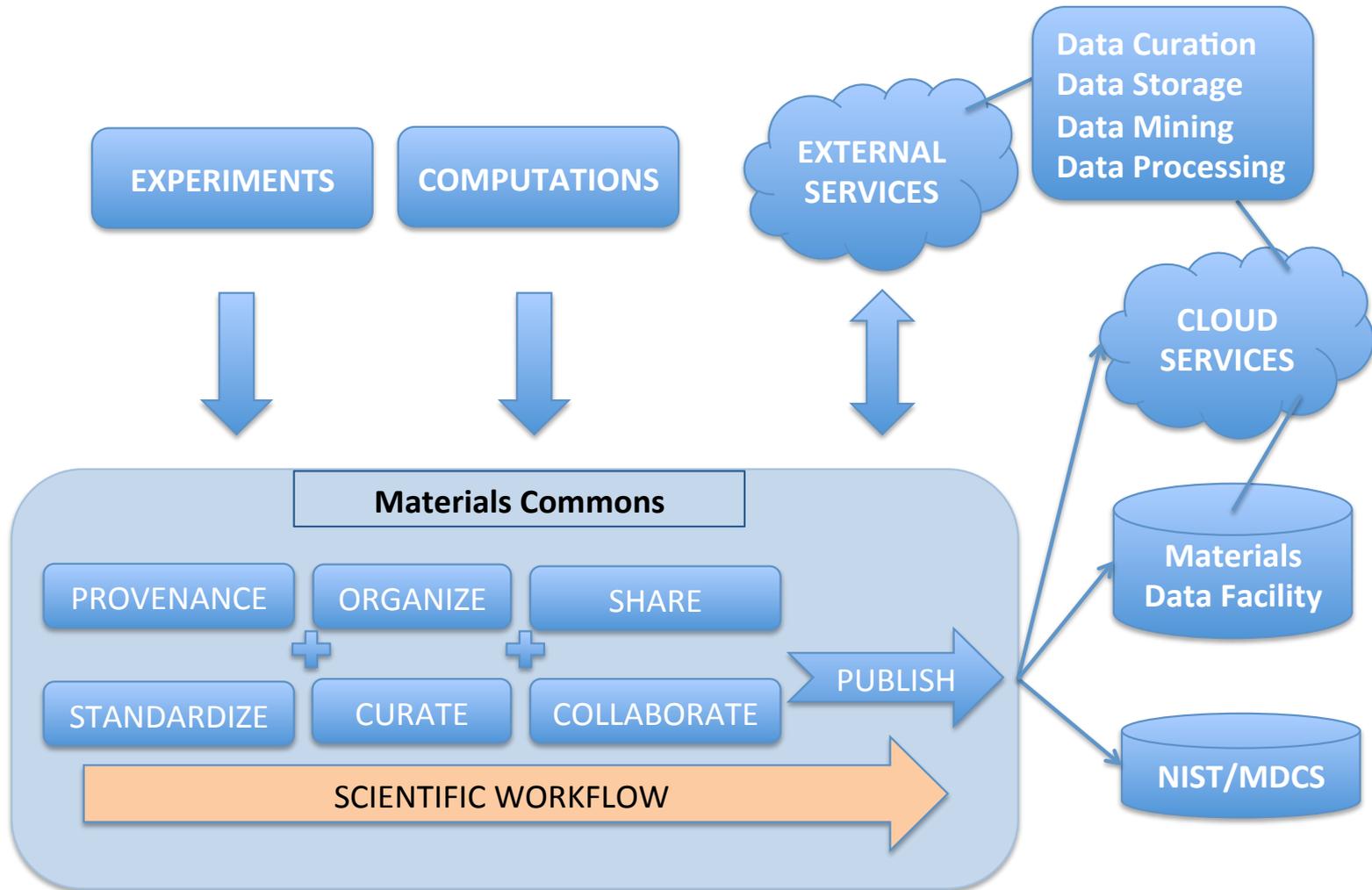
- Collaboration
- Experimental & Simulation Information
- Seamless, Continuous Workflow
- Provenance Tracking
- Accelerate model building and validation

# The Materials Commons

The Materials Commons consists of:

- 2 full time professional staff
- >25 PRISMS faculty and grad students as users.
- A 390 TB Isilon storage cluster data repository
- A website for uploading and downloading data, adding provenance, sharing and searching for data
- An application installed on your computer for uploading and downloading data
- A REST based API to access and extend the capabilities of the repository
- Interacting with NIST, ICE, CHiMaD etal, (and you?) to share best practices, schemas, etc.

# The Materials Commons - Workflow



# The Materials Commons:

<https://materialscommons.org>

## Materials Commons

Login



PROVENANCE



SAMPLES



REVIEWS



FILES



NOTES

# The Materials Commons - Interface

The screenshot displays the Materials Commons web interface. At the top, the user is logged in as 'johnea@umich.edu'. The main navigation bar includes 'Home', 'Help', 'Reviews 0', and 'Machines'. Below this, a breadcrumb trail shows 'AZ91 Precipitation Evolution' with a count of '10'. The dashboard features several buttons: 'Dashboard', 'Processes', 'Files', 'Samples (2)', 'Reviews (0)', and 'Notes (0)'. A sidebar on the left contains a 'Userguide (Draft)' section with a 'User Start' heading. The 'User Start' section includes a paragraph explaining the purpose of User Start and an 'Overview' section. The 'Overview' section contains a flowchart illustrating the process of creating provenance. The flowchart starts with 'Create Project', which branches into 'Upload files', 'Set up access', 'Create sample', and 'Add notes'. These four steps lead to a decision diamond labeled 'Create provenance'. From this diamond, the process can either 'Save as draft' (which leads to 'Submit to review') or 'Submit prov'.

**User Start**

When you login for the first time into Materials Commons we expect you will have no projects / data entered. User Start is intended to help all the new users who are trying to understand what MaterialsCommons does.

**Overview**

Provenance is the heart of the Materials Commons. The flowchart below is designed to help user understand the workflow of how to create Provenance. Areas highlighted in grey are the essential steps to create provenance

```
graph TD; A[Create Project] --> B[Upload files]; A --> C[Set up access]; A --> D[Create sample]; A --> E[Add notes]; B --> F{Create provenance}; C --> F; D --> F; E --> F; F --> G([Save as draft]); F --> H([Submit prov]); G --> I[Submit to review]; I --> G;
```

On-Line User Guide (currently in draft form) helps walk new users through the process

# The Materials Commons - Collaboration

The screenshot displays the Materials Commons interface for a project named "CPFEA Collaboration". The top navigation bar includes "Home", "Help", "Reviews 0", "Machines", and a user profile for "sriramg@umich.edu". The project header shows "CPFEA Collaboration" with a count of 5 and a welcome message for the user. Below the header are navigation buttons for "Dashboard", "Processes", "Files", "Samples (0)", "Reviews (0)", and "Notes (0)".

**Project Permissions:**

| User who have access to the project |
|-------------------------------------|
| agithens@umich.edu (owner)          |
| agithens@umich.edu                  |
| johna@umich.edu                     |
| sriramg@umich.edu                   |

**Processes (0)**

No processes.

**Files (1)**

Search files...

- CPFEA Collaboration (Tue Apr 21 2015)
  - WE43 T5 (as-recieved) (Tue Apr 21 2015)
    - CPFE Simulation (Mon Apr 27 2015)
    - Strain Maps (Tue Apr 21 2015)
      - FOV01\_1.25%\_strain.mat (378 MB, Tue Apr 21 2015)
      - FOV01\_2.98%\_strain.mat (370 MB, Tue Apr 21 2015)
      - FOV01\_5.16%\_strain.mat (365 MB, Tue Apr 21 2015)
      - FOV04\_4.44%\_strain.mat (370 MB, Tue Apr 21 2015)
    - undefined (Tue Apr 21 2015)
      - T5\_stress\_strain.mat

**Reviews (0)**

Search reviews...

No reviews.

**Notes (0)**

Search notes...

No notes.

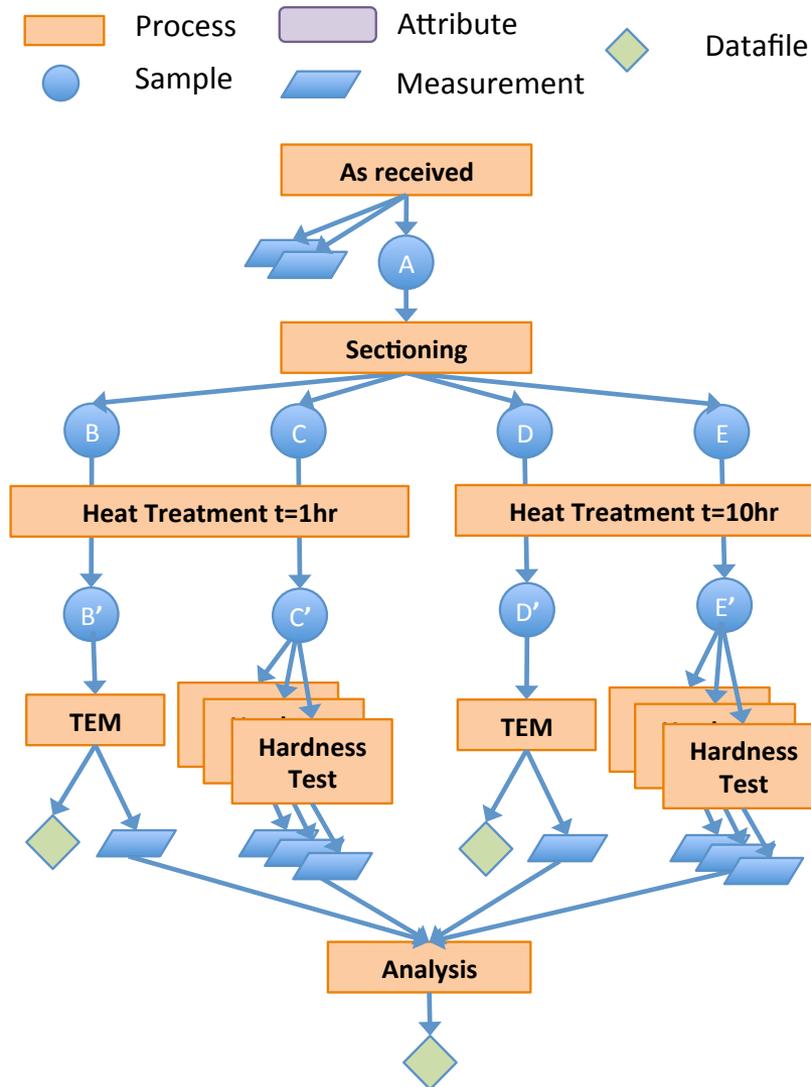
**Teams work in projects**

- They share files through the website
- Reviews provide a mechanism to get input and track responses
- Notes are an additional mechanism for knowledge sharing and capture

# Materials Commons Data Model

- *Sample*: Representations of real or virtual things
  - materials, experimental or virtual specimens, etc.
- *Sample Attributes*: Measured attributes of a sample
  - grain size, weight, modulus, concentration, diffusivity, etc.
- *Measurements*: A measured value of a sample attribute.
  - may be one or many per sample attribute
- *Process*: Representations of actions
  - APT, SEM, Fatigue Test, Monte Carlo Calculation, Continuum Plasticity Calculation, etc.
  - specifies if a process transforms a sample attribute
  - templates specify what provenance information must / should / can be given

# Materials Commons Data Model

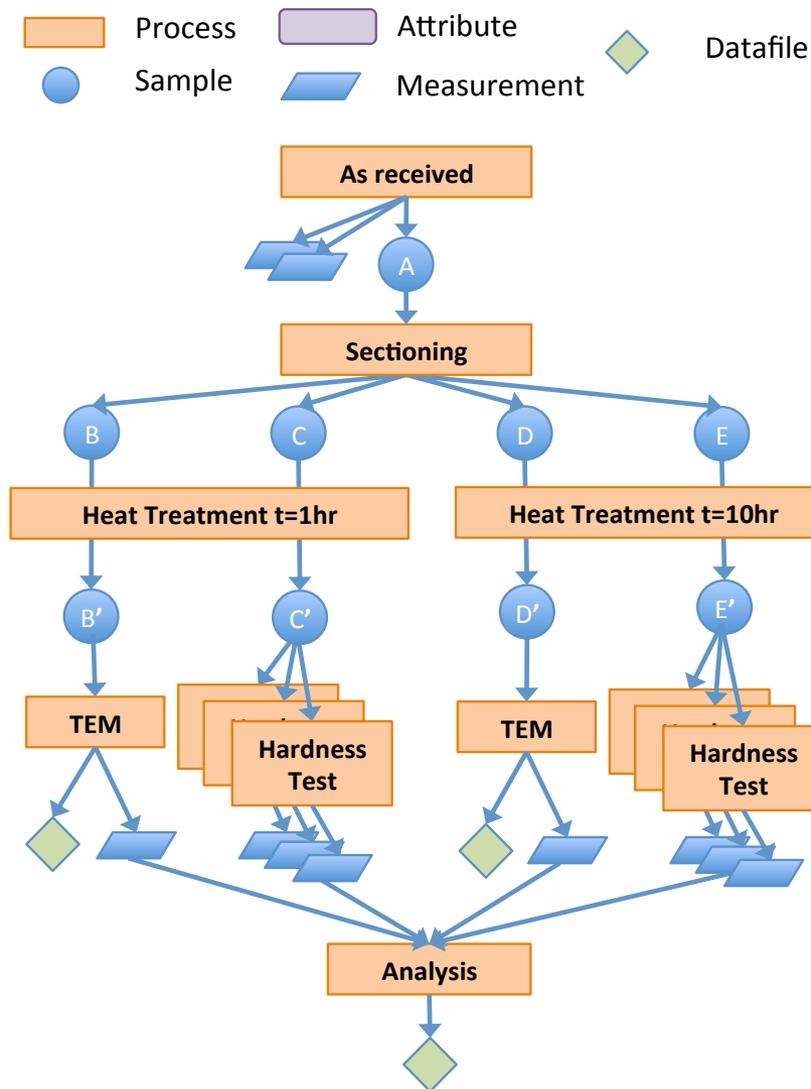


Example experiment:

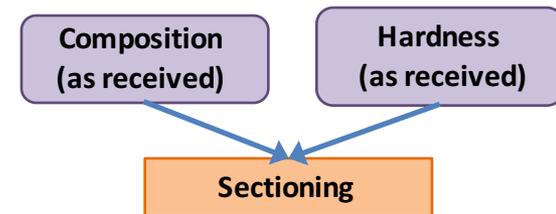
Ppt. size & hardness = f(heat treatment)

- Recieve material
- Section into specimens
- Measure ppt. size by TEM
- Measure hardness by indentation tests
- Analyze the results, fit them to a model, and create plots.

# Materials Commons Data Model

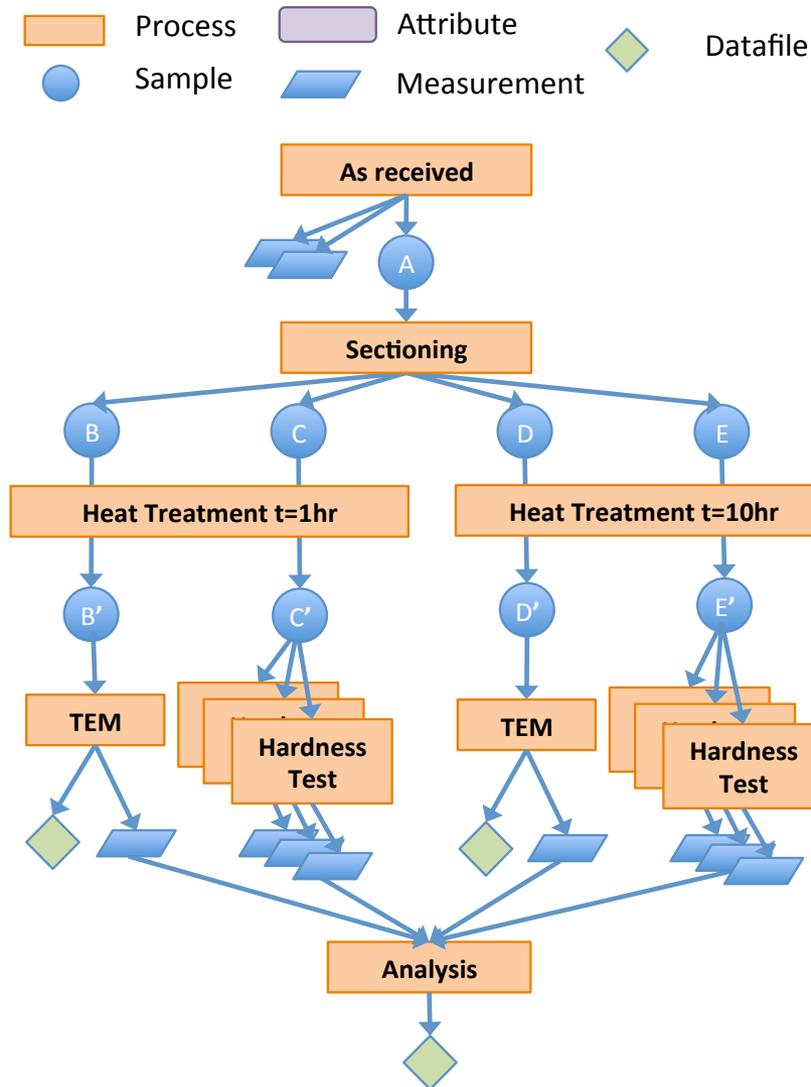


Scientific knowledge is encoded in the process data:

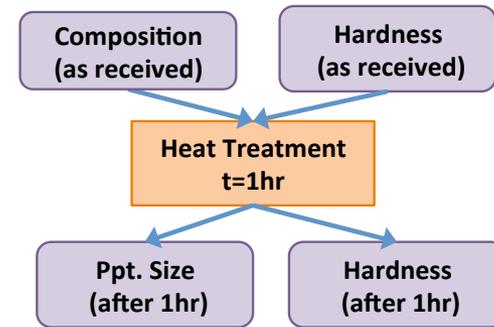


- Sectioning creates new samples, but the created samples have the same composition and hardness.
- Measurements on one sample (A, B, C, D, or E) can be applied to all samples.

# Materials Commons Data Model

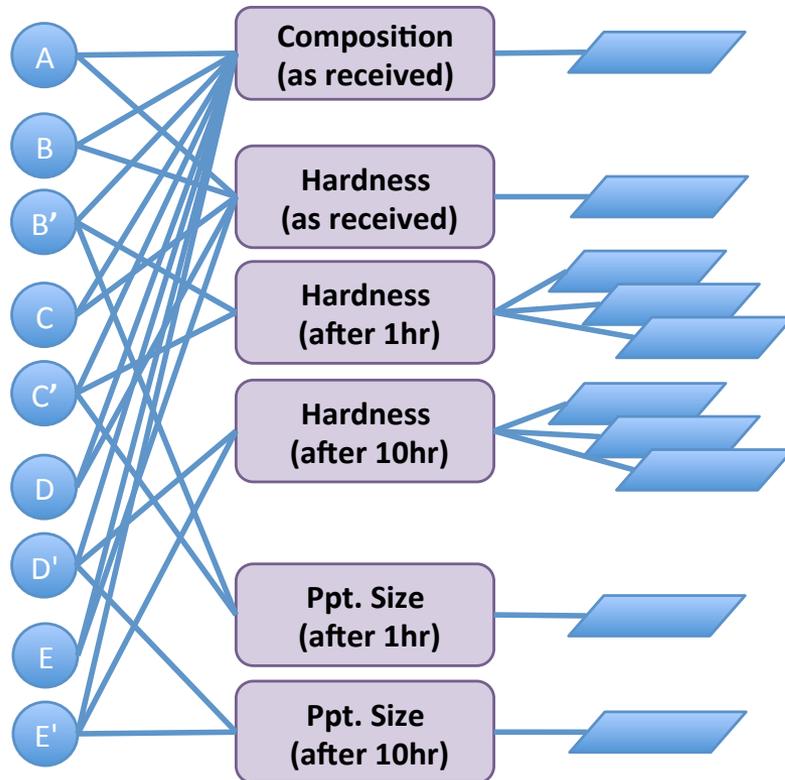


Scientific knowledge is encoded in the process data:



- Heat treatment (ex. B -> B') does not transform the composition, but it does transform the hardness and ppt. size.

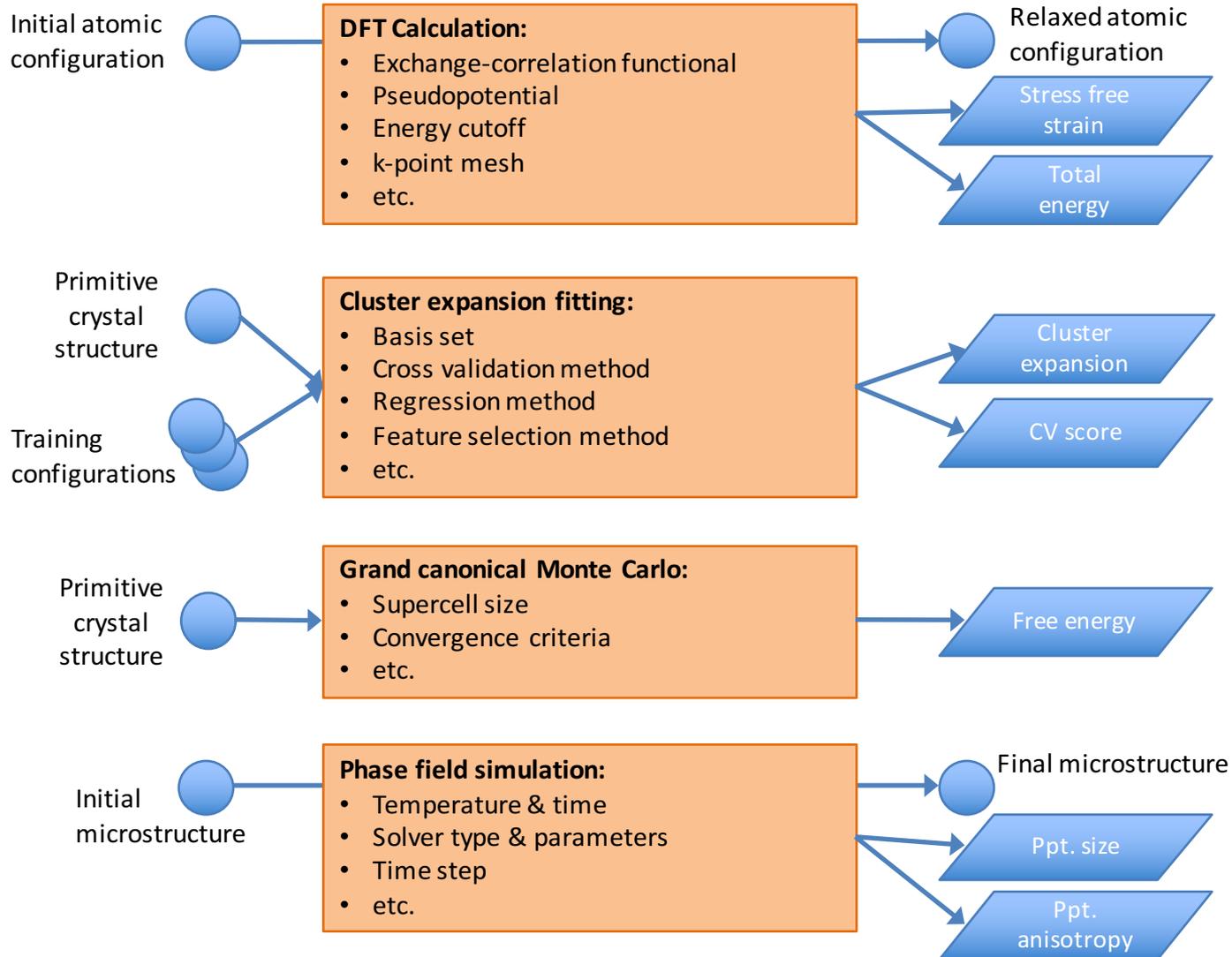
# Materials Commons Data Model



Attributes may be shared across samples, so measurements are immediately applied to all relevant samples.

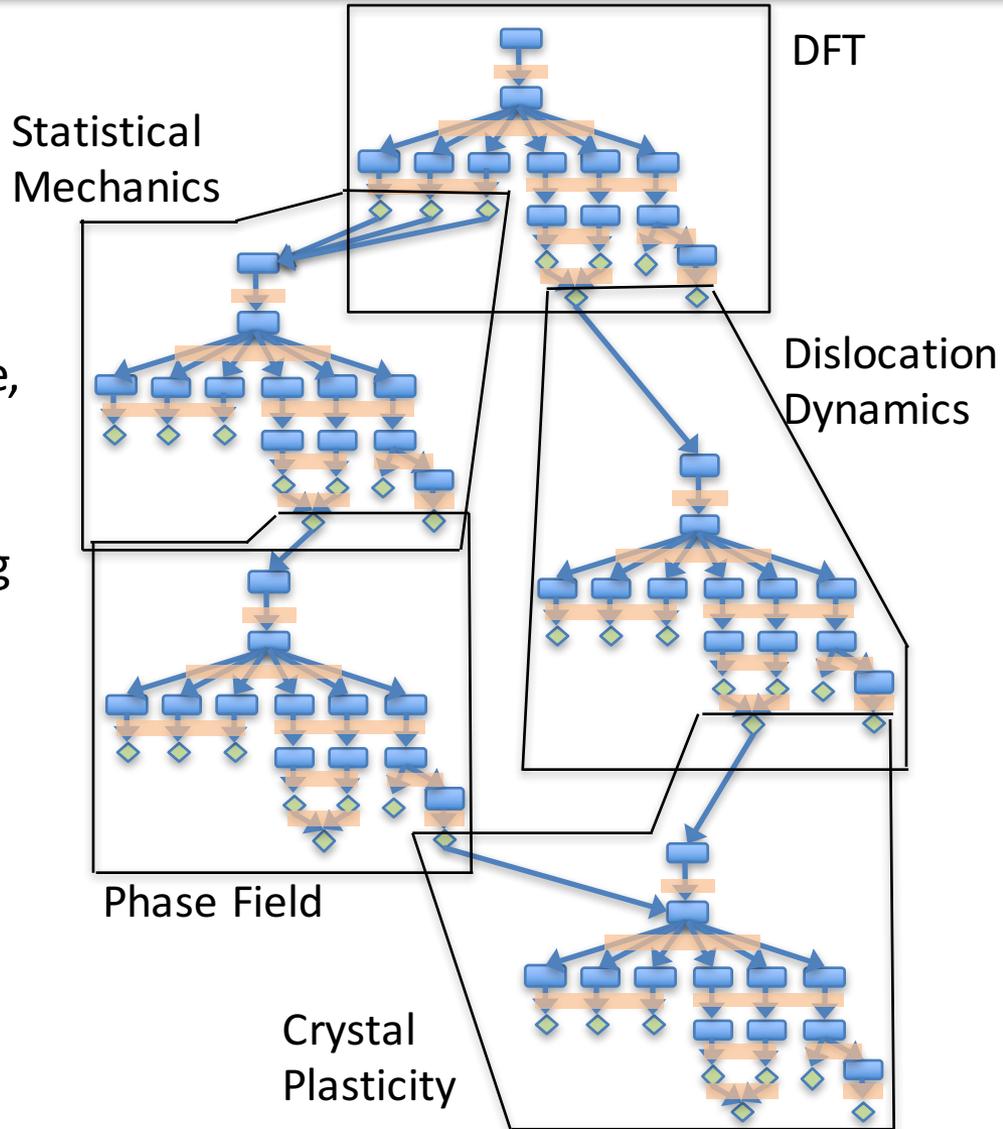
- Composition:
  - shared by all samples
- Hardness (as received)
  - shared by {A, B, C, D, E}
- Hardness and Ppt. size (after 1hr):
  - shared by {B', C'}
- Hardness and Ppt. size (after 10hr):
  - shared by {D', E'}

# Provenance for Multi-scale Modeling



# Share provenance

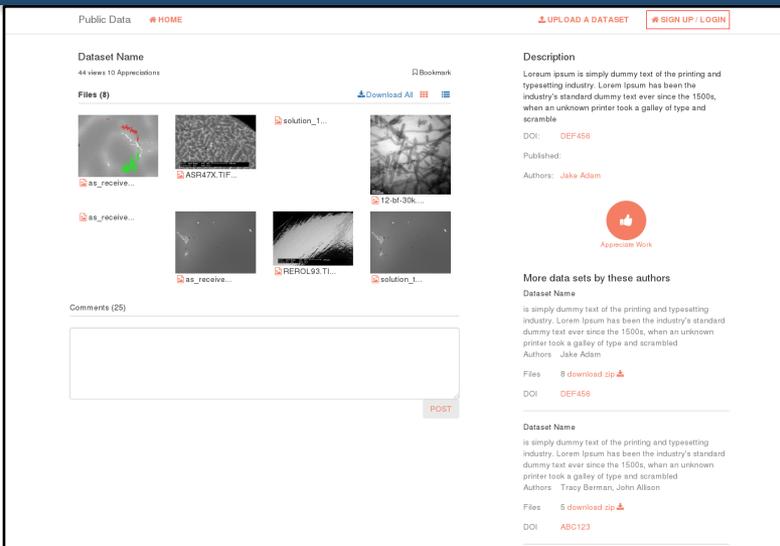
- Once a project or dataset is shared with you, you may include it's sample, processes, etc. in another project.
- Data may be re-used while maintaing provenance



# Use Data

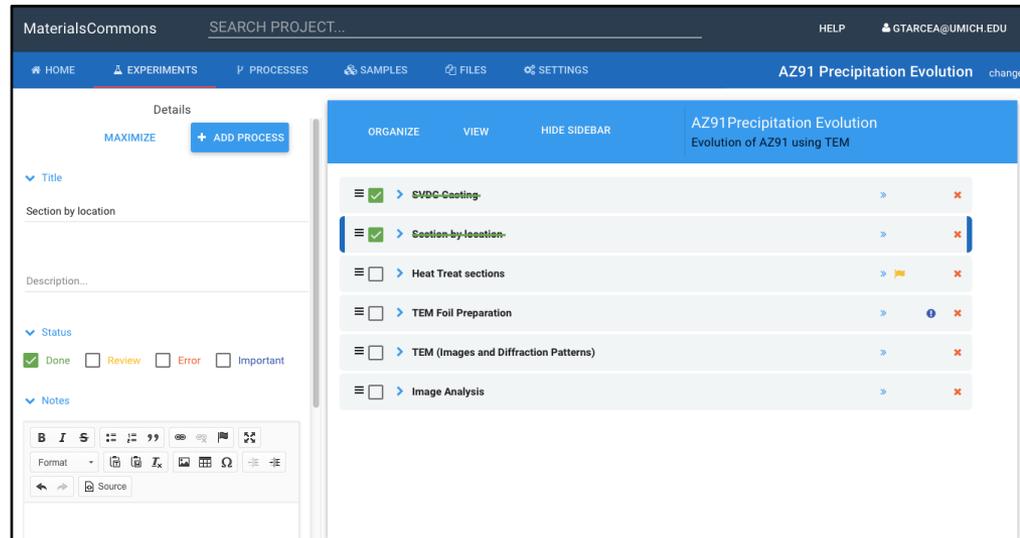
- Flexibly visualize and analyze
- Fit constitutive models
- Fit process models
- Provenance graph -> workflow -> optimization:
  - Multi-scale integration
  - Material design
  - Model reduction
  - Optimal design of experiments and computations

# The Materials Commons Upcoming Features



- Make public datasets:
  - Persistent identifiers
  - Search and browse for datasets
  - Give / get feedback, usage stats
  - Download formatted data
  - Re-use data in new projects, maintaining provenance

- New "Experiments" interface:
  - Organize 'task' or 'todo' list
  - Document as you go
  - Provenance graph is built for you
  - Ready to make public when finished
- Python API
  - Write scripts to upload & download data and provenance



# Conclusions

- The Materials Commons is a novel information repository and collaboration platform for the materials community
  - Focus on PRISMS technical emphasis areas
- Aim to be a seamless part of the scientific workflow
  - Structured data storage, with provenance
  - Collaborative
- Enables data use for science and engineering



# Materials Commons Technologies

- API/REST Services are all JSON based
- Data Sync is moving to MsgPak, Curve, and ZeroMQ
- Website uses Socket.IO for real time updates, REST/JSON for service access
  - Just started integrating this in
- JSON Document store on backend (RethinkDB)
- Website written using AngularJS
- Backend is a mix of Python and Go
- Some Erlang mixed in
- RabbitMQ used for pipeline processing

# Architecture Block Diagram

