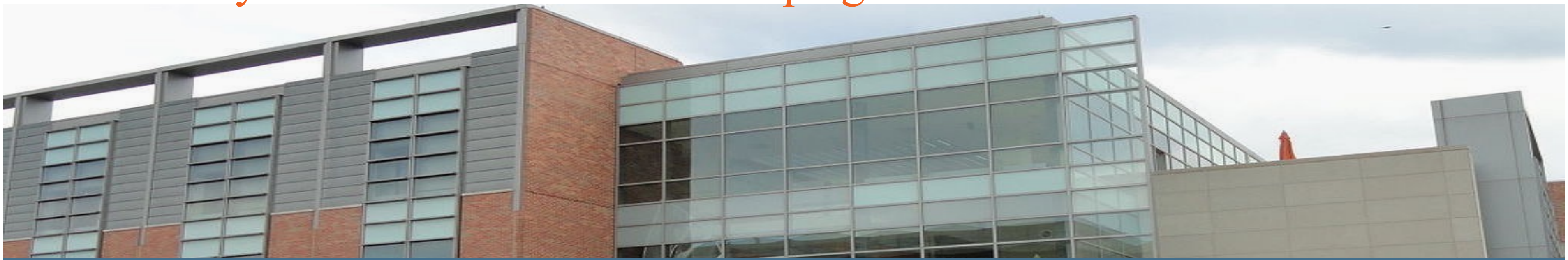
T2C2

Accelerating Science via Smart and Joint Cyber-Infrastructure for Materials and Semiconductor Fabrication Data and Metadata

Klara Nahrstedt
University of Illinois at Urbana-Champaign

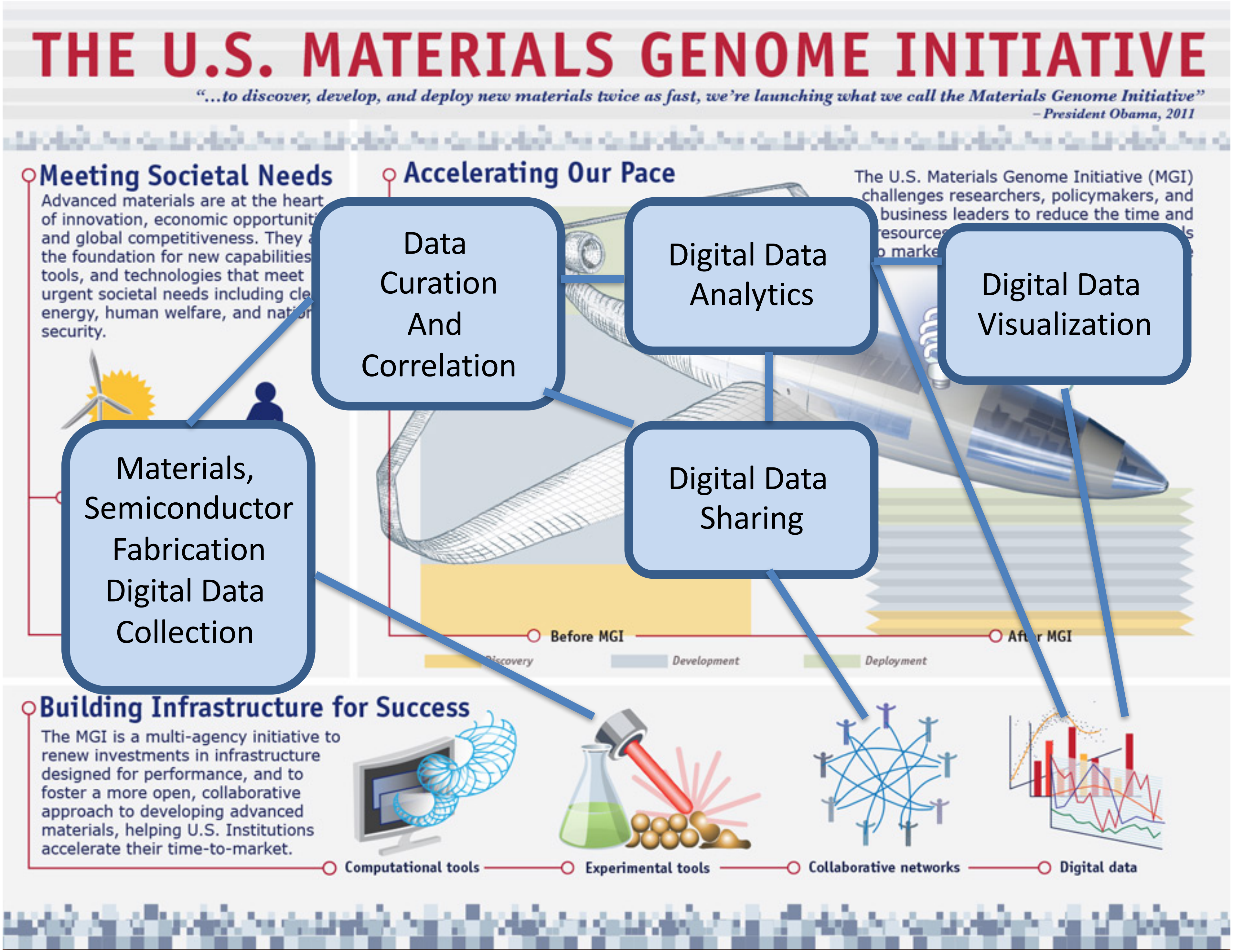


A timely and trusted curator and coordinator of scientific data

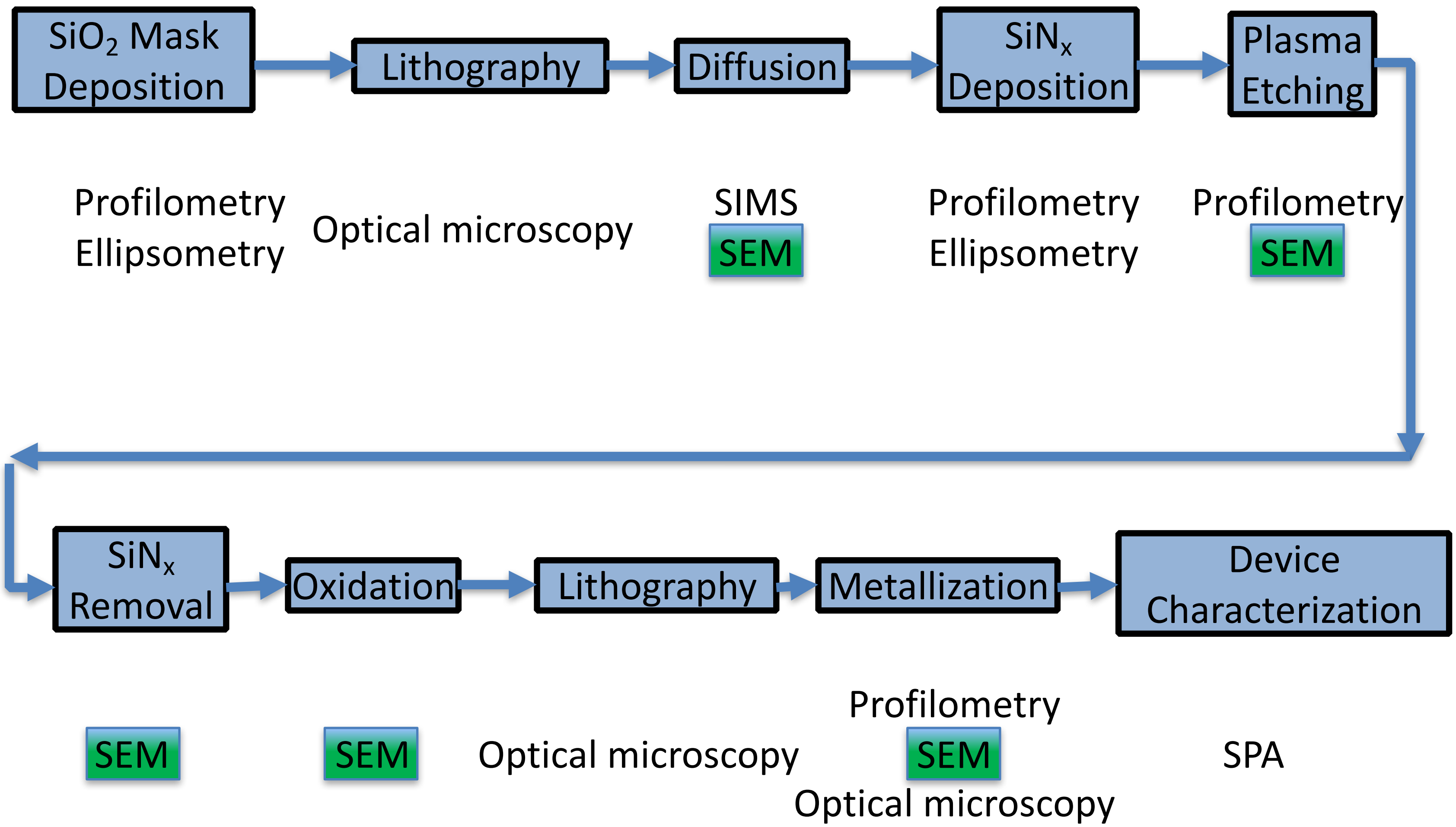
MRL FREDERICK SEITZ MATERIALS RESEARCH LABORATORY

MNTL micro+nanotechnology lab

CSL: COORDINATED SCIENCE LAB

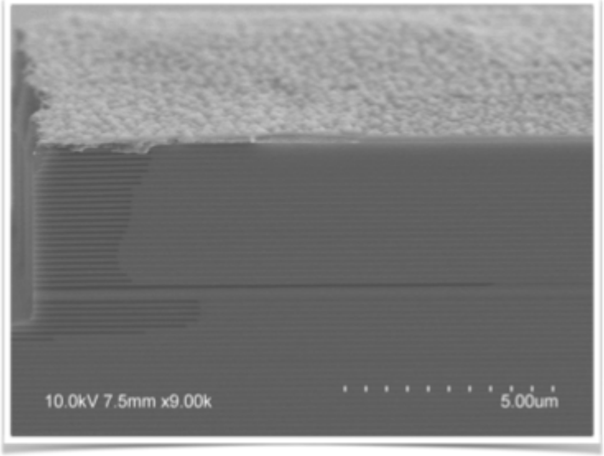

THE U.S. MATERIALS GENOME INITIATIVE
"...to discover, develop, and deploy new materials twice as fast, we're launching what we call the Materials Genome Initiative"
– President Obama, 2011
Meeting Societal Needs
Advanced materials are at the heart
of innovation, economic opportunit
and global competitiveness. They a
the foundation for new capabilities
tools, and technologies that meet
urgent societal needs including cle
energy, human welfare, and nati
security.
Accelerating Our Pace
The U.S. Materials Genome Initiative (MGI)
challenges researchers, policymakers, and
business leaders to reduce the time and
Before MGI
After MGI
Discovery
Development
Deployment
Building Infrastructure for Success
The MGI is a multi-agency initiative to renew investments in infrastructure designed for performance, and to foster a more open, collaborative approach to developing advanced materials, helping U.S. Institutions accelerate their time-to-market.
Computational tools
Experimental tools
Collaborative networks
Digital data
Data Curation And Correlation
Digital Data Analytics
Digital Data Visualization
Materials, Semiconductor Fabrication Digital Data Collection
Digital Data Sharing

Example of Semiconductor Device Fabrication Process



Example of Collected Data

An example of the result from an experiment at MNTL



Result image of 07302013-Oxidation experiment

Experimental setting:

Time 13min
Temp 425 C

(Structured meta data)

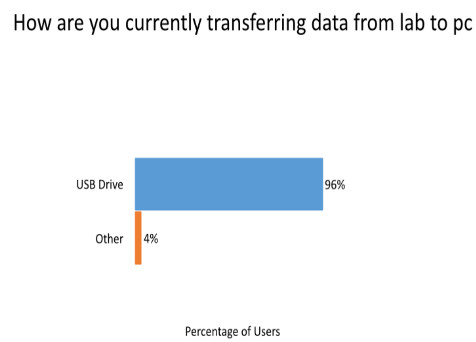
Notes:

Oxidation depth is about 12um. Oxidation layer composed of Al(0.98)GaAs with thickness of 30 nm. Furnace in 2111 MNT L, 2" diameter quartz.

(Free text)

Current State of Data Capture in Materials and Semiconductor Domains

- Current situation for experimental data involves manual processes for data capture and storage leading to poor documentation of results



- Data transfer is often done via "sneaker-net" techniques using flash drives or email
- No data file conversion is available

- "Best" results and images are kept, but what is "best" is determined by a narrow, specific scientific objective. "Imperfect" data is often discarded or not available for others to review.

- Data only correlated only through publications

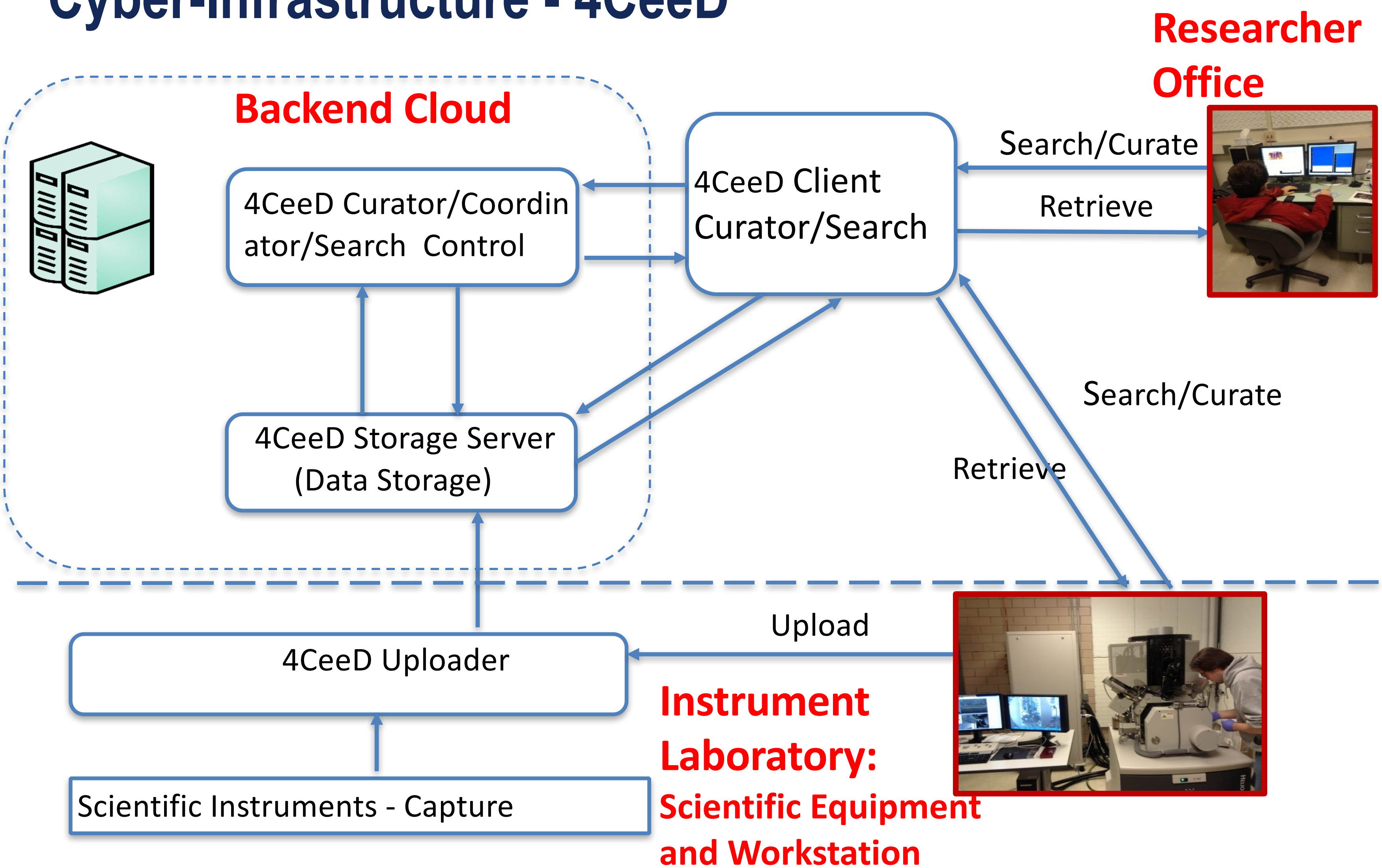
Effects of Current State

- Measurements on multiple instruments for a new material may not be well correlated due to mechanisms to encode the linkages between measurements.
- Novel device prototypes can be difficult to reproduce due to a lack of proper capture of “recipes” used.
- In addition, previous experiments in the deposition systems may affect subsequent experiments.
- Curation of system information can greatly improve the reproducibility and understanding of results.

Important Question

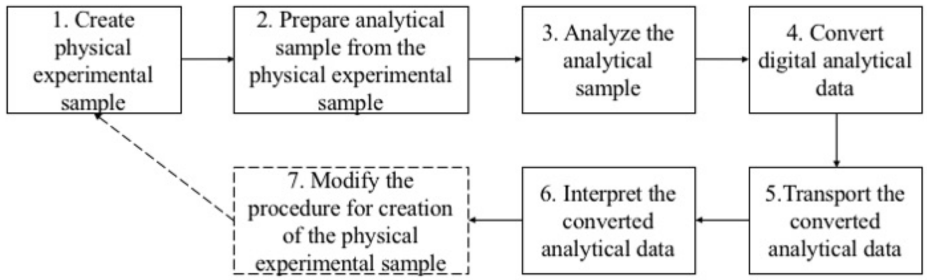
- How do we accelerate pace of material science and semiconductor fabrication ?
- Approach: MUST ACCELERATE AT MULTIPLE LEVELS!!
  - Accelerate Lab Sessions for scientists at microscopes
  - Make it easy to curate multi-modal science data (allow free text)
  - Enable correlation of material data and semiconductor fabrication processes using AI techniques to enable easy and fast search of correlated multi-modal data

Our Approach: Distributed Real-Time "Smart" Data Cyber-Infrastructure - 4CeeD

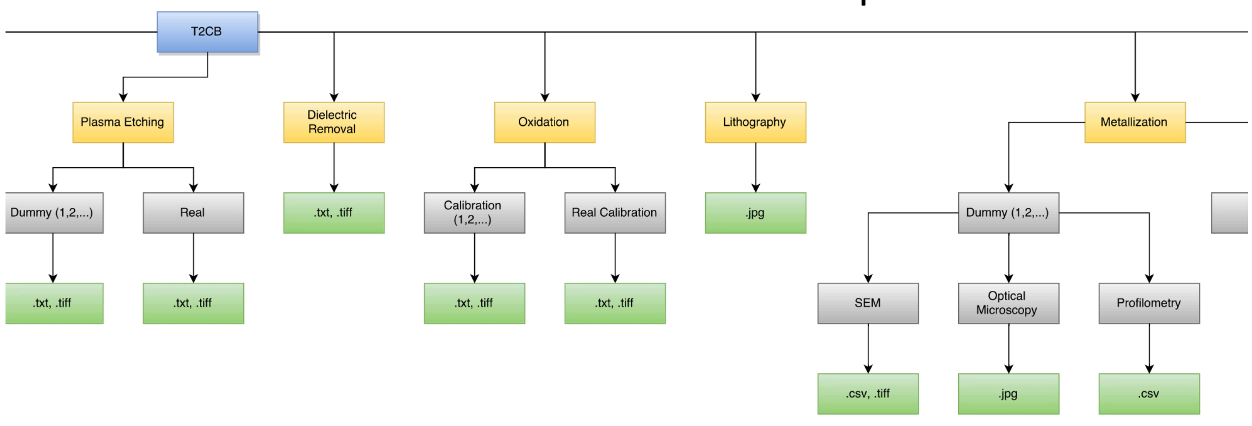


Acceleration of Science - User Process and Data Model

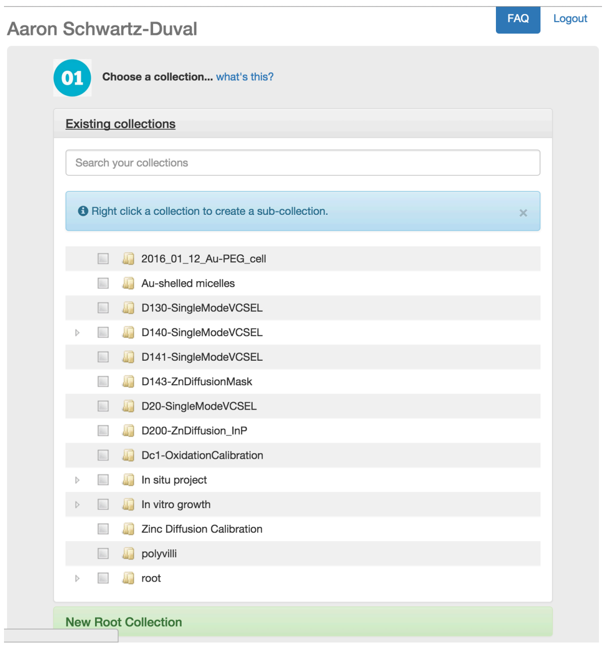
Example of an experiment task flow

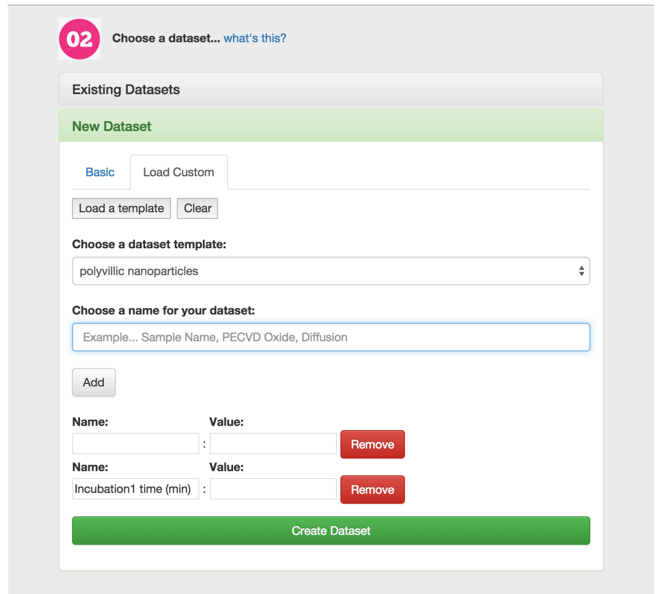


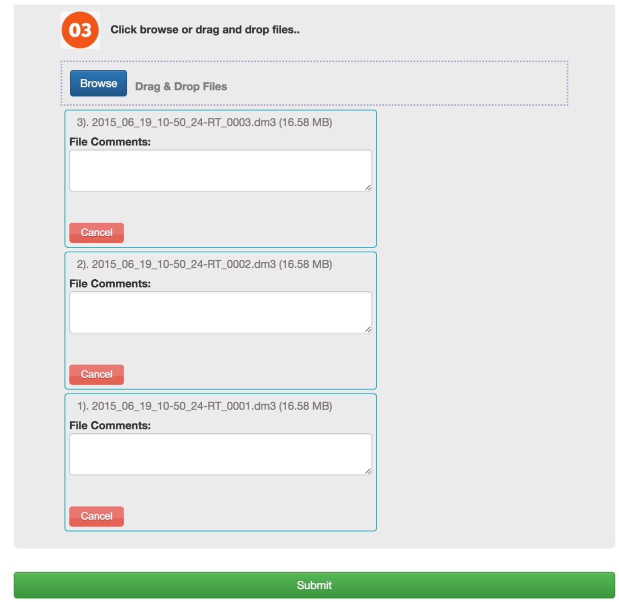
4CeeD Data Model organizes projects into collections, datasets, and files. These can then be shared in spaces.



Acceleration of Users' Lab Sessions - 4CeeD Uploader







Create or Select A Collection

Create or Select Dataset

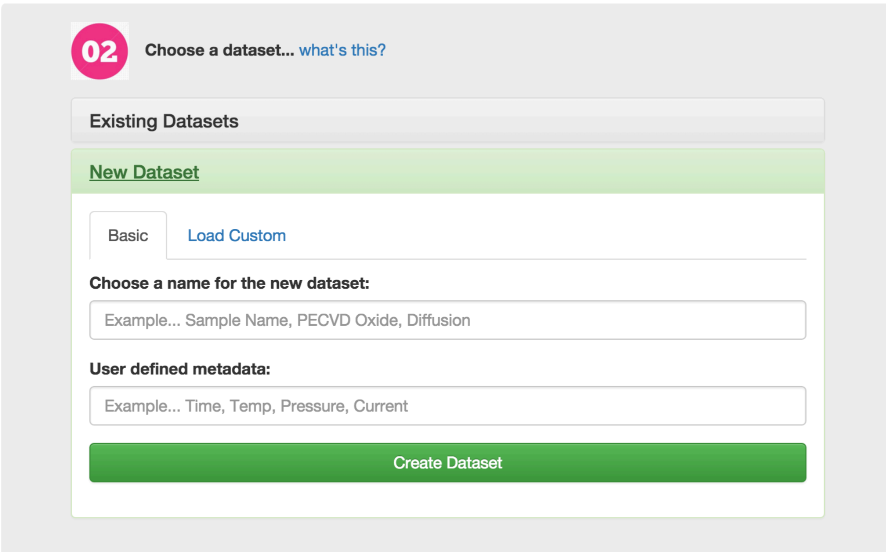
Optional: Choose template and enter metadata

Upload Files

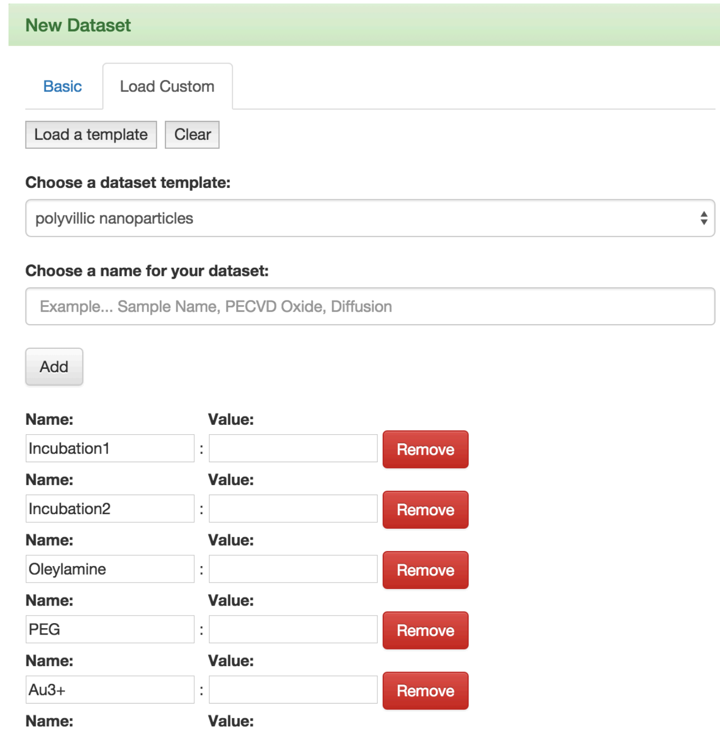
CLIENT Video

Acceleration of Users' Lab Sessions – Adding User Templates

Plain text only



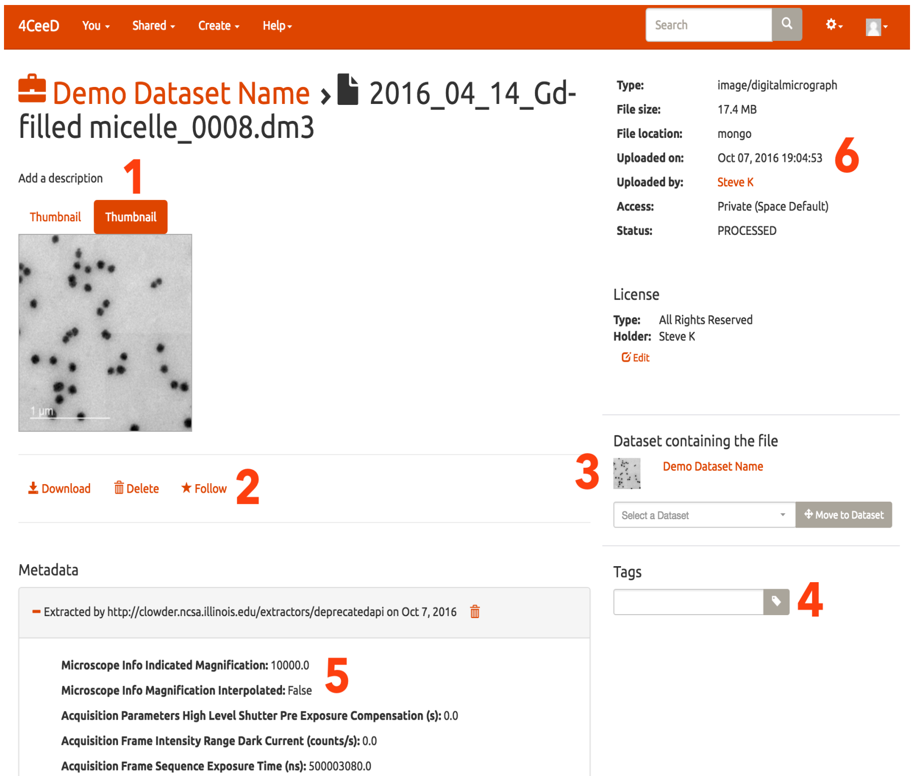
Custom Select/Create Template



Many user-defined datasets incorporate a dozen or more key/value pairs. Allowing users to select a global template or save their own template saves time and avoids errors.

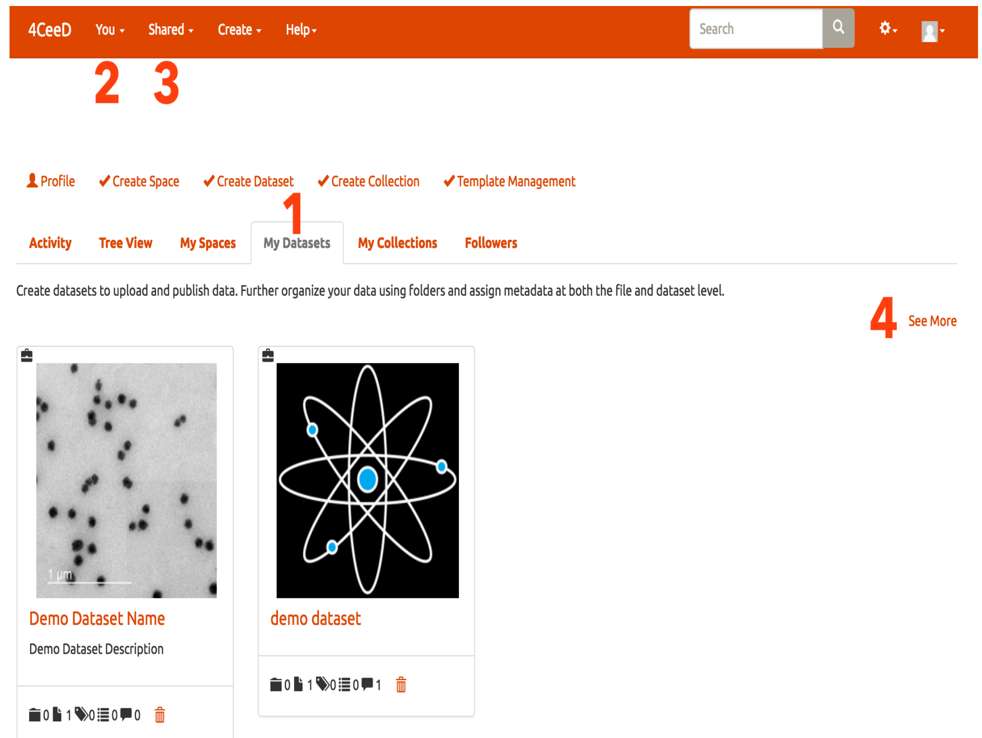
Acceleration of Data Manipulation - 4CeeD Curator

File View



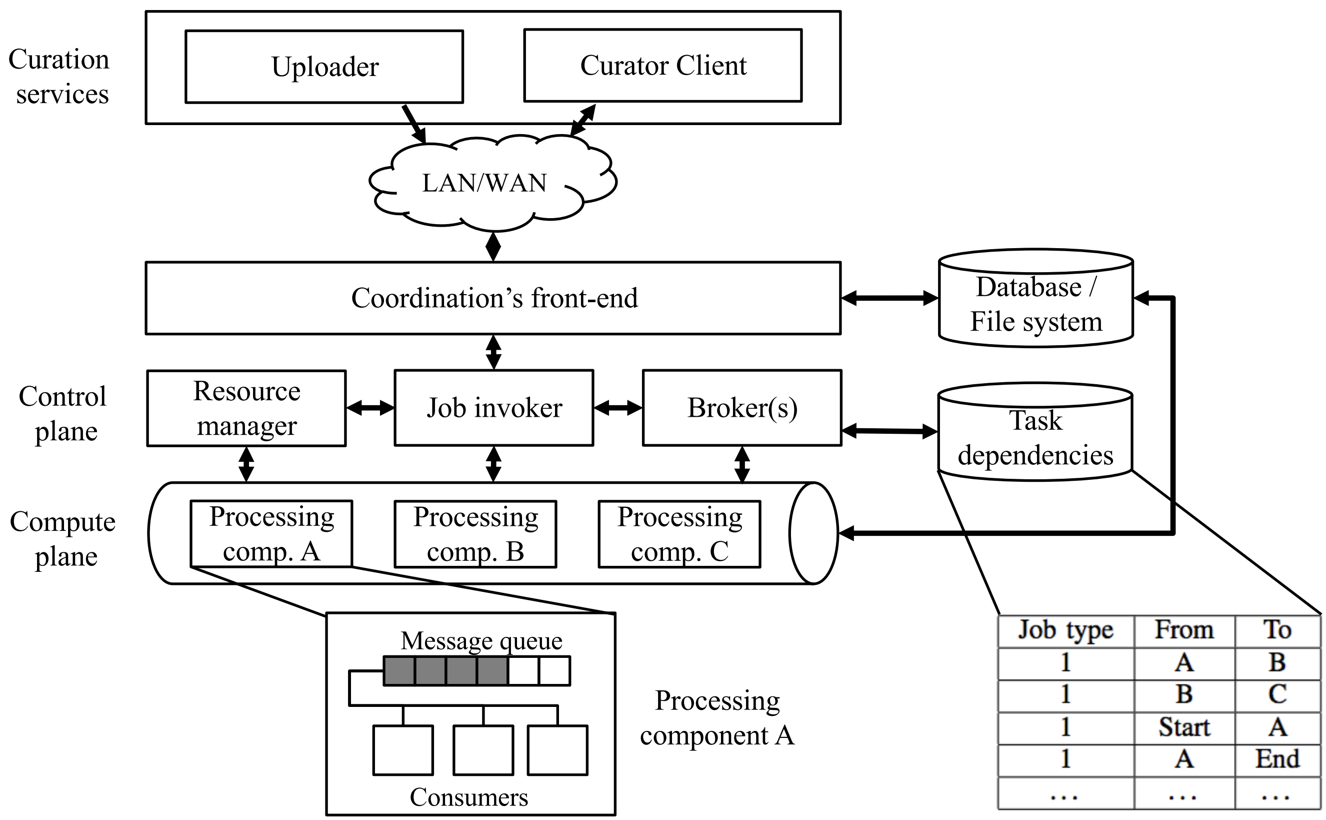
[Preview, annotate, download, extracted metadata]

Dashboard View



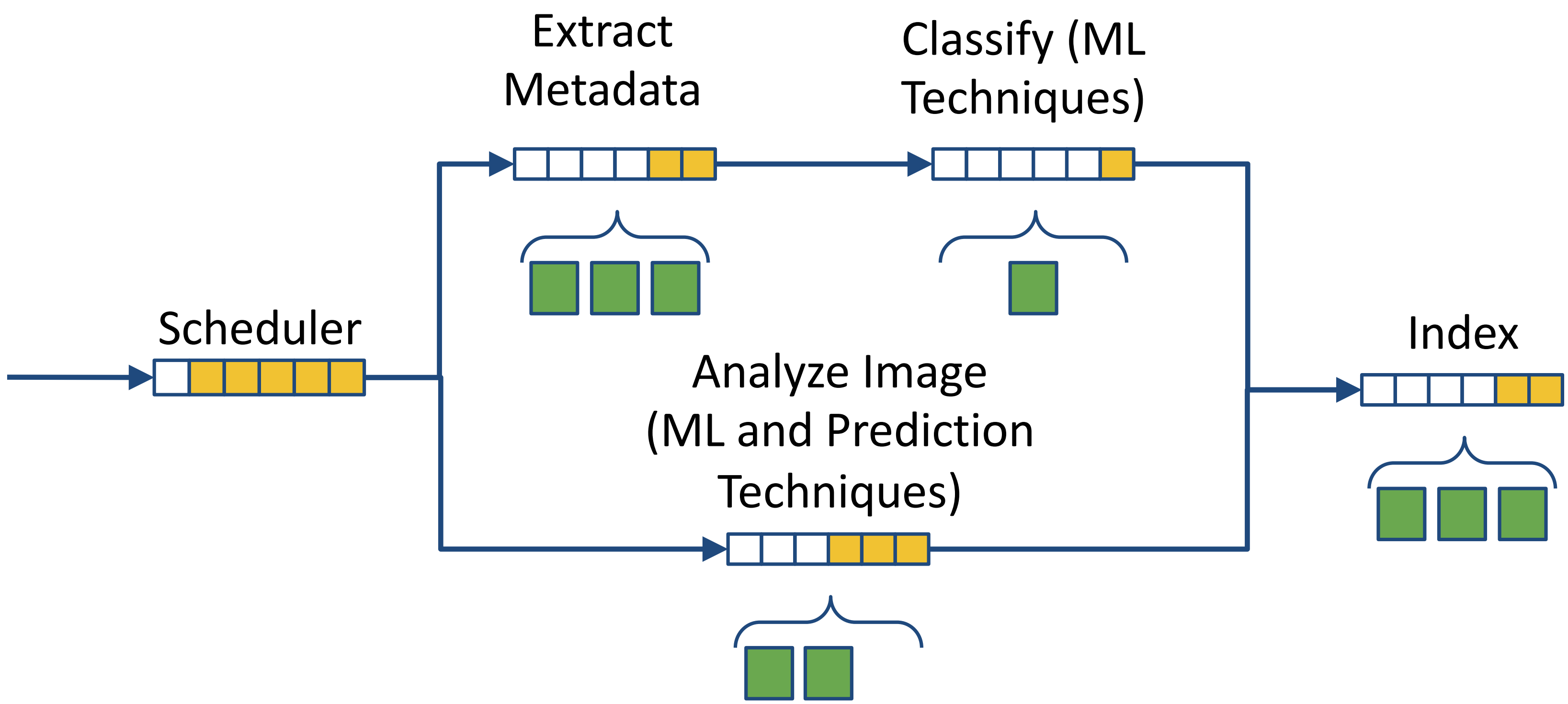
[Dashboard management]

Acceleration of Science - 4CeeD Cloud Coordinator



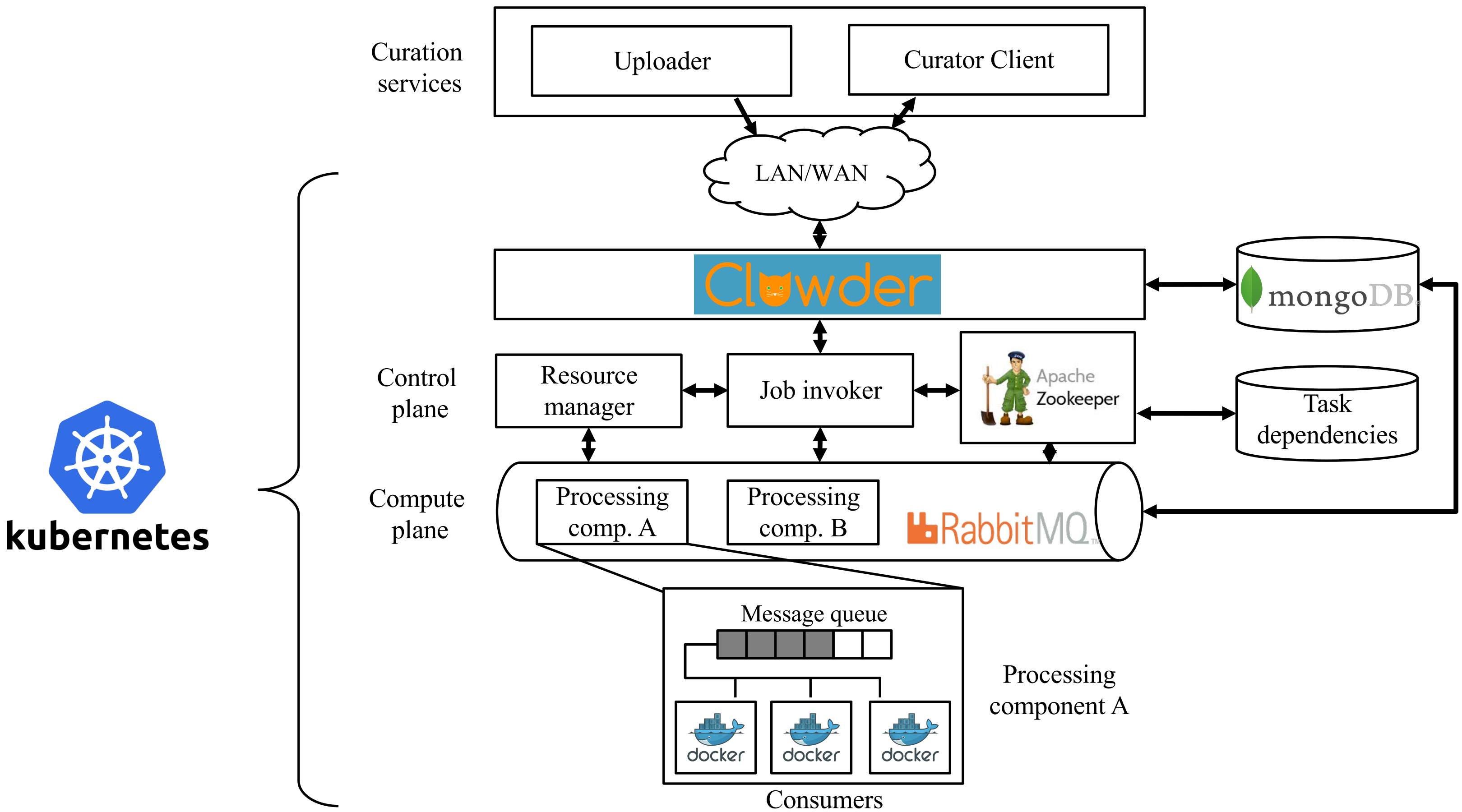
Workflow Execution with AI Tasks on Science Data

(Example of Type 1 workflow)

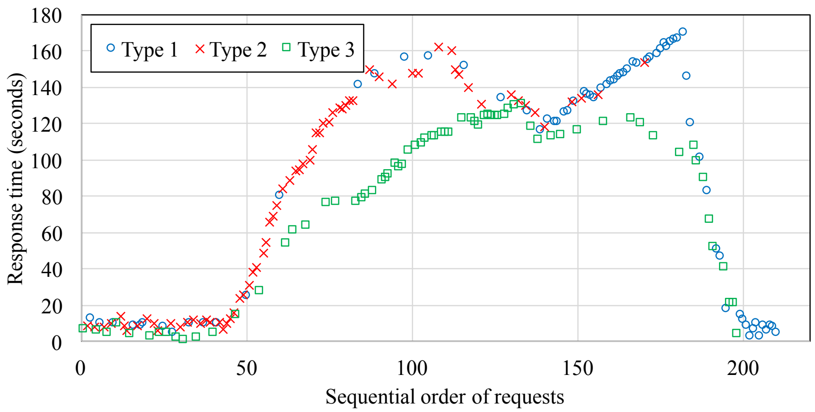


Other Tasks in Workflow: Natural Language Processing (NLP) of Free Text describing experiment; Filtering, Clustering, Tagging

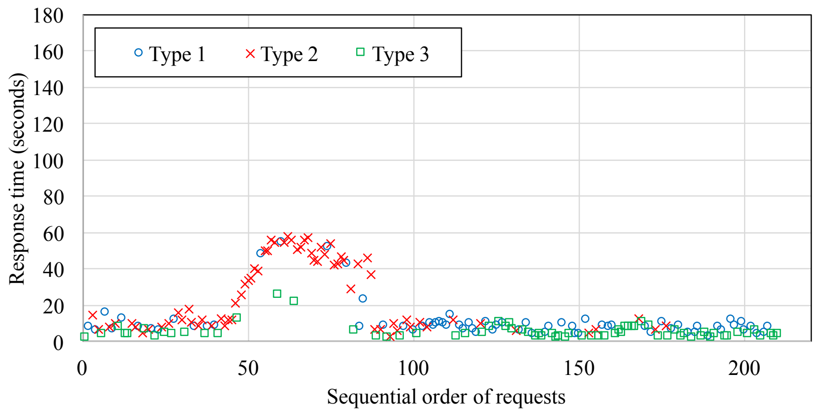
4CeeD Coordinator Implementation



Acceleration of Task Workflow Processing with Speed-Up in Response Time

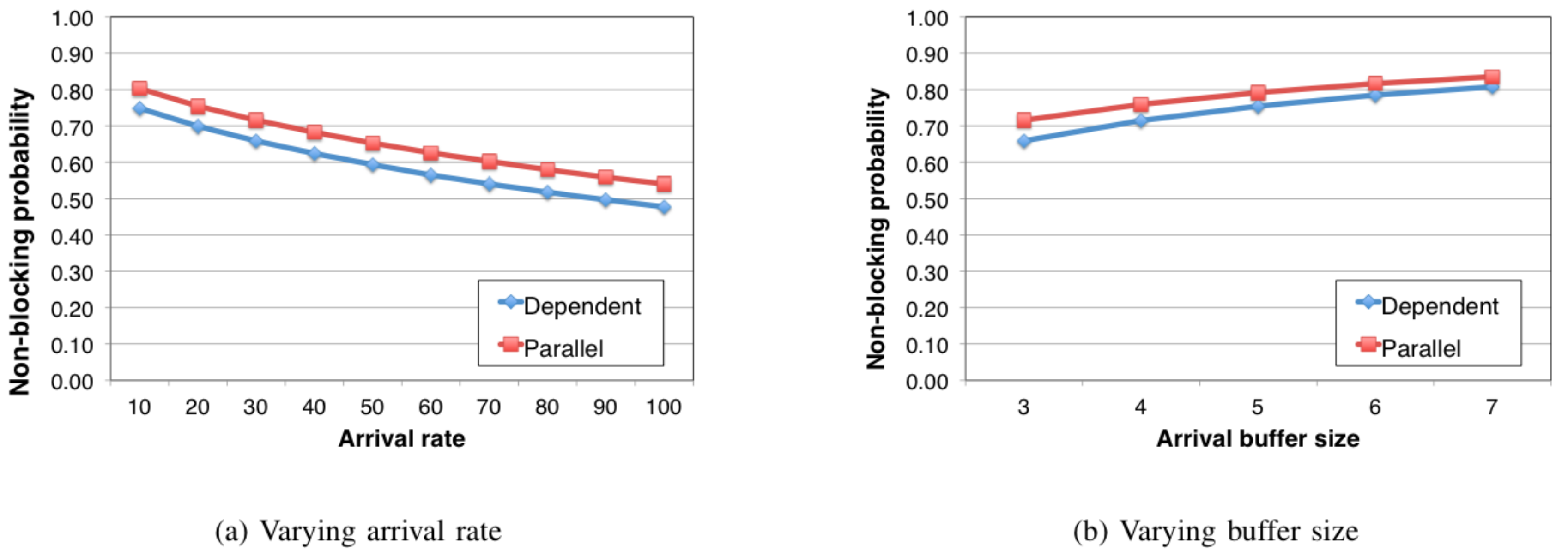


Jobs' average response time without resource adaptation



Jobs' average response time using our elastic resource adaptation

Acceleration of Jobs Processing with Efficient Global Resource Provisioning Strategies



"Parallel execution when possible"

"From global to local bottleneck"

Conclusions

- Acceleration of Lab Session for Users
  - Saving time and money (the same amount of lab data is now processed in 20 minutes instead of 60 minutes)
  - Producing more data (going from Mbytes to Gigabytes)
  - Preserving more metadata (richer metadata available)
- Acceleration of Cloud Processing on Science Data (e.g., Response time speed up from 160 sec to 60 sec)
- Developed Training and Installation Material
  - Repository and installation instructions at github.com/4ceed
  - Product page: 4ceed.github.io

Acknowledgment

- NSF ACI DIBBs Funding

Joint Work with

  - Steve Konstanty
  - Todd Nicholson
  - Phuong Nguyen
  - Tim Spila
  - Michael Chan
  - Kenton McHenry
  - Tommy O'Brien
  - Aaron Schwarz-Duval

- NSF ACI DIBBs Funding

Joint Work with

  - Paul Braun
  - Brian Cunningham
  - Roy Campbell
  - Indranil Gupta
  - Narayana Aluru
  - John Rogers