

4CeeD: Real-Time Operating Infrastructure for Capturing, Curating, Correlating, Coordinating and Distributing Materials- related Data

Klara Nahrstedt

University of Illinois at Urbana-Champaign

<http://t2c2.csl.illinois.edu>

Co-Authors: Phuong Nguyen, Steven Konstanty, Todd Nicholson, Thomas O'Brien, Patric Su, Aaron Schwartz-Duval, Tarek Elgamal, Tim Spila, Roy Campbell, Indy Gupta, Michael Chan, Kenton McHenry, Norman Paquin,

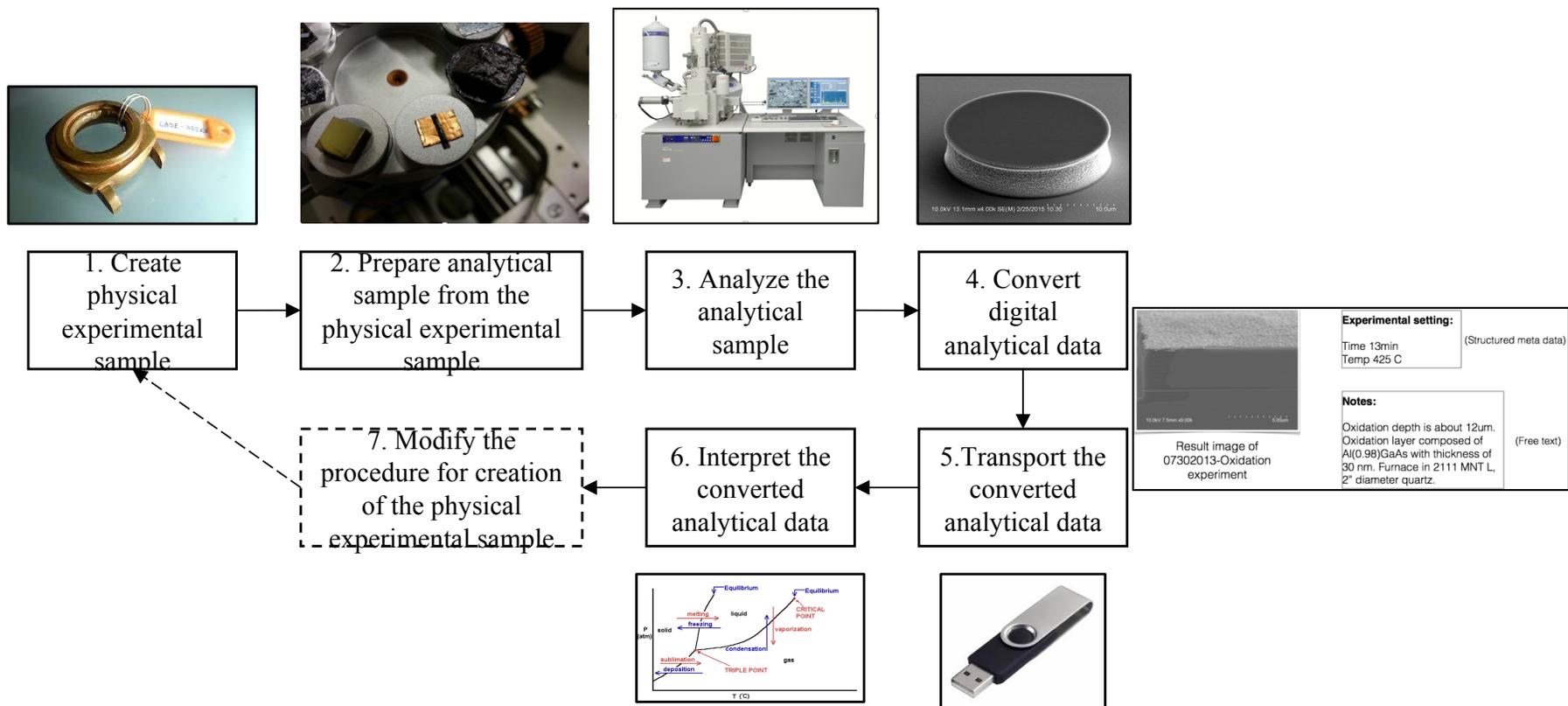


Outline

- Overall Problem Description
- 4CeeD Solution
 - System Perspective
 - User Perspective
- BRACELET in 4CeeD
 - System Perspective
- Conclusion

Background: Long process from material discovery to device fabrication

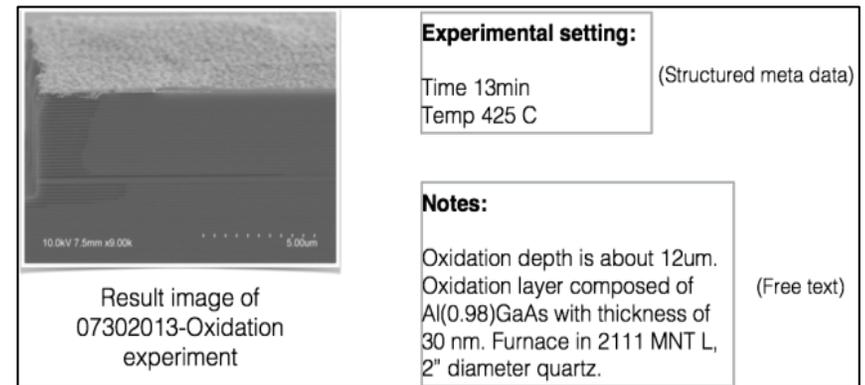
- It typically **takes 20 years** to go from the discovery of new materials to fabrication of new and next-generation devices*



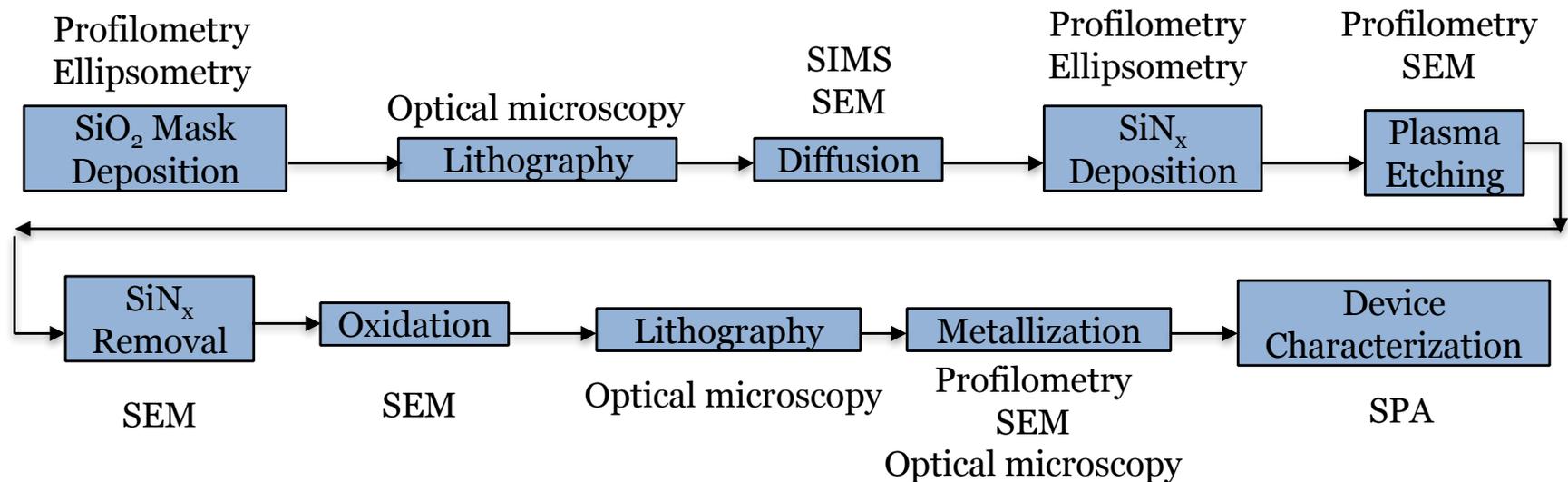
* Holdren, J.P. "Materials genome initiative for global competitiveness". National Science and Technology Council OSTP 2011.

Scientific Digital Data Acquisition and Workflow Challenges

- Data come of various types and multimodal formats
 - Each type requires a different data processing workflow



Sample output data from SEM microscopy



Current issues with scientific data capture, management and sharing



- Data acquisition & transfer
 - “*Sneaker-net*” data transfer
 - No data conversion is available during data acquisition

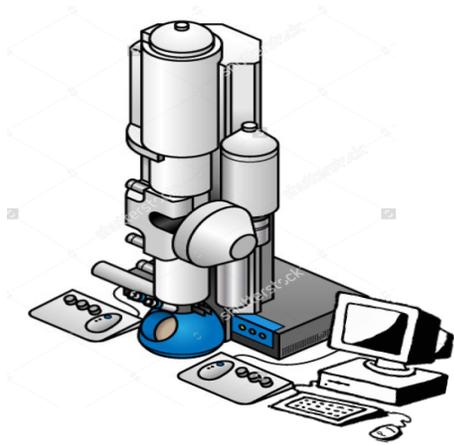


- Data management & processing
 - Using local file system, or general-purpose cloud storage services (e.g., DropBox, Google Drive, etc.)



- Data access & sharing
 - Materials science & semiconductor fabrication areas have never been digitally connected
 - Only “*best*” results and data are kept for publishing

Current State of Data Capture



**Instrument
(MRL/MNTL)**

- Fabricate experimental sample
- Prepare analytical sample
- Bring sample to instrument for analysis
- **Extract data (File conversion)**
- **Transport data to office computer**
- **Extract data (File conversion)**
- **Analyze data**
- **Transport data to office computer**
- **Analyze data**
- Repeat per iteration



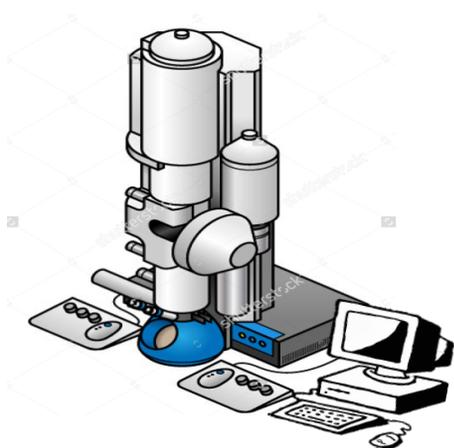
Flash Drives



Office

- Limits of home analysis
- 'Sneakernet'
- Metadata Loss
- Only what you have
- Security risk at both ends
- Excessive file name schemes
- Connect notes to data
- Very limited transport space
- Manual notes (after time delay)
- Lost or forgotten flash drives
- What metadata is important?
- Retrospective not reactive

Our Goal for Data Capture



**Instrument
(MRL/MNTL)**

- Fabricate experimental sample
- Prepare analytical sample
- Bring sample to instrument for analysis
- **Extract data (NO FILE CONVERSION)**
- **Transport data to office computer (DIRECT)**
- **Analyze data (REAL TIME)**
- Repeat per iteration



Real time interface

- High data rate with metadata
- More simple file names
- Reactive interpretation
- Easy automatic note taking
- Easy interpretation/searching
- All metadata included



Laptop



Campus PC



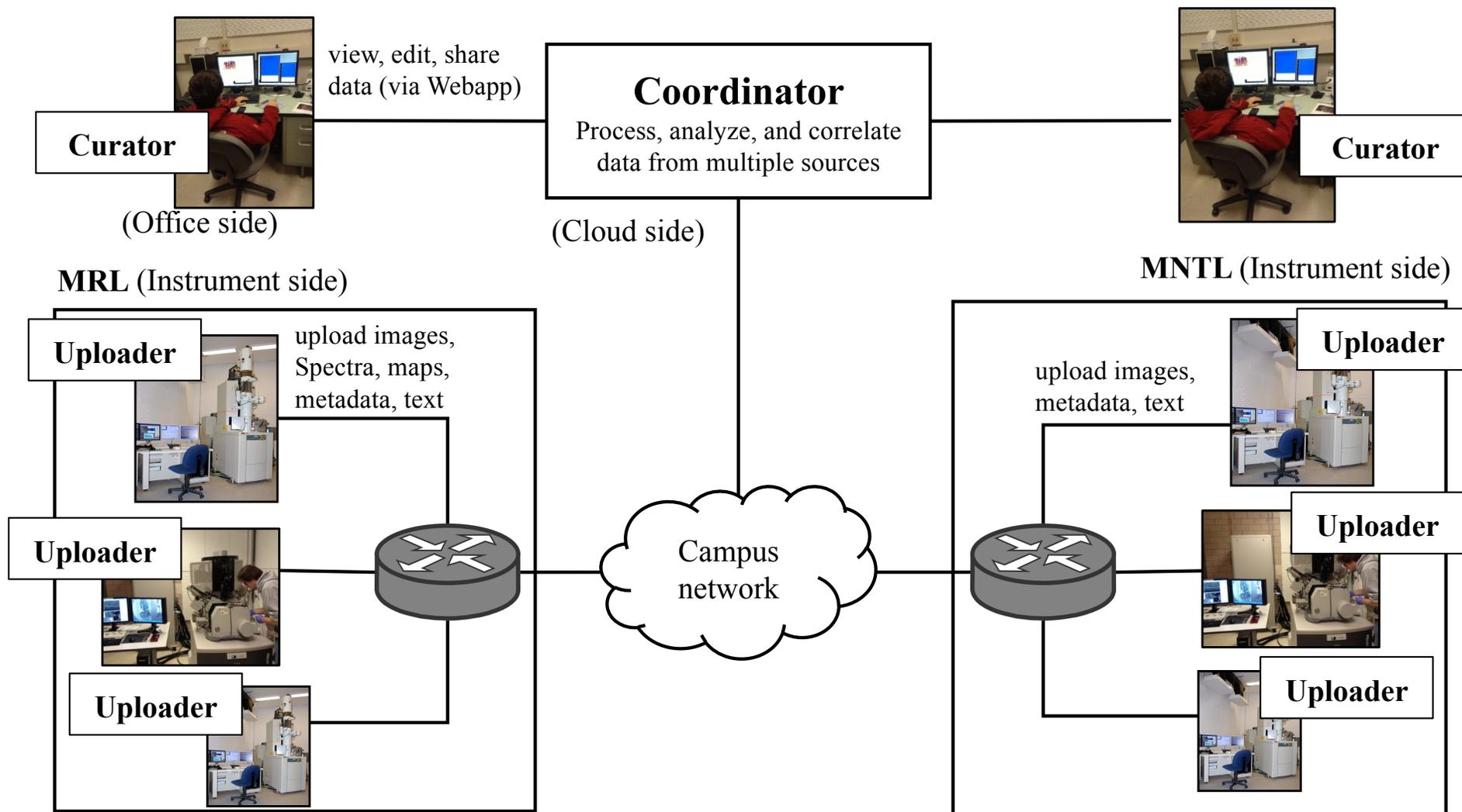
Collaborators



Office

Our Approach: 4CeeD System Perspective

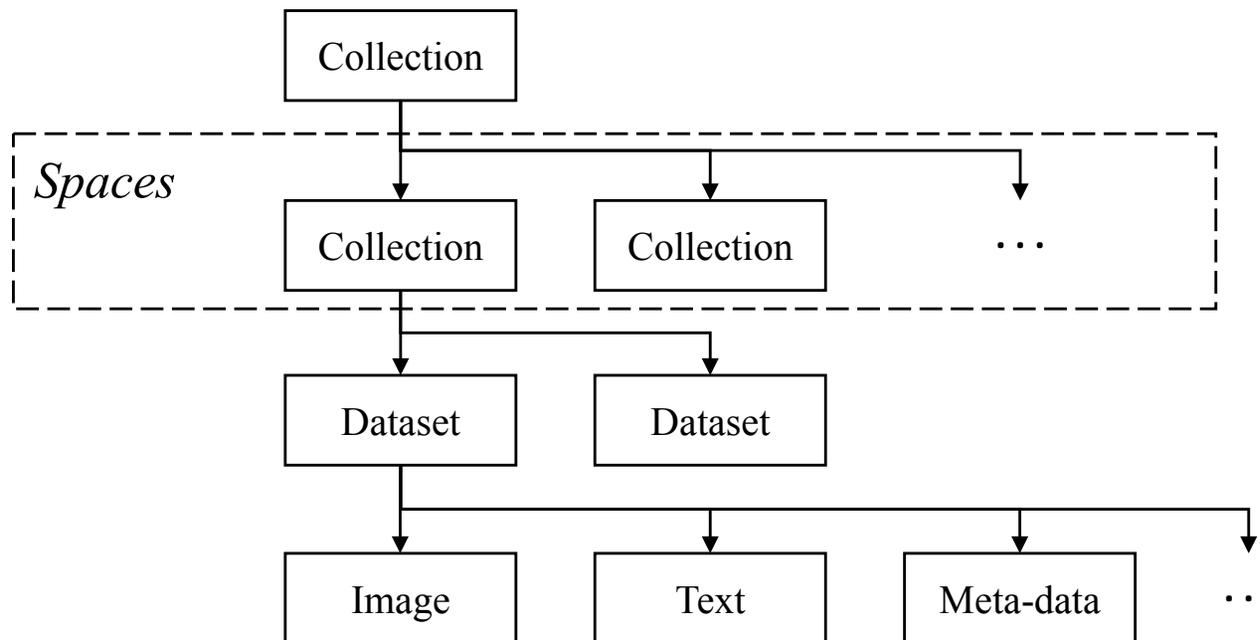
Our approach: 4CeeD Private Cloud Compute and Storage Service



* **4CeeD** stands for **C**apture, **C**urate, **C**oordinate, **C**orrelate, and **D**istribute material-related experimental data.

4CeeD Data Model

- Extendable data model
 - **Vertical** organization: Using concepts of *collections*, *datasets*, and *files* to provide flexible data organization
 - **Horizontal** organization: Using *spaces* to support sharing of data



4CeeD Data Uploader Service

- Uploader offers a simple 3-step interface, following the data model

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

Existing collections

Search your collections

Right click a collection to create a sub-collection.

- 123
- Au-shelled micelles
- Gd filled micelle project
- In vitro growth
 - Au-Brij-M
 - Au-PEG-M
- polyvilli

New Root Collection

Create collection

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

Existing Datasets

New Dataset

Basic Load Template Create Template Load Previous

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

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

Dataset Description:

Add New Field Clear Template

Name:	Unit Type:	Data Type:	Default Value:	Remove
Brij mass (mg)		String		Remove
What's internalized		String		Remove
mass of internalized molecules (mg)		String		Remove
What outer shell aminated polymer w		String		Remove

Create dataset &
associated meta-
data

03 Click browse or drag and drop files..

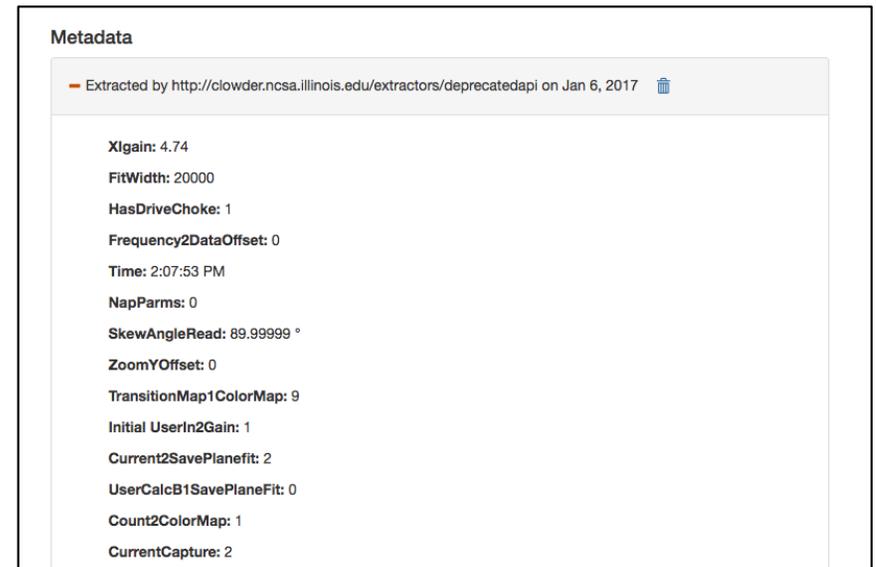
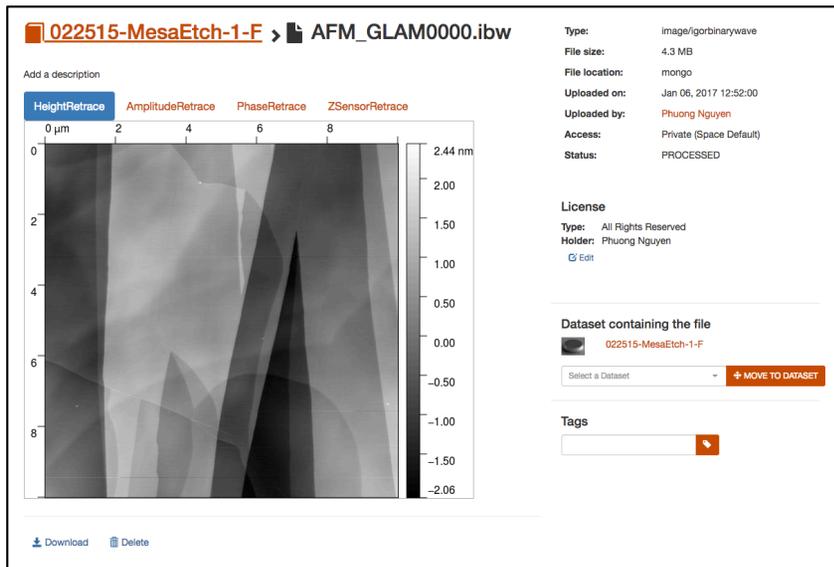
Browse Drag & Drop Files

Submit

Select raw data files
to upload

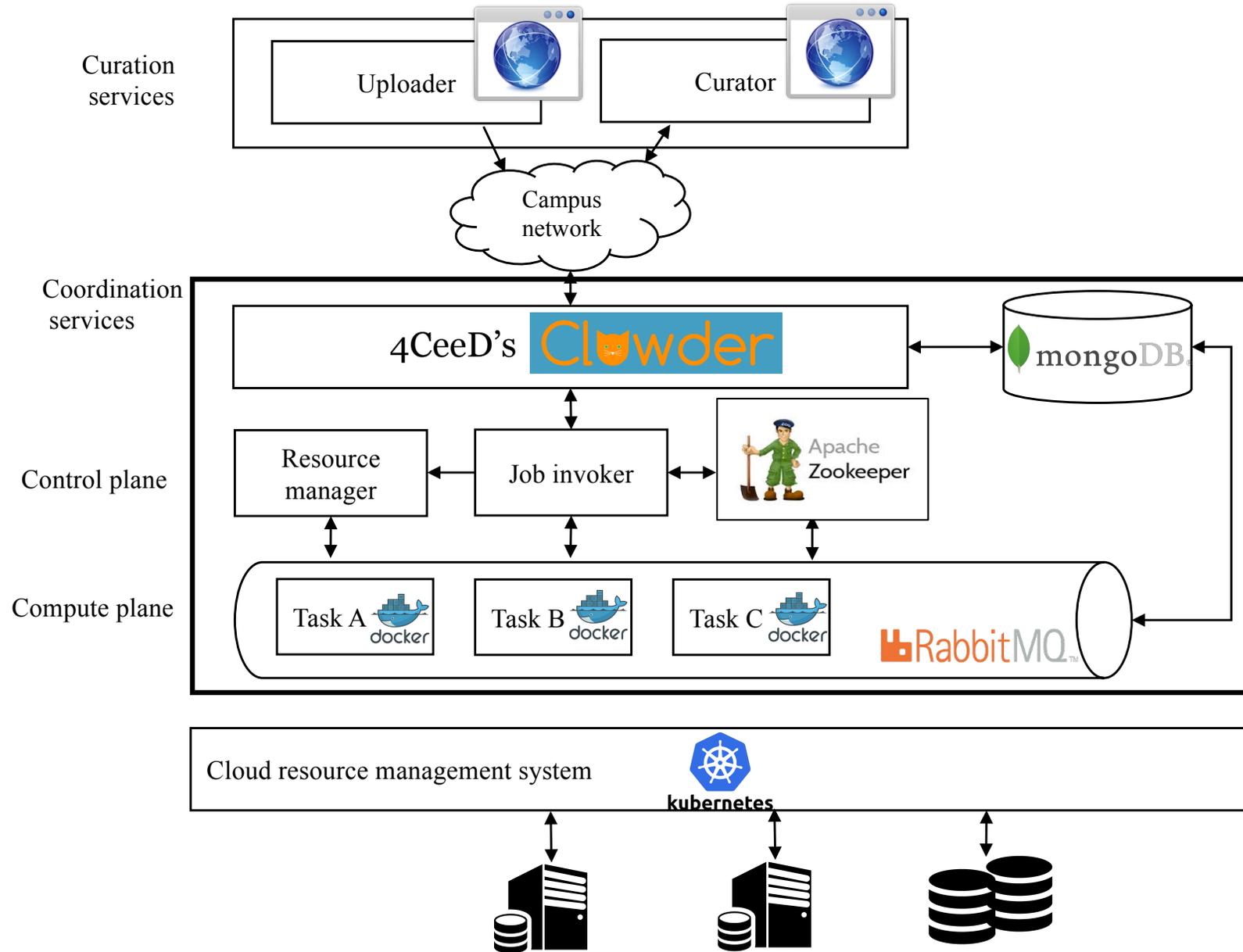
4CeeD Data Curator Service

- Uploaded raw data is processed and available for curation (tagging, annotation) and sharing
- Processed data is indexed and available for search

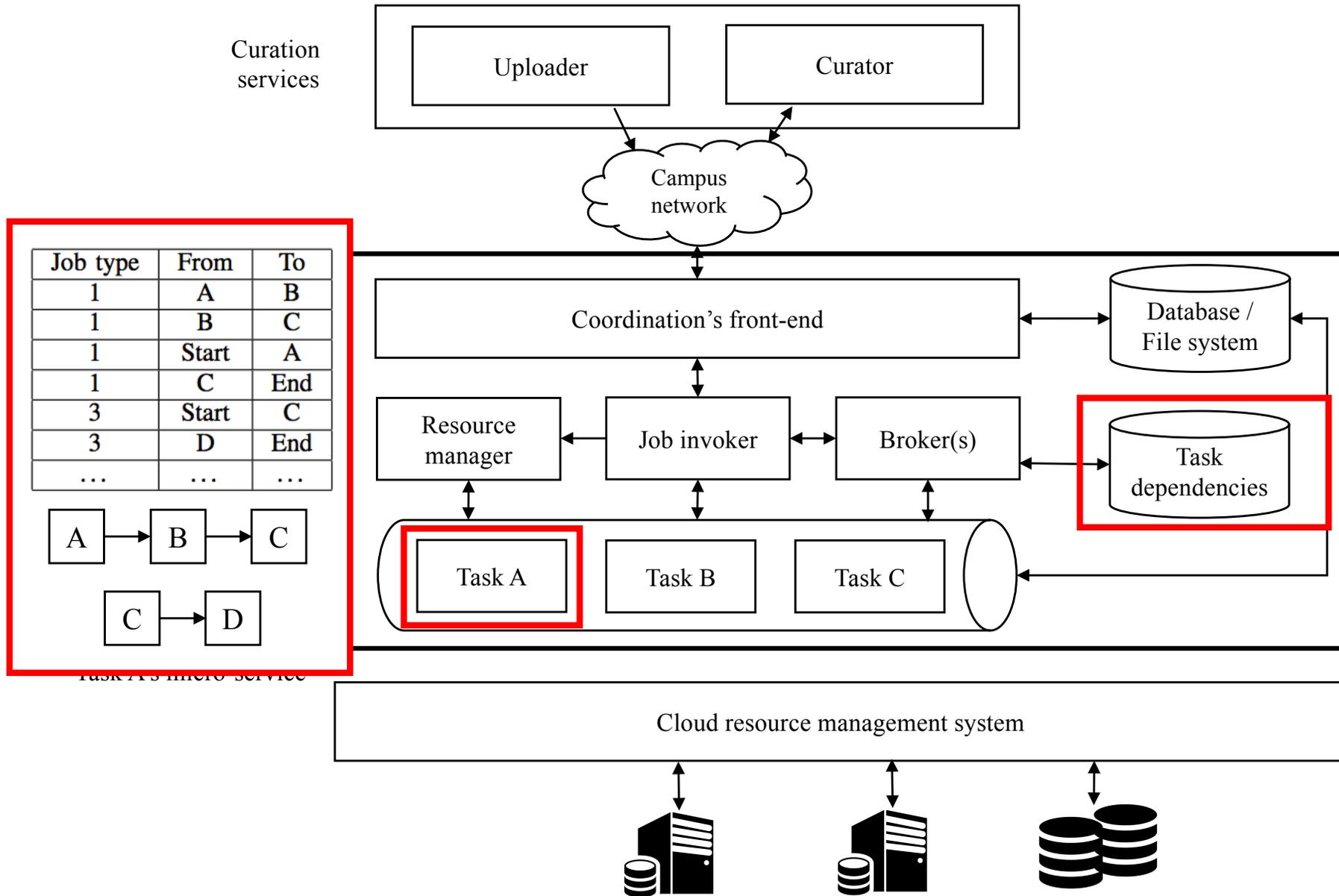


Example of processed AFM experiment data with gray-shade preview & meta-data

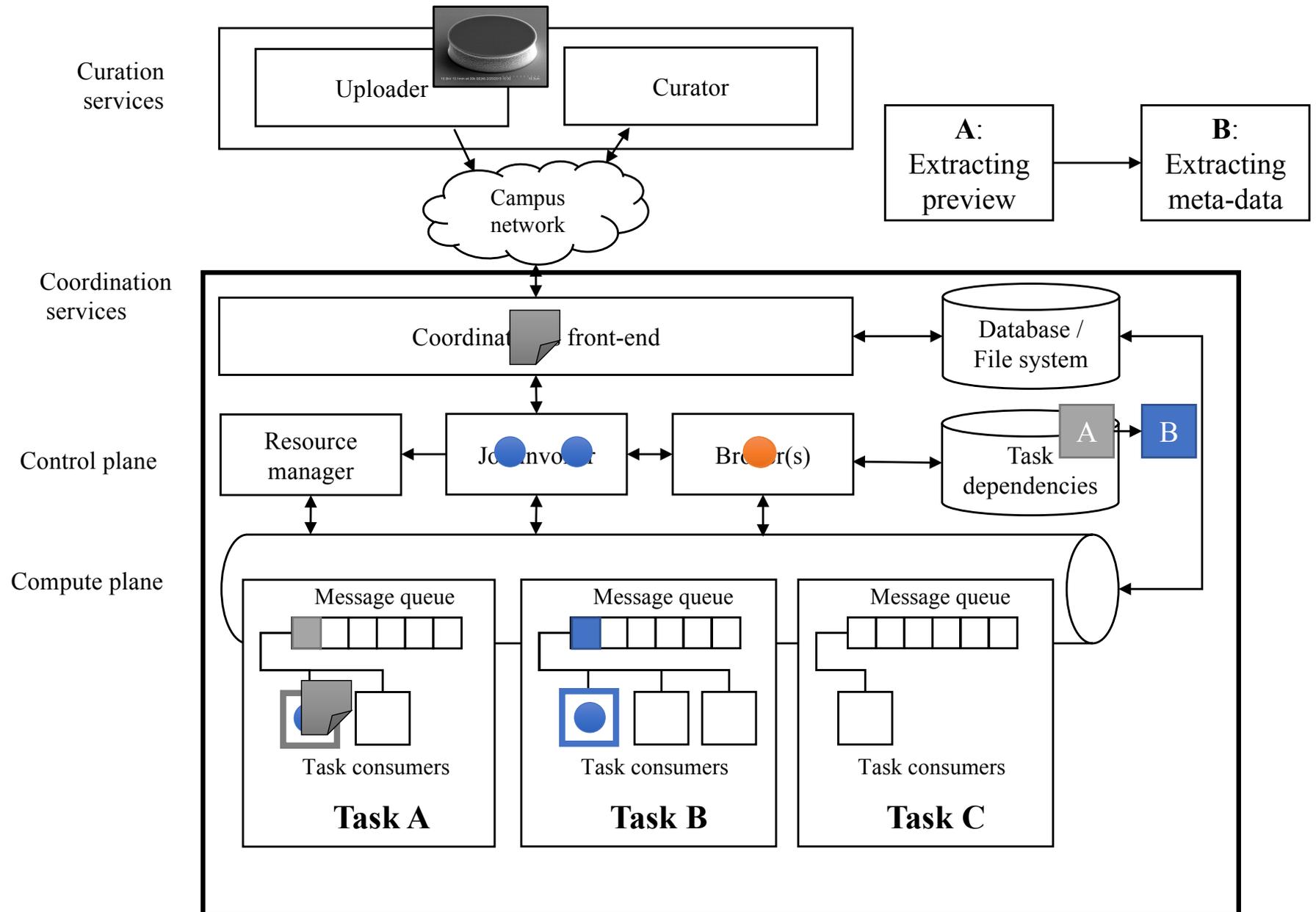
4CeeD's Coordination Service



4CeeD Coordination Service - Architecture



4CeeD's Coordination Service



Coordination Service - Consumer Scheduling (White Box Approach)

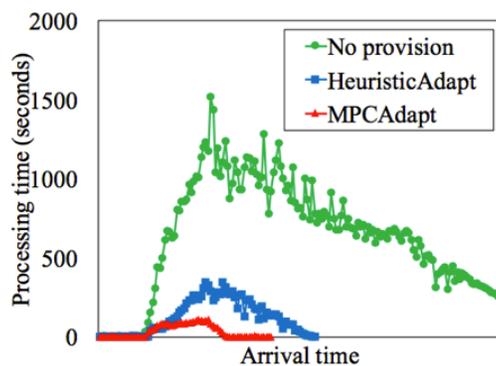
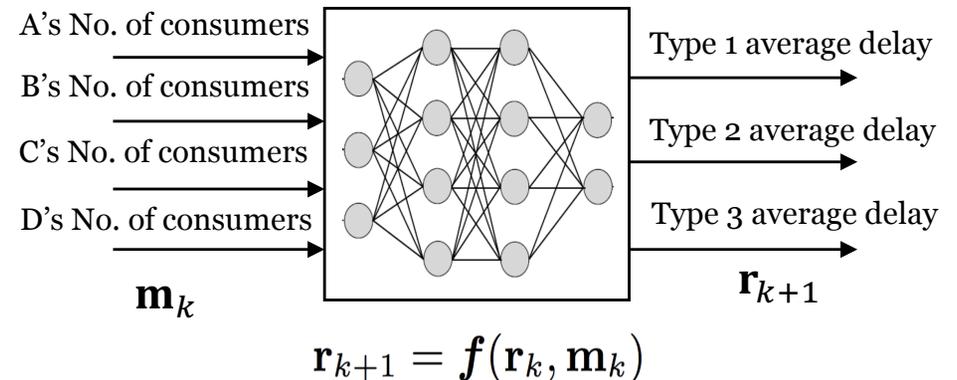
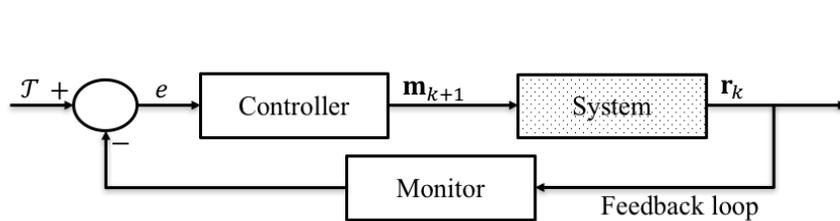
- Objective:
 - To **decide how many consumers** should be allocated to each task's **micro-service** to optimize the performance, subjected to resource constraint
- (a) white box approach: We use **work-in-progress** as **system performance metric**:
 - $WIP = \sum_j WIP_j$ (WIP is proportional to response time, via Little's law)
 - Leverage our previous work* on performance model of elastic pub/sub system to represent WIP_j as a function of number of consumers of task j -th: $WIP_j(m_j)$

$$\operatorname{argmin}_{\{m_j\}} \sum_j WIP_j(m_j)$$
$$\sum_j m_j \leq c$$

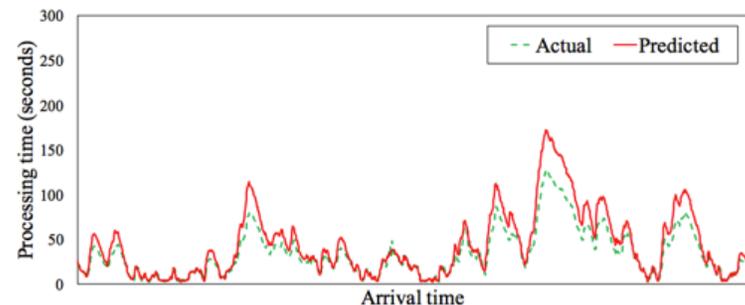
- (b) black box approach: We use model-predictive control

Resource Management (Black Box Approach)

- Robust resource management using model predictive control (from control theory) and neural network-based performance model



Effectiveness of feedback control-based adaptation algorithms when dealing with bursty workloads



Neural network-based system identification can accurately predict system performance

Our Approach: 4CeeD User Perspective (Example from MNTL)

<http://t2c2.csl.illinois.edu/demo/>

4CeeD Capabilities

- Key Features

- Usage Cases

Welcome to 4CeeD
4CeeD is a scalable data repository where you can share, organize and analyze data.

Trusted Capture Metadata Extraction Smart File Management Share and Discover

4CeeD You Shared Create Help Search

Datasets

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

✓ Create [Grid/View Icons]

<p>Silicon (Si) Cryo Etch v2 Carrier Wafer : Silicon (Si) with Crystal Bond Time : 3 min Substrate Material : Silicon (Si) Mask Material : Nickel (Ni) + Thermal Silicon Dioxide (SiO2) Gas 1 : SF6 ...</p>	<p>Silicon (Si) Bosch Etch v2 Substrate Material : Silicon (Si) Mask Material : Nickel (Ni) + Silicon Dioxide (SiO2) Gas 1 : SF6 Gas 1 Etch Flow : 100 sccm Gas 2 : ClF8 Gas 2 Etch Flow : 5 sccm Gas 1 Dep Flow ...</p>	<p>Three-Gas Etch v2 Time : 15 min Carrier Wafer : Alumina Deposition Tool : STS PECVD Dielectric Material : Silicon Nitride (SiNx) Substrate Material : Silicon (Si) Mask Material : Nickel (Ni) Gas 1 ...</p>	<p>SF6 only v1 Carrier Wafer : Alumina with Crystal Bond Time : 1 min Deposition Tool : STS PECVD Dielectric Material : Silicon Nitride (SiNx) Substrate Material : Silicon (Si) Mask Material : Nick...</p>

latasets/577beba2d9fa0ac2ef34f

4CeeD Capabilities

- Key Selected Features
 - Metadata
 - Template
 - File Sharing/Spaces
 - File Extraction

- Usage Cases

Dallesasse Group Shared Space

Shared collections of Dallesasse Group recipes. [Not your personal stuff unless you want to include it here for the group to see.]

[Delete](#) [Create Dataset](#) [Create Collections](#) [Copy Space to New Space](#)

Datasets in the Space

[View All Datasets](#)

[Manage Users](#)

[Edit Space](#)

Templates

[Create Template](#) [Update Template](#) [Delete Template](#)

Update Template

My Templates:

MNTL: PlasmaLab Freon RIE

Template name:

MNTL: PlasmaLab Freon RIE

Template tags (separate by comma):

<Example> SEM, Diffusion

[ADD NEW FIELD](#)

[CLEAR TEMPLATE](#)

Share this template with others?

Name:	Unit Type:	Data Type:	Value:	Required:	
<input type="text" value="CF4"/>	<input type="text" value="%"/>	<input type="text" value="String"/>	<input type="text" value="0"/>	<input type="text" value="No"/>	REMOVE
<input type="text" value="O2"/>	<input type="text" value="%"/>	<input type="text" value="Number"/>	<input type="text" value="50"/>	<input type="text" value="No"/>	REMOVE
<input type="text" value="SF6"/>	<input type="text" value="%"/>	<input type="text" value="Number"/>	<input type="text" value="0"/>	<input type="text" value="No"/>	REMOVE
<input type="text" value="Pressure"/>	<input type="text" value="mT"/>	<input type="text" value="Number"/>	<input type="text" value="100"/>	<input type="text" value="No"/>	REMOVE
<input type="text" value="Power (RF)"/>	<input type="text" value="%"/>	<input type="text" value="Number"/>	<input type="text" value="20"/>	<input type="text" value="No"/>	REMOVE
<input type="text" value="Time"/>	<input type="text" value="Seconds"/>	<input type="text" value="Number"/>	<input type="text" value="600"/>	<input type="text" value="No"/>	REMOVE

Key Feature: Metadata/Template

Templates

Create Template Update Template Delete Template

Update Template

My Templates:

MNTL: PlasmaLab Freon RIE

Template name:

MNTL: PlasmaLab Freon RIE

Template tags (separate by comma):

<Example> SEM, Diffusion

ADD NEW FIELD CLEAR TEMPLATE

Share this template with others?

Name:	Unit Type:	Data Type:	Value:	Required:	
CF4	%	String	0	No	REMOVE
O2	%	Number	50	No	REMOVE
SF6	%	Number	0	No	REMOVE
Pressure	mT	Number	100	No	REMOVE
Power (RF)	%	Number	20	No	REMOVE
Time	Seconds	Number	600	No	REMOVE

- Tracking and keeping fabrication recipes
- Easy to enter in all of your data at once
- MNTL Tool templates already made for users:

Trion Minilock PECVD,
STS PECVD,
Oxford Freon ICP RIE
Oxford ICP RIE
Lithography (Standard, Image Reversal)
Ellipsometer
and more!

MNTL Template for Plasmalab Freon RIE

Key Feature: Metadata/Template

2. Initial Oxidation

Created by Patrick Su
Created on Oct 29, 2017
Access: Space Default (Private) Private
 Public

All Rights Reserved Patrick Su

SiO₂, Oxide

[+ Add Files](#) [Download All Files](#) [Delete](#) [Collaborators](#)

[Files](#) [Metadata](#) [Comments \(0\)](#)

— Extracted by templateMetadata on Oct 29, 2017

[+ EDIT](#)

author: psu8@illinois.edu

name: 2. Initial Oxidation

— terms:

key: Mode

units: N/A

value: Med Dep

key: Pre-Dep Time

units: Seconds

value: 120

key: Deposition Time

units: Seconds

value: 240

key: Clean Time

units: Seconds

value: 1000

Space containing the Dataset



4CeeD Demo
4 datasets | [Remove](#)

Select a Space [+ ADD](#)

Copy Dataset to Spaces

Select a Space [+ COPY](#)

Collections containing the Dataset

Standard ECE 444 Transistor
3 datasets | [Remove](#)

Select a collection [+ ADD](#)

Tags

[TAG](#)

- Tracking and keeping fabrication recipes
- Easy to enter in all of your data at once
- MNTL Tool templates already made for users:

Trion Minilock PECVD,
STS PECVD,
Oxford Freon ICP RIE
Oxford ICP RIE
Lithography (Standard, Image Reversal)
Ellipsometer
and more!

Metadata for SiO₂ Deposition

Key Feature: Group Spaces/File Extraction

Dallesasse Group Shared Space

Shared collections of Dallesasse Group recipes. [Not your personal stuff unless you want to include it here for the group to see.]

[Delete](#) [Create Dataset](#) [Create Collections](#) [Copy Space to New Space](#)

Datasets in the Space

[View All Datasets](#)

The screenshot displays a grid of six dataset cards. Each card contains a title and a '0 Files' indicator at the bottom. The titles are:

- 10 Ratio 0.6 (Wafer 7)
- 9 MF Ratio 0.4 (Wafer 8)
- 8 MF Ratio 0.5 (Wafer 4)
- 7 MF Ratio 0.25 (Wafer 2)
- 6 Oxford ICP RIE InP Etch (1.5 um)
- 5 PG Remover / Degrease (no methanol) / O2 Plasma / Alpha Step

[Manage Users](#)

[Edit Space](#)

Copy Contents of Space to existing Space

Select a Space

Shared Spaces

- Group spaces facilitate updated coordination amongst collaborators
- Fabrication flows and experiments are always up to date
- Easy visual analysis of all updated data

Key Feature: Group Spaces/File Extraction

4CeeD Zip Uploader

Do you want to place this data in a shared space?

Yes No

Choose a shared space... [What's this?](#)

Existing spaces

Your Shared Spaces:

4CeeD Demo

Create New Shared Space

BROWSE Drag & Drop Files

The contents and metadata of large zip files may take time to populate. An email will be send to you when this is complete.

- Zip Uploader and File Extractor allows easy transition to 4CeeD
- Capable of extracting large files (record is 30 GB from single user)

4CeeD Zip Uploader

4CeeD Capabilities

■ Key Features

■ Usage Cases

- Fabrication Flow
- Process Analysis
- Device Archiving

Welcome to 4CeeD

4CeeD is a scalable data repository where you can share, organize and analyze data.



Trusted Capture



Metadata Extraction



Smart File Management



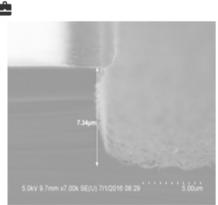
Share and Discover

4CeeD You ▾ Shared ▾ Create ▾ Help ▾

Datasets

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

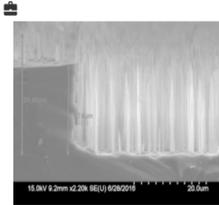
[Create](#)  



Silicon (Si) Cryo Etch v2

Carrier Wafer : Silicon (Si) with Crystal Bond
Time : 3 min
Substrate Material : Silicon (Si)
Mask Material : Nickel (Ni) + Thermal Silicon Dioxide (SiO2)
Gas 1 : SF6
...

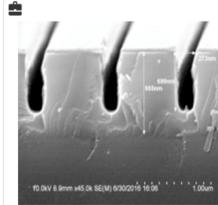
   6  0  1 



Silicon (Si) Bosch Etch v2

Substrate Material : Silicon (Si)
Mask Material : Nickel (Ni) + Silicon Dioxide (SiO2)
Gas 1 : SF6
Gas 1 Etch Flow : 100 sccm
Gas 2 : C4F8
Gas 2 Etch Flow : 5 sccm
Gas 1 Dep Flow ...

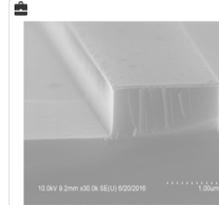
  4  0  1 



Three-Gas Etch v2

Time : 15 min
Carrier Wafer : Alumina
Deposition Tool : STS PECVD
Dielectric Material : Silicon Nitride (SiNx)
Substrate Material : Silicon (Si)
Mask Material : Nickel (Ni)
Gas 1 ...

  11  0  1 



SF6 only v1

Carrier Wafer : Alumina with Crystal Bond
Time : 1 min
Deposition Tool : STS PECVD
Dielectric Material : Silicon Nitride (SiNx)
Substrate Material : Silicon (Si)
Mask Material : Nick...

  27  0  0 

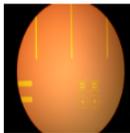


[latasets/577beba2df6fa00ac2ef34f](#)

Usage Case: Fabrication Flow

Child Collections in the Collection

[✓ Create Child Collection](#)



1-Mesa Lithography

Optical lithography with S1813 on SiO2. Training with Patrick Su.

 1 Dataset

[✕ Remove](#)

Created by [Thomas O'Brien](#)

Created on Oct 03, 2016



2-Mesa Mask Definition

Created by [Thomas O'Brien](#)

Created on Oct 03, 2016

 1 Dataset

[✕ Remove](#)



3-Stripe Planarization

Created by [Thomas O'Brien](#)

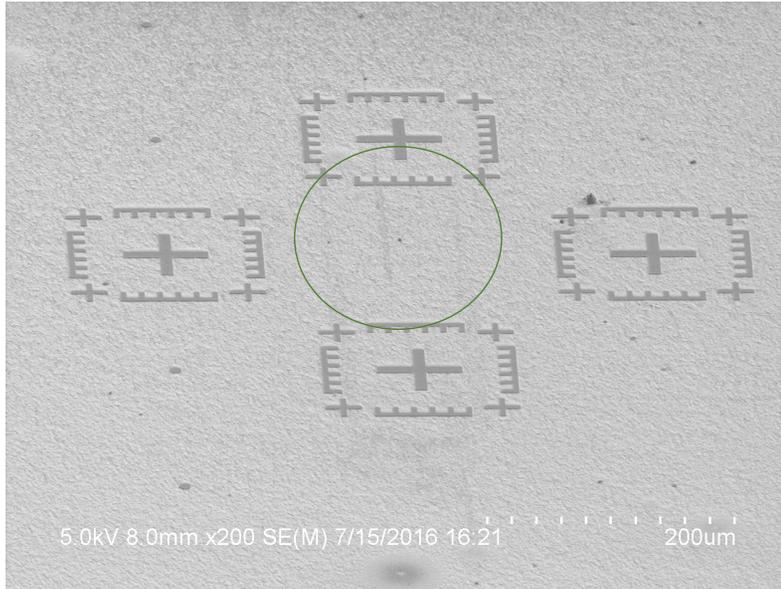
Created on Oct 12, 2016

 3 Datasets

[✕ Remove](#)

- Allows easy track of fabrication process flows
- Facilitates ramp-up for new members on projects/group
- Visual thumbnail image is an easy way to verify if their sample is on track

Usage Case: Process Analysis



- 4CeeD digitalizes data that can be searched
- Facilitates failure diagnostics
- Data that extends back to the first user of 4CeeD
- Baked on polymer example

4CeeD You Shared Create Help

Search Results baked polymer

Datasets



Name: Diffusion Mask Definition
Descriptions: SiNx diffusion mask definition. Images show etching failure due to baked on polymer, most likely residual from PR removal.
Collection name(s): Failure Example

Usage Case: Device Archiving

Collections

Use collections and sub-collections to organize multiple datasets and their associated files.

🌲 Tree View

The screenshot displays a grid of collection cards. Each card has a title, a description, and a summary of child collections and datasets. Below the grid is a navigation bar with five tabs: Activity, Tree View, My Spaces, My Collections, and My Datasets. The 'Tree View' tab is currently selected.

Collection Name	Child Collections	Datasets
Standard ECE 444 Transistor	0	3
GaN MZM 002	0	4
GaAs MZM 001	0	2
SiNx - Si Stress Measurement Collection (4 in Si Wafer)	2	0
VCSEL001	0	3
D151- LateralDiffusion_GaAs	0	0

Navigation Bar: Activity | Tree View | My Spaces | My Collections | My Datasets

Group/Projects

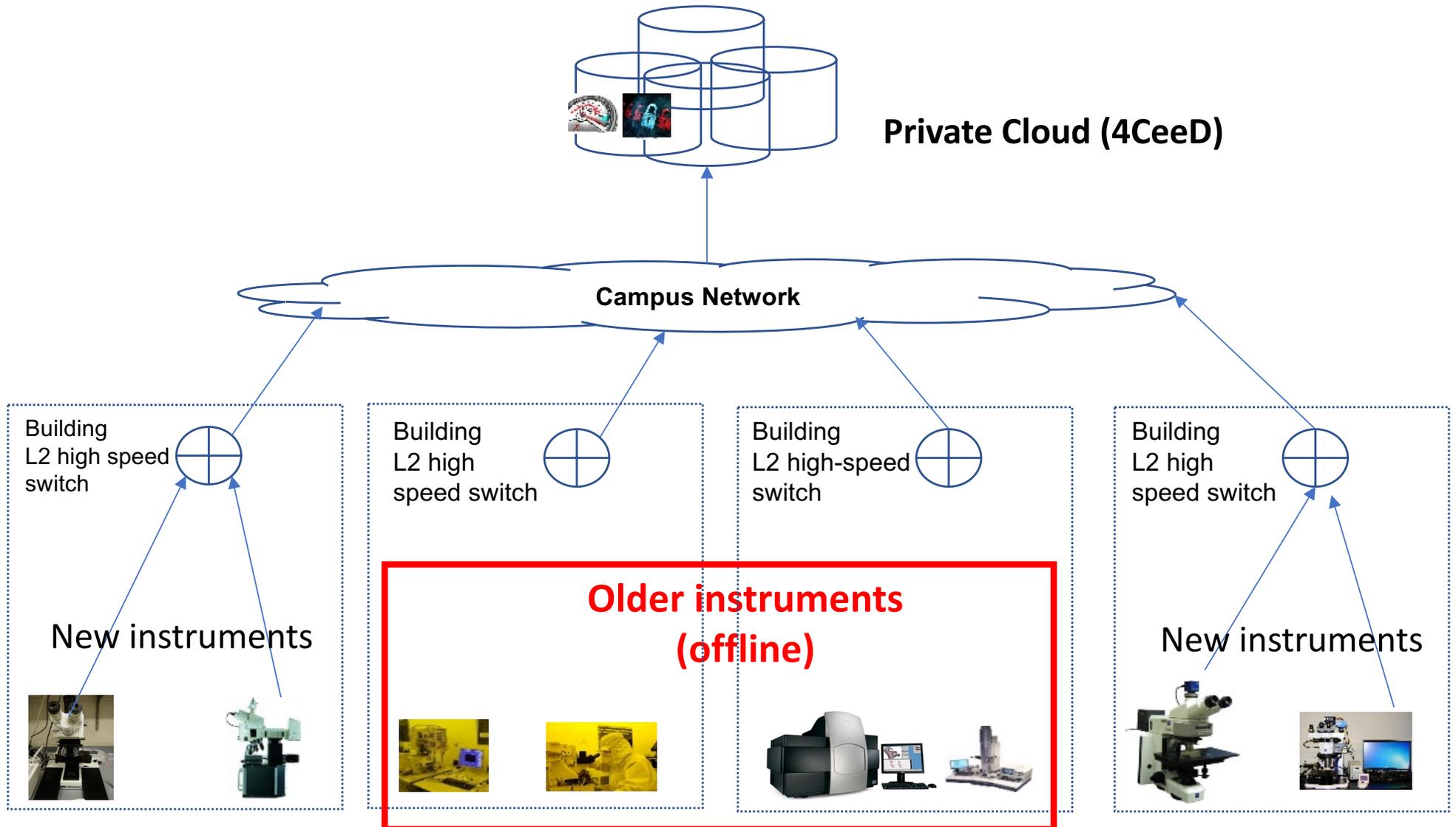
Devices/Experiment

Process Flow/Experiment Step

- Easily store old device fabrication flows
- Can be referred to by newer generation of group members
- Personal File Hierarchy:
Spaces: Group/Project
Collections: Devices
Dataset: Process Flow

BRACELET in 4CeeD

Current situation in campus cyber- infrastructure



Current challenges with old instruments

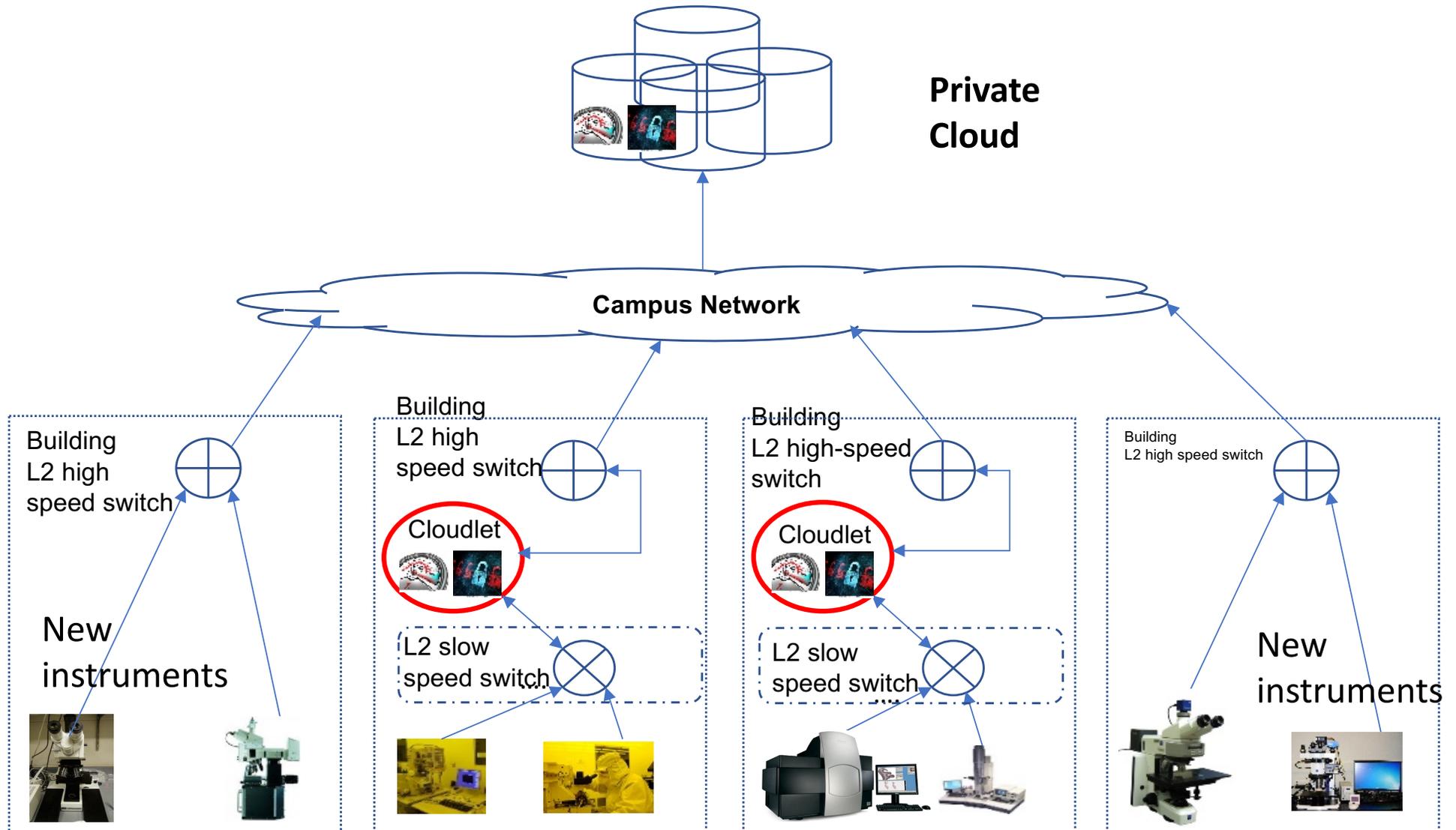


- **Performance mismatch:** Older instruments' Windows NT or XP runs network protocols at lower bandwidth speeds (10Mbps or 100Mbps)

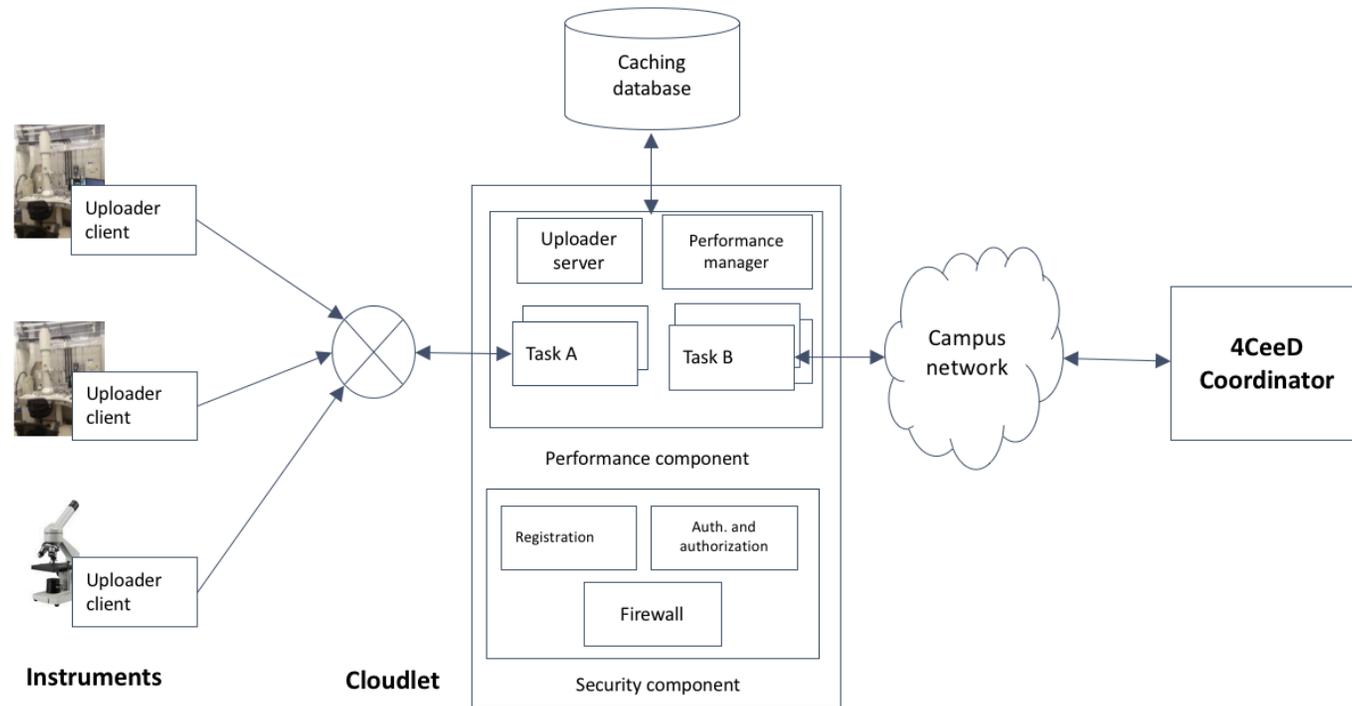


- **Obsolete security:** Older devices and their OS systems cannot be patched, hence being vulnerable & taken offline

Approach: Putting edge device between older instruments and private cloud



Approach: Putting edge device between older instruments and private cloud



Performance:

- Have two network interfaces configured at different speeds
- Perform traffic shaping from slower to faster network

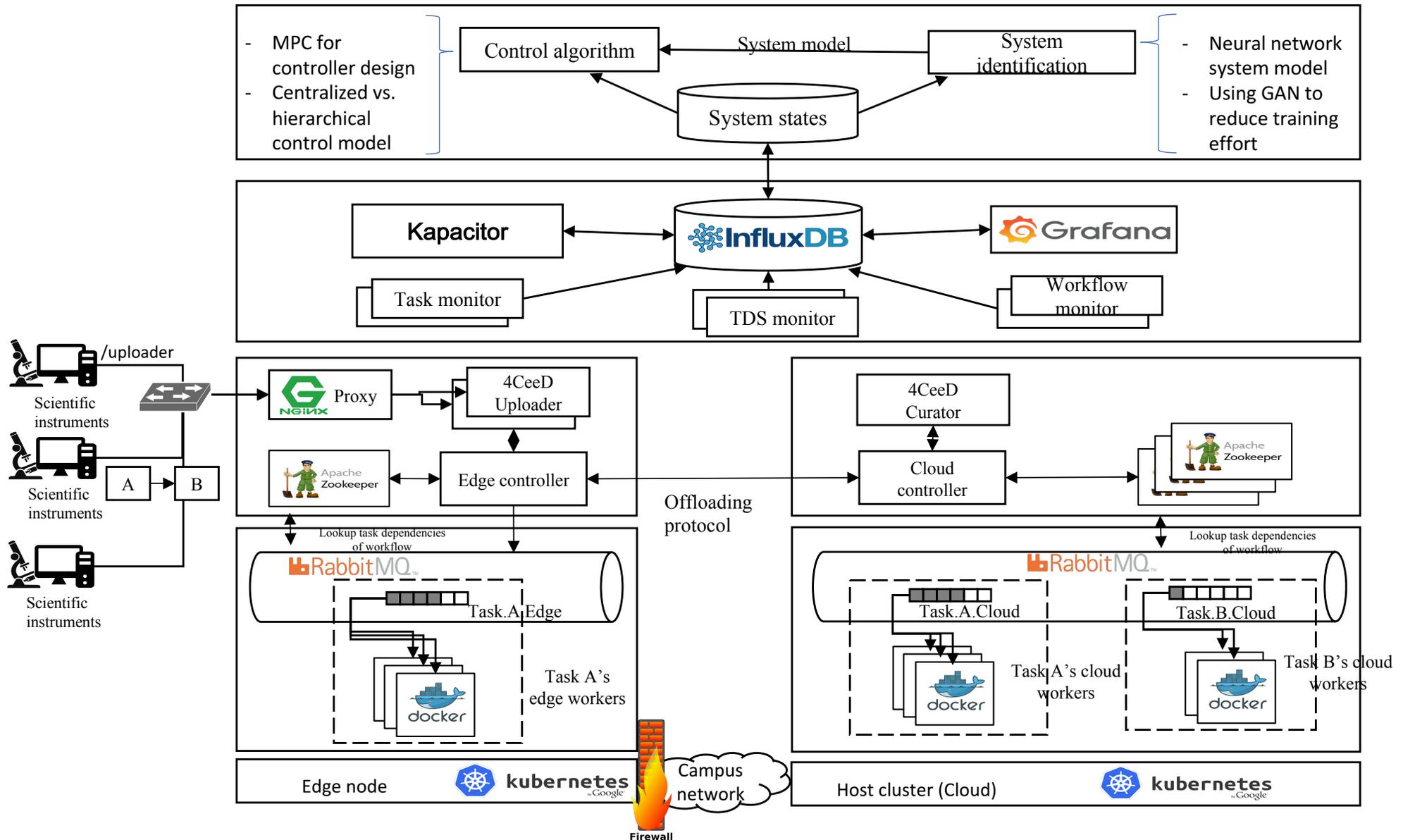


Security:

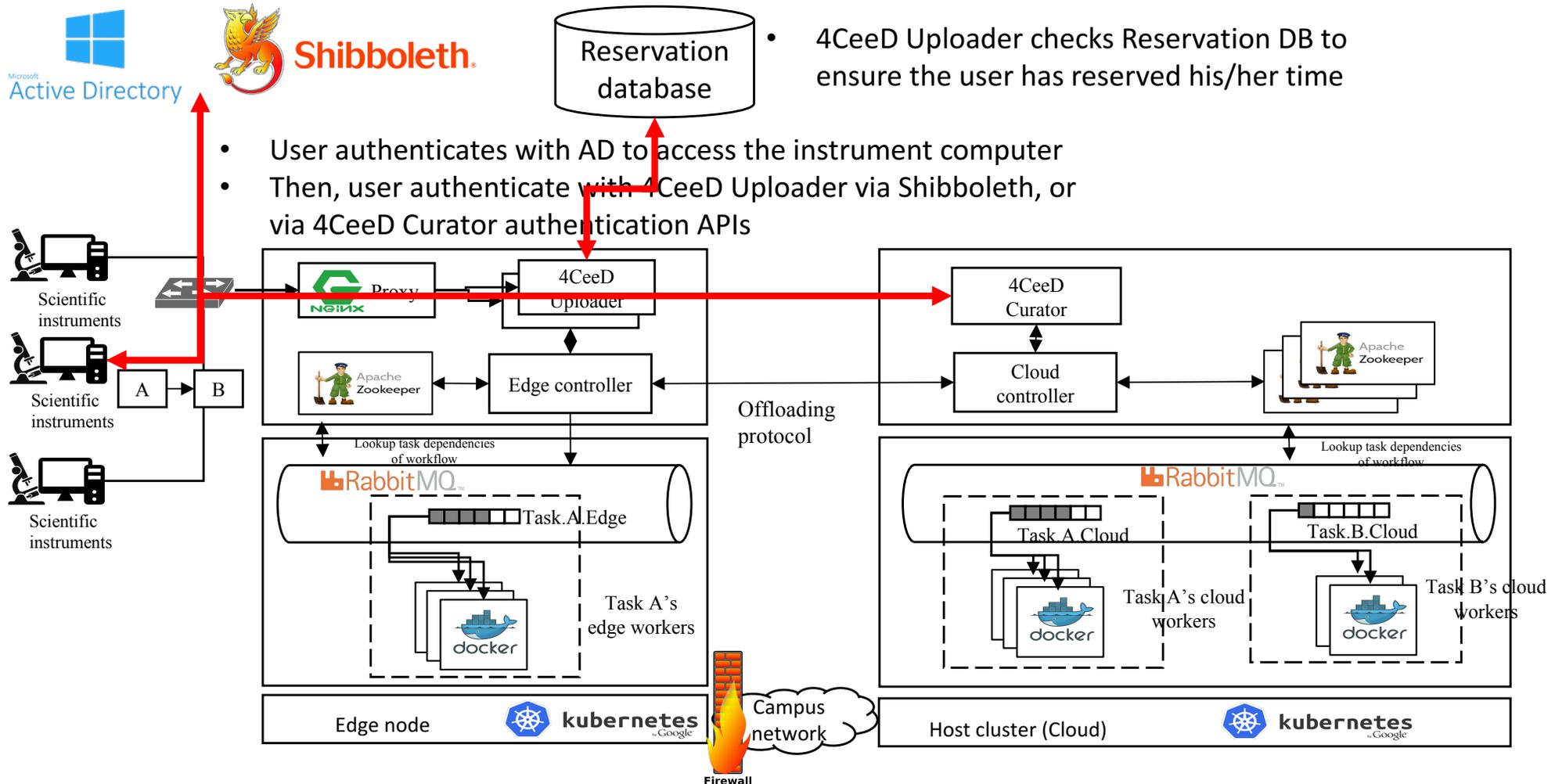
- User & instrument registration
- Data encryption during upload
- Firewall to protect against external threats



Cloudlet & private cloud integration

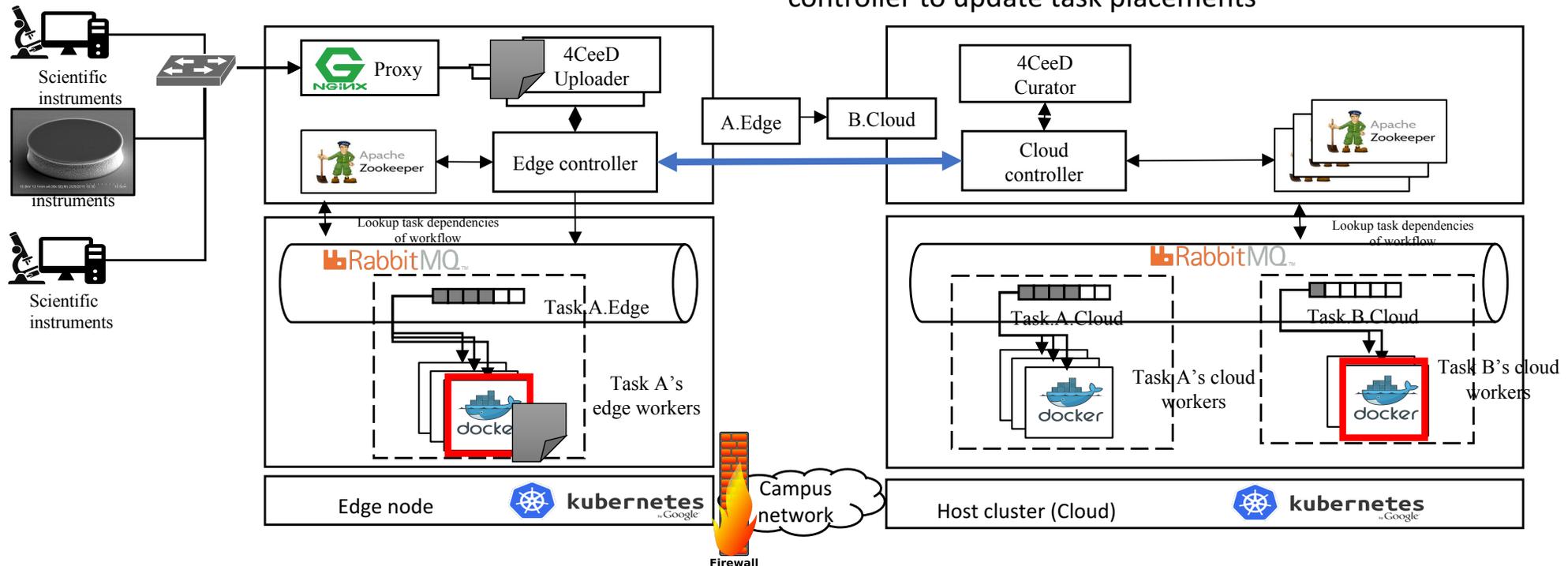


Cloudlet & private cloud integration (authentication)



Cloudlet & private cloud integration (data upload and processing)

- After learning about the request, the data consumer forwards processing requests to the next task (for so long as current placement)
- 4CeeD Uploader communicates with local Edge controller to learn about where to send request to
- Edge controller periodically communicates with cloud controller to update task placements



Conclusion – 4CeeD Advantages

- **One stop location** for research data
- **Sharable** between scientists, and her research partners
- **Visual interpretation of data** without having to open specialty software
- Cloud based → **access data anywhere**
- Available **machine-specific meta-data** on measurement for future replication of measurements
- Available **templates** which allow for the entry of metadata specific to scientist's project

Conclusions and impact

- *First* real-time data acquisition and cloud coordination framework for **materials-related and semiconductor fabrication areas**
- Major Savings:
 - **30% time saving** during MICROSCOPE experimental sessions (e.g., file transfer, save metadata, export previews, etc.)
 - At least **\$30000 SAVINGS PER GRADUATE STUDENT** (initial estimates)
 - The time saving translates into **saving \$25 to \$30 each hour of lab session cost** and each grad student spends at **least 1000 hours in lab during his/her PhD time.**
 - **Shorten time from digital capture to curation, interpretation & insights**
 - Traditionally took **2-3 hours** to capture, upload and curate data via memory stick, notebook, sneaker net to upload data, now it takes **few seconds.**
 - Other benefits include better **data preservation, exploration, and security !**

References

- **“MONAD: Self-adaptive Micro-service Infrastructure for Heterogeneous Scientific Workflows”**
Phuong Nguyen and Klara Nahrstedt
The 14th IEEE International Conference on Autonomic Computing (ICAC) 2017, Columbus, OH, USA.
- **“4CeeD: Real-time Acquisition and Analysis Framework for Materials-related Cyber-Physical Environments”**
Phuong Nguyen, Steven Konstanty, Todd Nicholson, Thomas O’brien, Aaron Schwartz-Duval, Timothy Spila, Klara Nahrstedt, Roy H. Campbell, Indranil Gupta, Michael Chan, Kenton McHenry, and Normand Paquin
The 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid) 2017, Madrid, Spain.
- **“Resource Management for Elastic Publish Subscribe Systems: A Performance Modeling-based Approach”**
Phuong Nguyen, Klara Nahrstedt
The 9th IEEE International Conference on Cloud Computing (CLOUD) 2016, San Francisco, CA, USA.