T2C2

4CeeD DIBBs Challenges and Solutions

Klara Nahrstedt

Coordinated Science Laboratory, University of Illinois at Urbana-Champaign

Panel at 1st NSF DIBBs Workshop, January 11-12, 2017

A timely and trusted curator and coordinator of scientific data

MRL FREDERICK SEITZ MATERIALS RESEARCH LABORATORY

MNTL micro+nanotechnology lab

CSL: COORDINATED SCIENCE LAB

Overarching Challenge

It takes 20 years from new material being developed to its usage in semiconductor fabrication of a device (National Academy Report)

MRL (Materials Research)

MNTL (Micro-and-Nano Technology Research)

Approach

Address Data Acquisition, Curation and Sharing prior to Publication of Results

Challenge 1

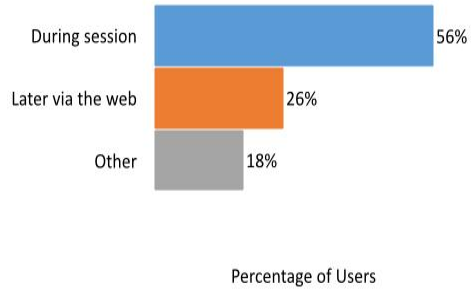
Highly Diverse Users

Challenge: Materials and Semiconductor Fabrication Device Scientists

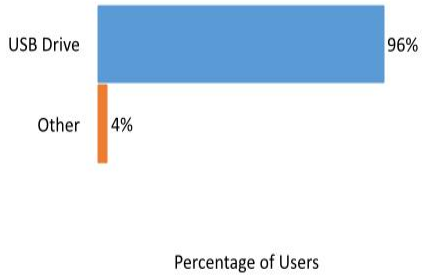
- Diverse Users - Students, Faculty, Lab Managers, Staff
- Diverse Usage of Scientific Instruments
  - Conducted Survey
  - Question examples:
    - How much data is uploaded per session?
    - How many sessions per week?
    - How and where do users capture and upload their data?
    - What instruments are used?

Approach: Scientific Data Management Survey

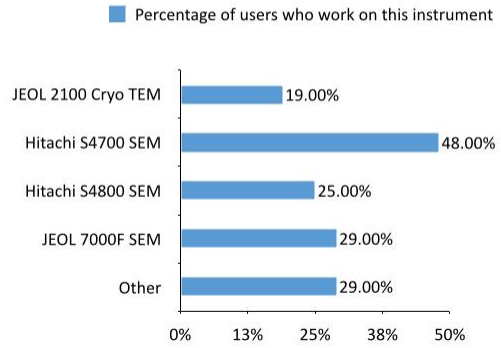
When would you prefer to upload your data?



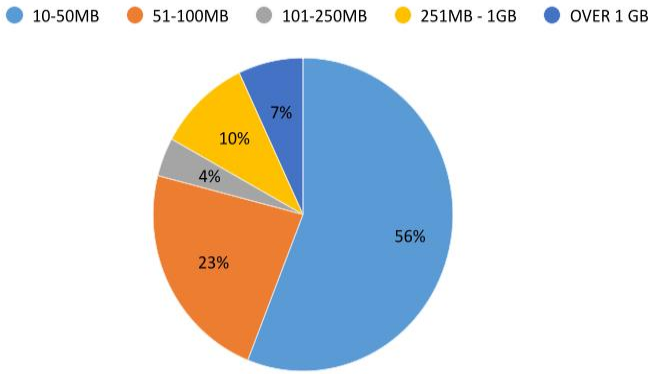
How are you currently transferring data from lab to pc



The data archiving service will start on a limited set of equipment. Do you work on any of the listed equipment?



How much data do you upload a session?



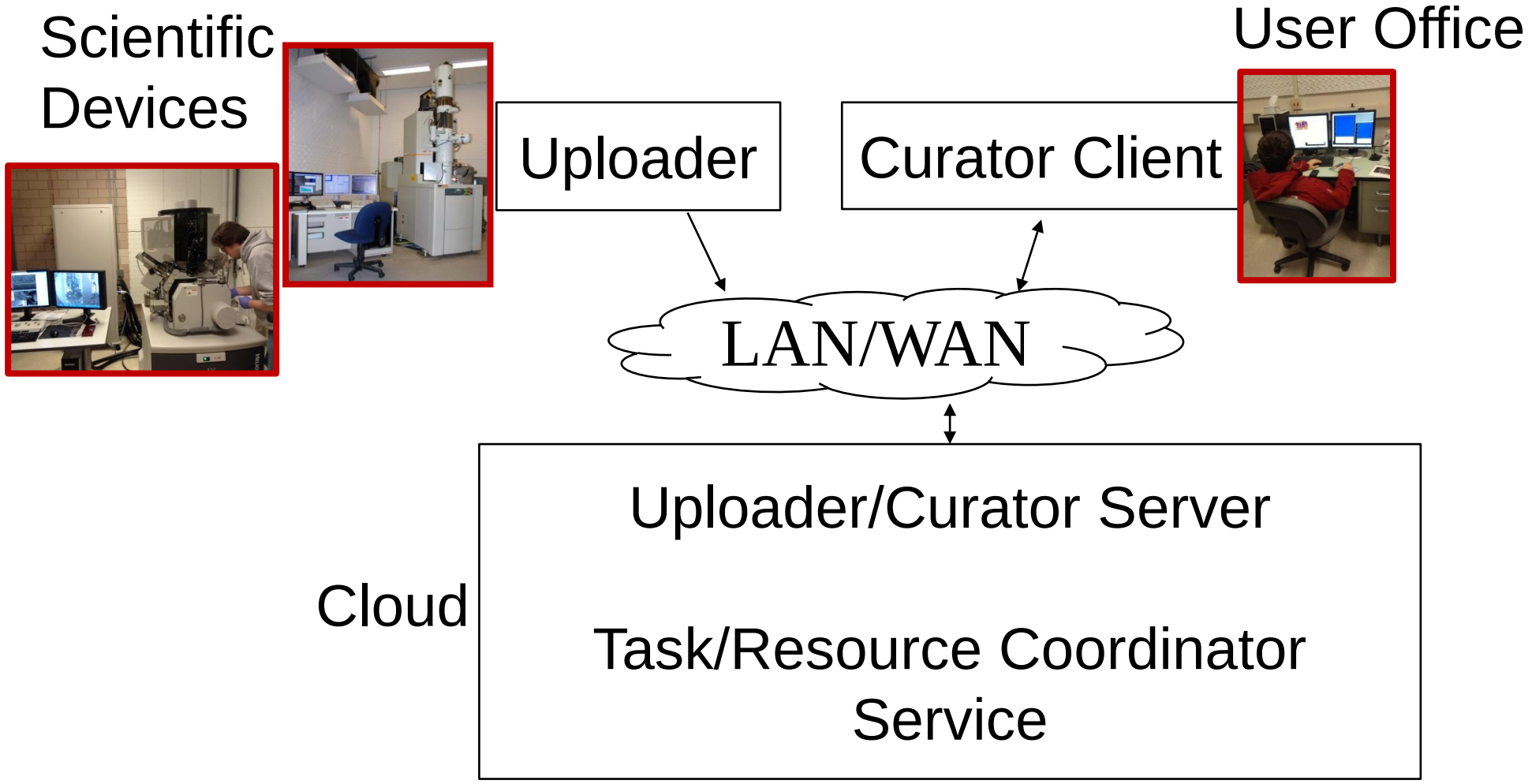
Challenge 2

System

Challenge: Real-Time and Trustworthy System

- Speed-up acquisition time at scientific instruments
- Speed-up curation time
- Accessible and simple to use to diverse users
- Real-Time and Trustworthy with cloud solutions
- Easy deployable
- Sustainable

Approach: 4CeeD Architecture

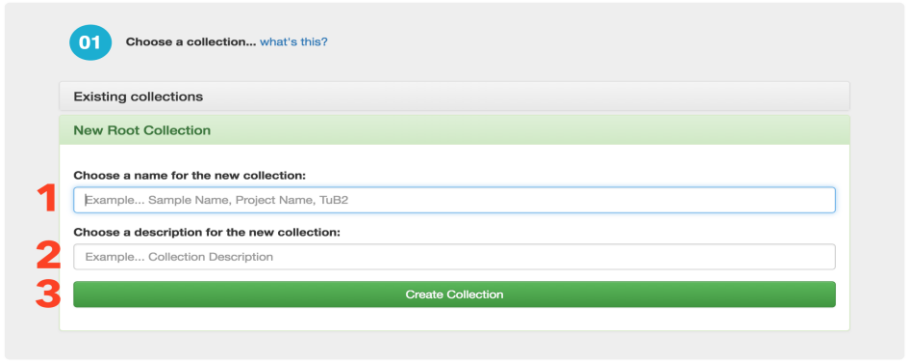


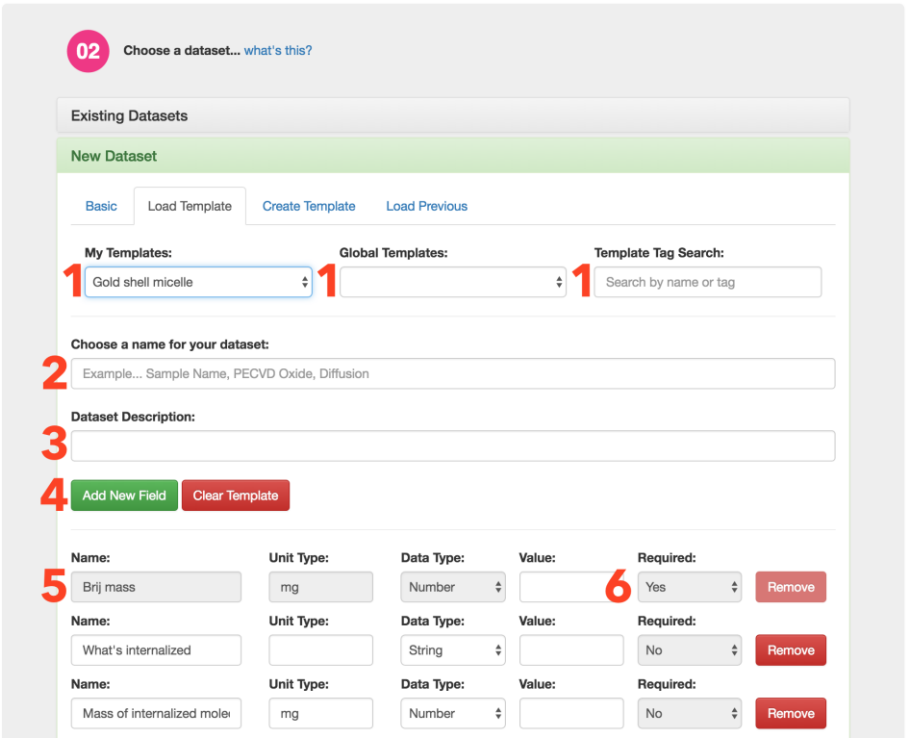
- github.com/4ceed

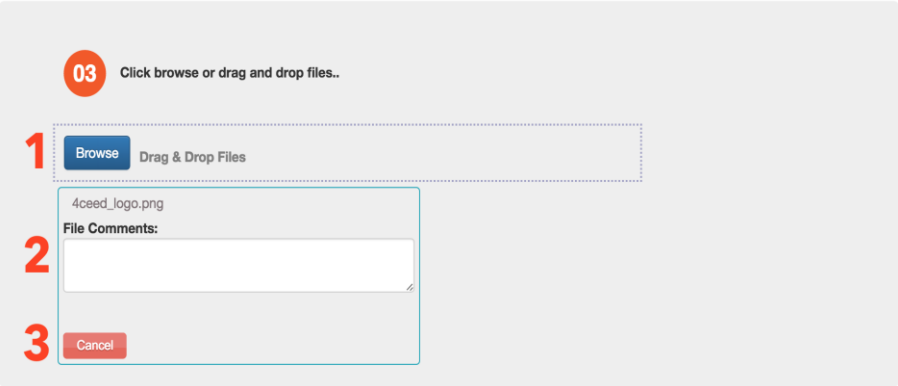
4CeeD Uploader

(Simple and Speed-Up Usage at Microscopes)

3 Simple steps, with support for advanced usage







1. Choose or select a collection.

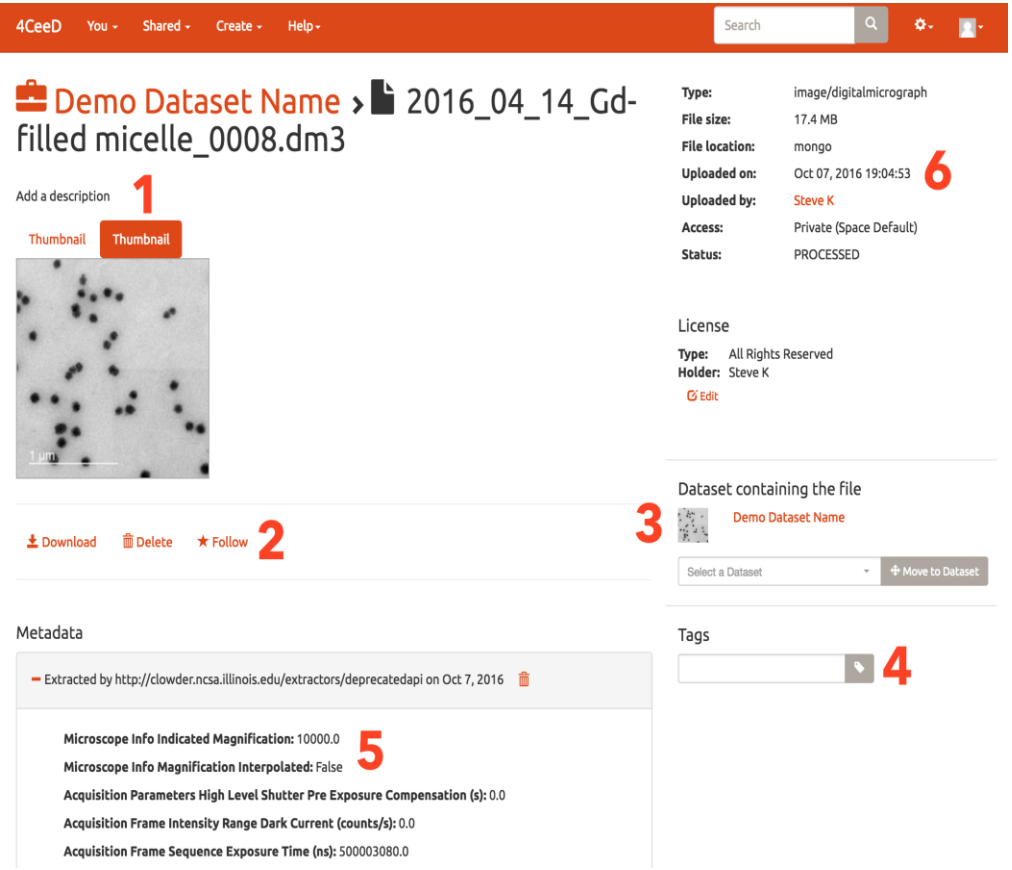
2. Load template and enter user defined metadata to create a dataset.

3. Upload files to cloud coordinator.

4CeeD Curator

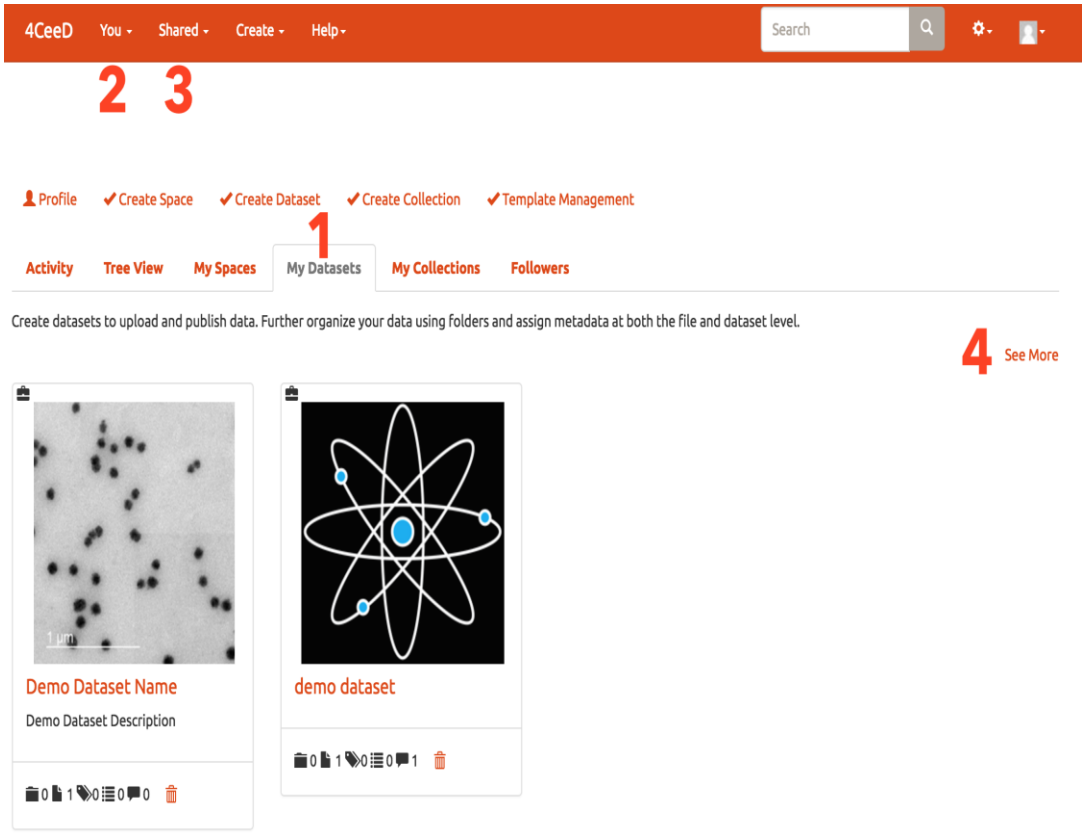
(Speed-Up Curation)

File View



[Preview, annotate, download, extracted metadata]

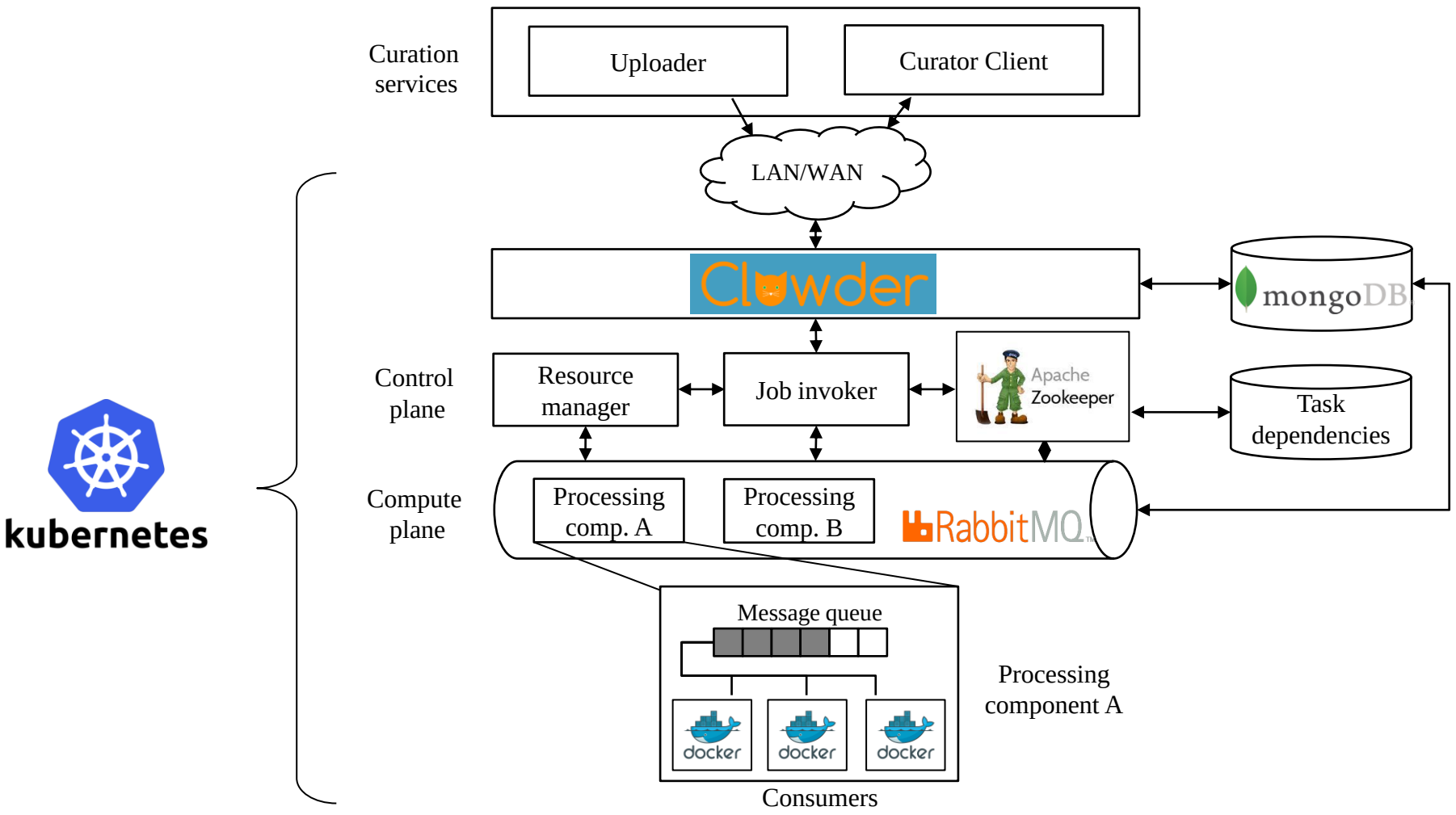
Dashboard View



[Dashboard management]

4CeeD Coordinator

(Real-Time, Trustworthy and Scalable Cloud Solutions)

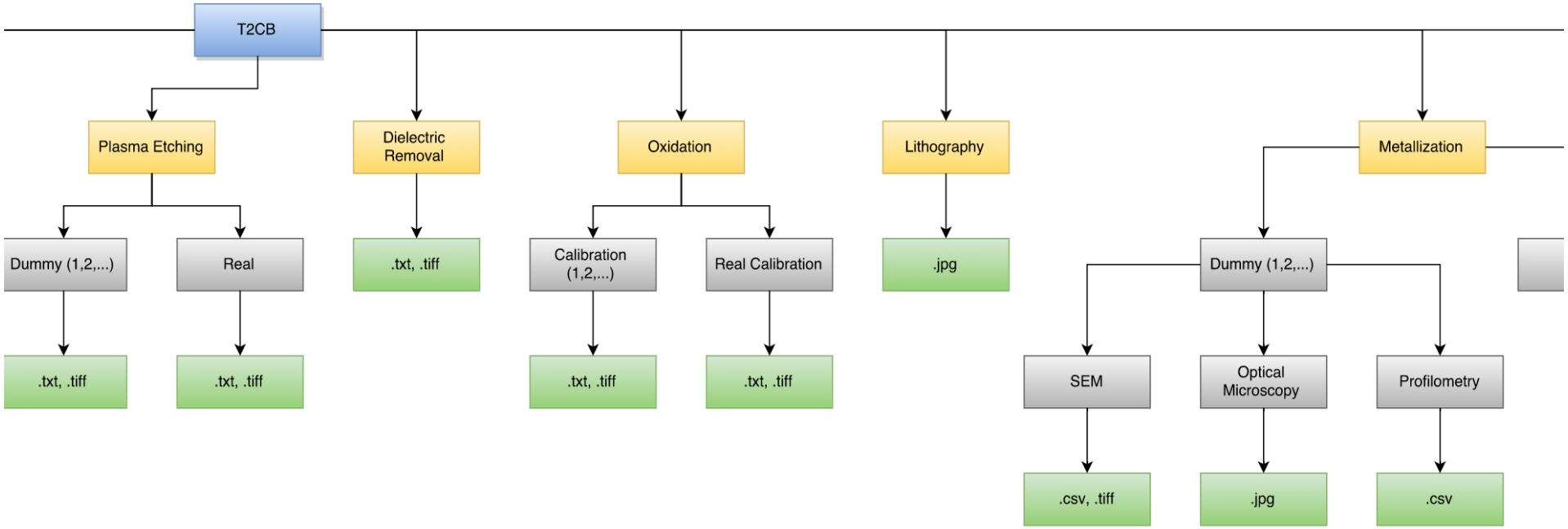


Challenge 3

Connection and Integration with other DIBBs Projects

Approach: Smart Data Management
(Integration and Enhancement of NCSA DIBBs Clowder/Brown Dog)

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



Challenge 4

Sustainability

Approach: Partner with University Engineering IT Organization

- 4CeeD private cloud for MRL/MNTL
  - Production cloud with Engineering IT staff maintenance
  - 40 TB cloud for Material Research and Semiconductor Research at UIUC
  - Purchase and Installation will start in November/December 2016
  - Operation starting January/February 2017 for both laboratories
- Aiming to have other universities and organizations to download 4CeeD (https://4ceed.github.io/)

Acknowledgement

- This research was funded by National Science Foundation NSF ACI DIBBs 1443013
- Joint Work with
  - Faculty: Roy Campbell and Indy Gupta (CS/CSL), David Nicol (CSL), Brian Cunningham (MNTL), John Rogers/Paul Braun (MRL)
  - Staff: Timothy Spila (MRL). Kenton McHenry (NCSA)
  - Research Programmers: Steve Konstanty, Todd Nicholson
  - PhD Students: Phuong Nguyen (CS), Tommy