

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

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

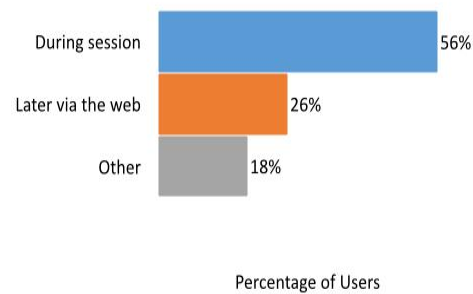
Highly Diverse Users

Challenge: Materials and Semiconductor Fabrication Device Scientists

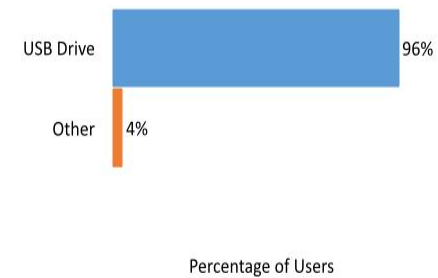
- Diverse Users - Students, Faculty, Lab Managers, Staff
- Diverse Usage of Scientific Instruments
 - Conducted Survey
 - Question examples:
 - How much data is uploaded per session?
 - How many sessions per week?
 - How and where do users capture and upload their data?
 - What instruments are used?

Approach: Scientific Data Management Survey

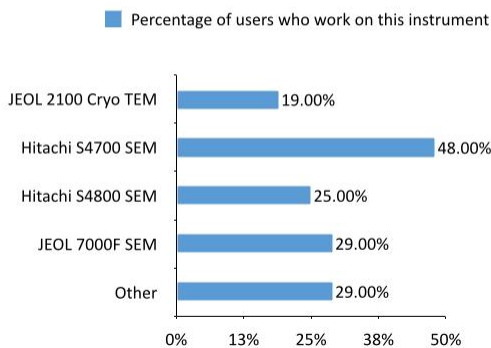
When would you prefer to upload your data?



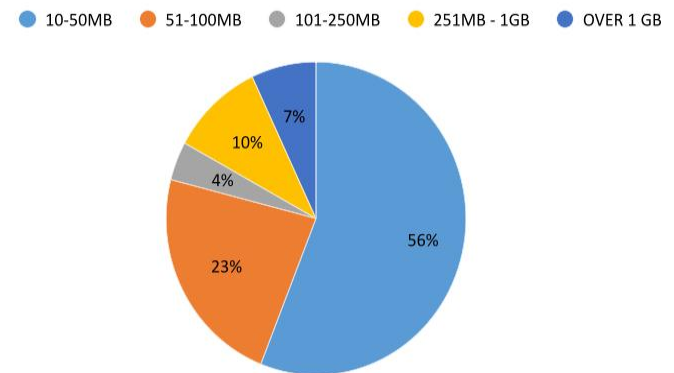
How are you currently transferring data from lab to pc



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?



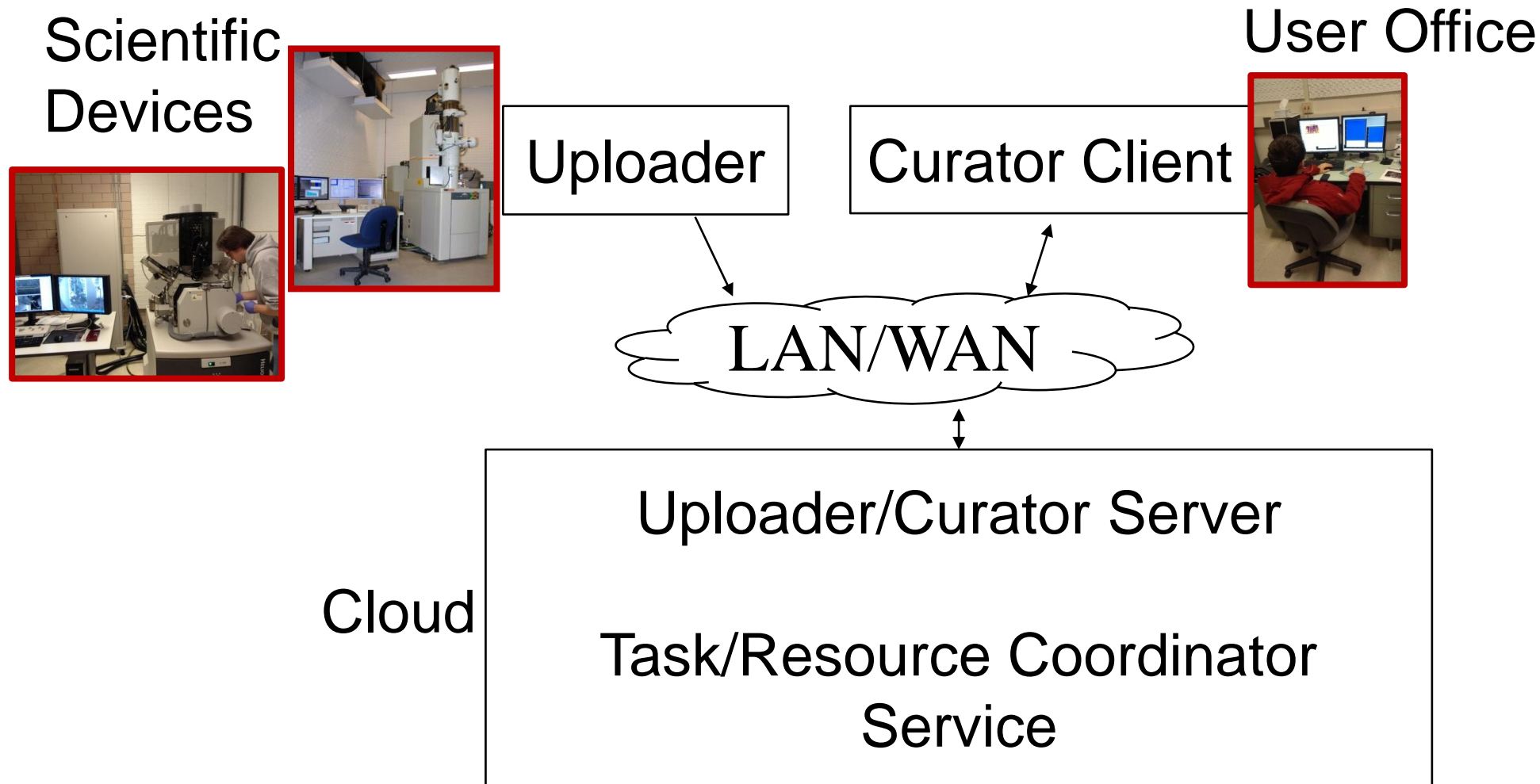
Challenge 2

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

01 Choose a collection... what's this?

Existing collections

New Root Collection

1 Choose a name for the new collection:
Example... Sample Name, Project Name, TuB2

2 Choose a description for the new collection:
Example... Collection Description

3 Create Collection

02 Choose a dataset... what's this?

Existing Datasets

New Dataset

Basic Load Template Create Template Load Previous

1 My Templates: Gold shell micelle **1** Global Templates: **1** Template Tag Search: Search by name or tag

2 Choose a name for your dataset:
Example... Sample Name, PECVD Oxide, Diffusion

3 Dataset Description:

4 Add New Field Clear Template

5 Name: Brij mass Unit Type: mg Data Type: Number Value: **6** Required: Yes Remove

Name: What's internalized Unit Type: Data Type: String Value: Required: No Remove

Name: Mass of internalized mole Unit Type: mg Data Type: Number Value: Required: No Remove

03 Click browse or drag and drop files..

1 Browse Drag & Drop Files

4ceed_logo.png

2 File Comments:

3 Cancel

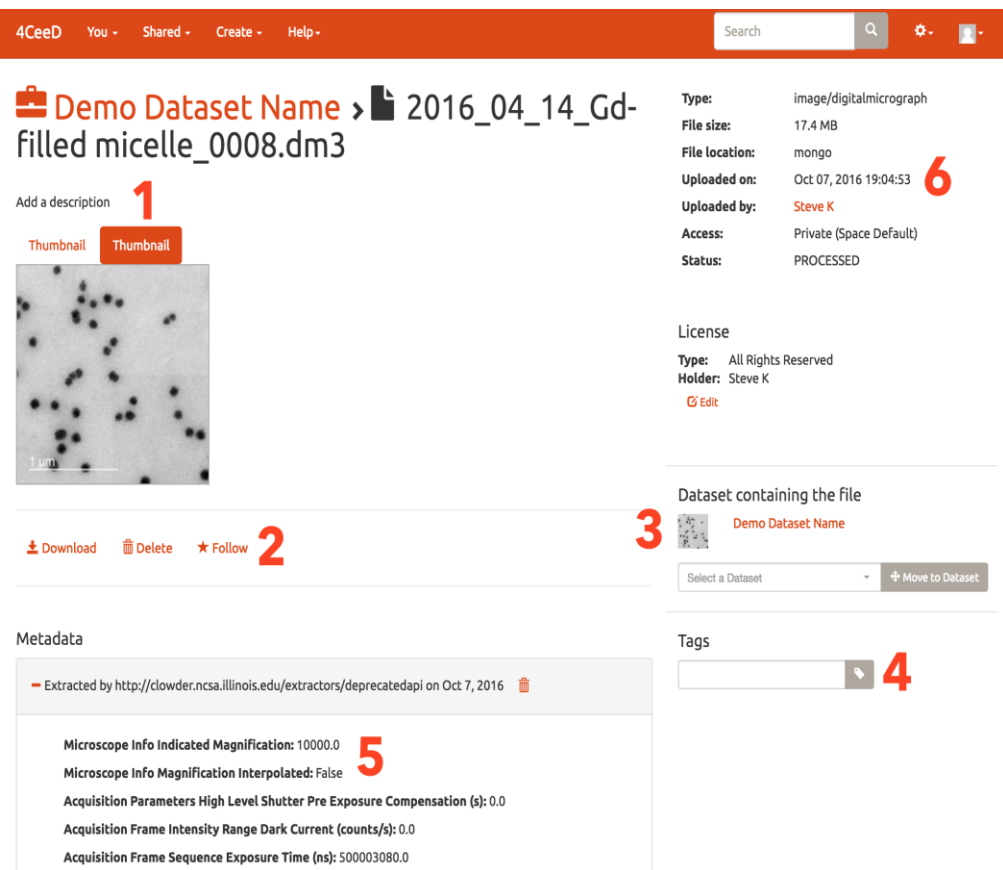
1. Choose or select a collection.

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



4CeeD You - Shared - Create - Help - Search

Demo Dataset Name > 2016_04_14_Gd-filled micelle_0008.dm3

1 Add a description

Thumbnail Thumbnail

2 Download Delete Follow

3 Dataset containing the file

4 Tags

5 Metadata

6 File details:

- Type: image/digitalmicrograph
- File size: 17.4 MB
- File location: mongo
- Uploaded on: Oct 07, 2016 19:04:53
- Uploaded by: Steve K
- Access: Private (Space Default)
- Status: PROCESSED

License: All Rights Reserved, Holder: Steve K

Dataset containing the file: Demo Dataset Name

Tags: []

Metadata: Extracted by http://cloudner.ncsa.illinois.edu/extractors/deprecatdapi on Oct 7, 2016

Microscope Info Indicated Magnification: 10000.0

Microscope Info Magnification Interpolated: False

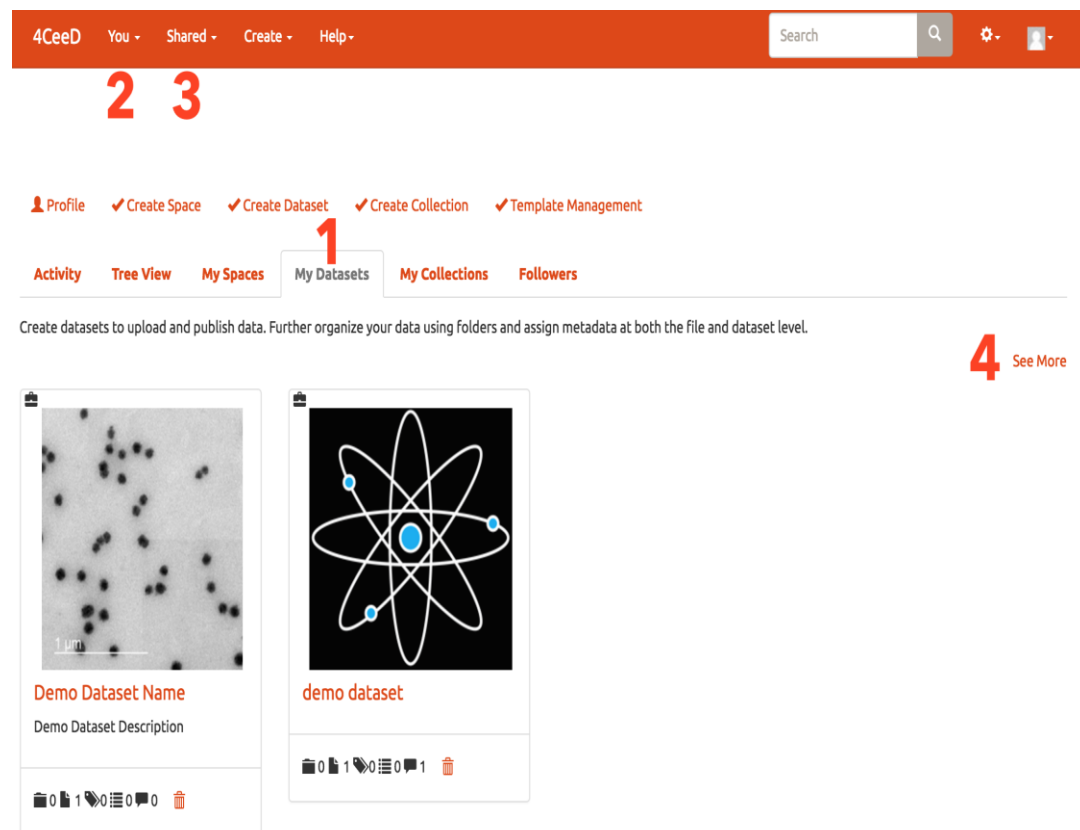
Acquisition Parameters High Level Shutter Pre Exposure Compensation (s): 0.0

Acquisition Frame Intensity Range Dark Current (counts/s): 0.0

Acquisition Frame Sequence Exposure Time (ns): 500003080.0

[Preview, annotate, download, extracted metadata]

Dashboard View



4CeeD You - Shared - Create - Help - Search

2 3

1 Profile Create Space Create Dataset Create Collection Template Management

Activity Tree View My Spaces My Datasets My Collections Followers

Create datasets to upload and publish data. Further organize your data using folders and assign metadata at both the file and dataset level.

4 See More

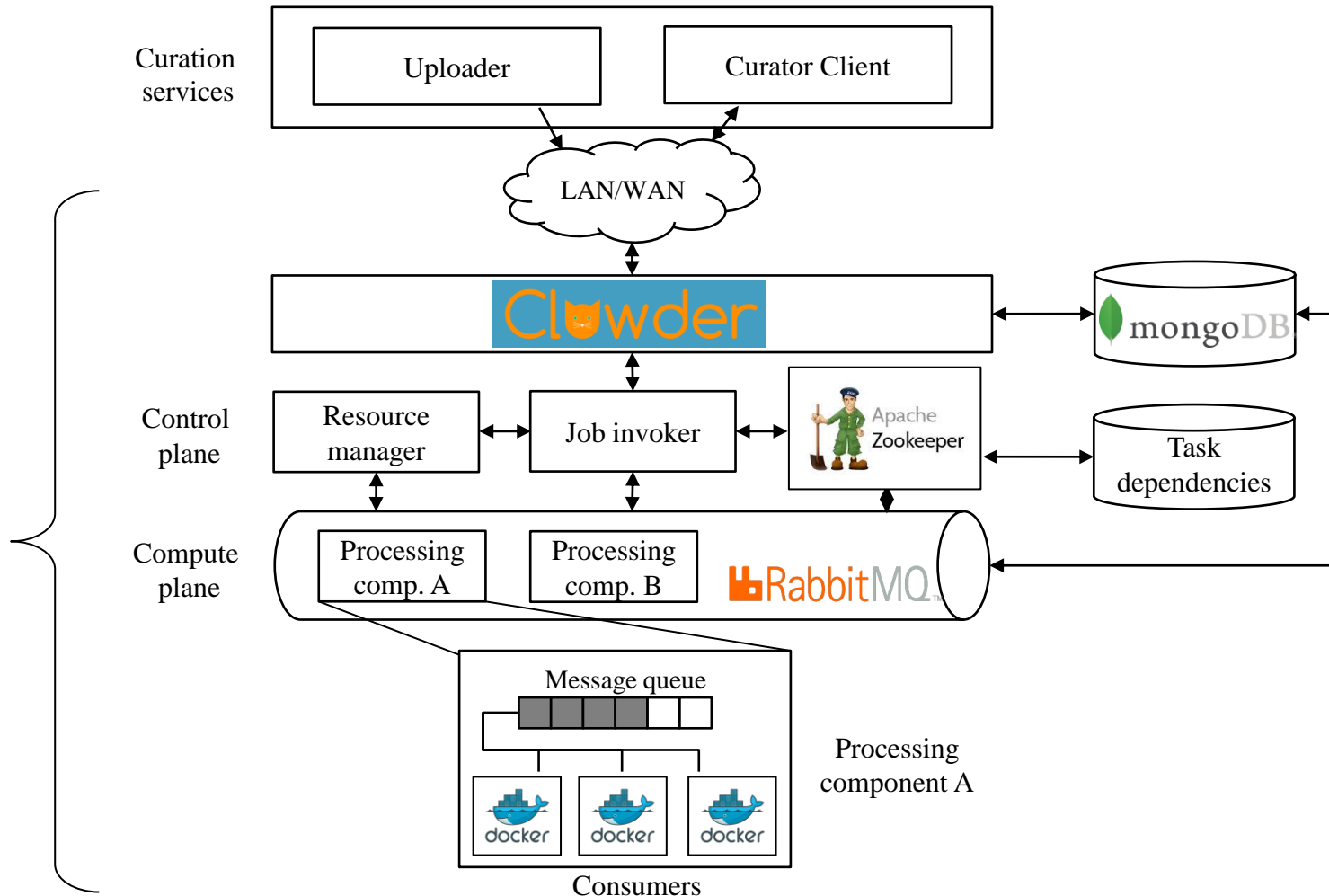
Demo Dataset Name

demo dataset

[Dashboard management]

4CeeD Coordinator

(Real-Time, Trustworthy and Scalable Cloud Solutions)

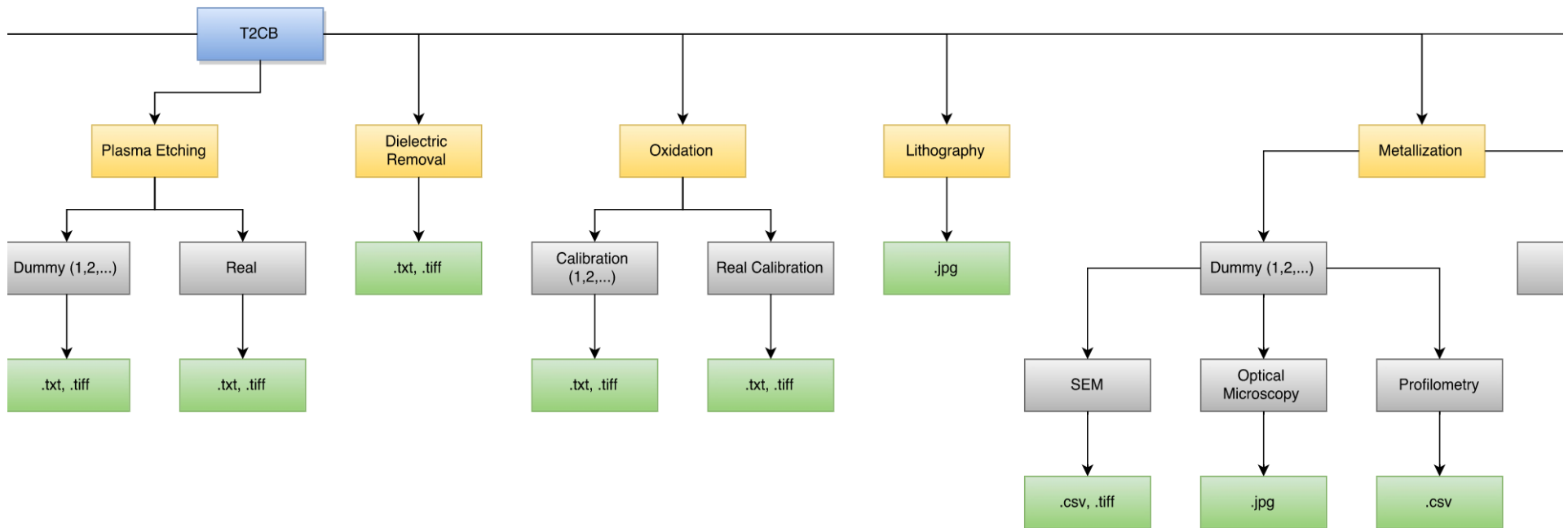


Challenge 3

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