# 4CeeD: Real-Time Data Acquisition and Analysis Framework for Material-related Cyber-Physical Environments

K. Nahrstedt, S. Konstanty, T. Nicholson, P. Nguyen, T. Spila, T. O'Brien, M. Chan, A. Schwartz-Duval; N. Aluru, P. Braun, R. Campbell, B. Cunningham, I. Gupta, K. McHenry, J. Rogers; University of Illinois at Urbana-Champaign

OVERVIEW

4CeeD framework for real-time **C**apture, **C**uration, **C**oordination, **C**ollaboration, and **D**istribution of material-related data.
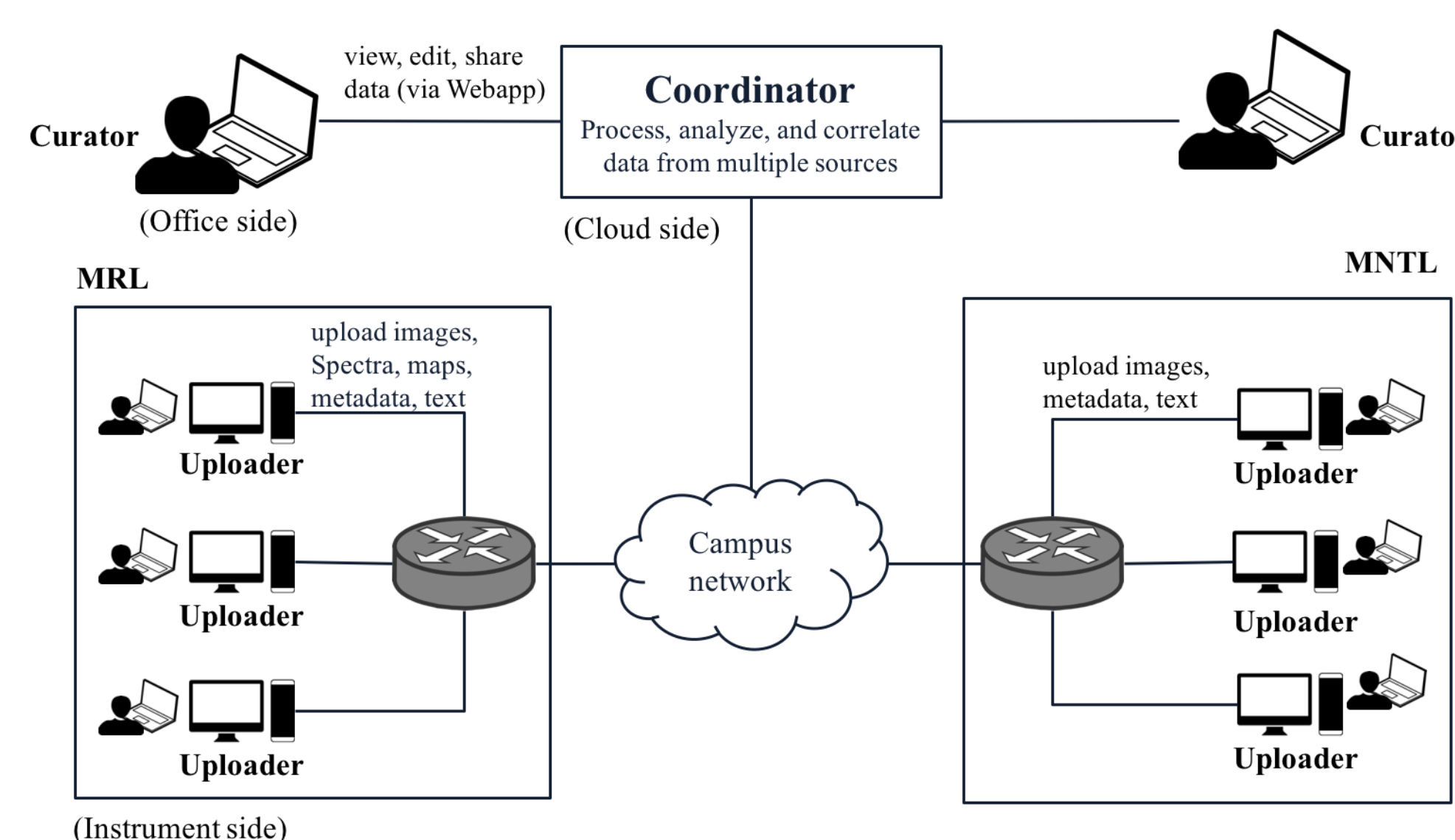
Overview of 4CeeD framework

Curation service

- **Uploader:** Simple, user-in- the-loop interface for uploading raw data generated from materials-making/characterization instruments

- **Curator**: A novel interface that allows users to annotate, add tags, remove erroneous captured data after lab sessions

Summary of curation service successes:

- Avoids using "sneakernet" & prevent data loss, malware spread, and accidental erasure of using USB drives

- Enables to view data and image previews without requiring access to proprietary software reduces time spent in lab.

- Eliminates need to convert file formats & potential loss of metadata in converting proprietary file formats to common image formats (.jpeg, .png, etc.)

- Improves data discovery by improving search ability & data sharing with other users

Coordination service

- Centralized cloud infrastructure for storing and processing uploaded data.

Summary of coordination service successes:

- Supports heterogeneous types of workflow-based data processing jobs, using a novel micro-service based architecture

- Represents highly scalable solution

- Handles variable and bursty workloads, using a novel dynamic resource management mechanism

UPLOADER DESIGN

- Provide streamlined interface for creating collections and datasets.

- Drag n' Drop file uploads.

- Create, share, and reuse metadata templates.

- Preview recent activity

- Enable lab staff to upload on behalf of experimenters.

1) Users can create and reuse collections and sub collections to organize their datasets.
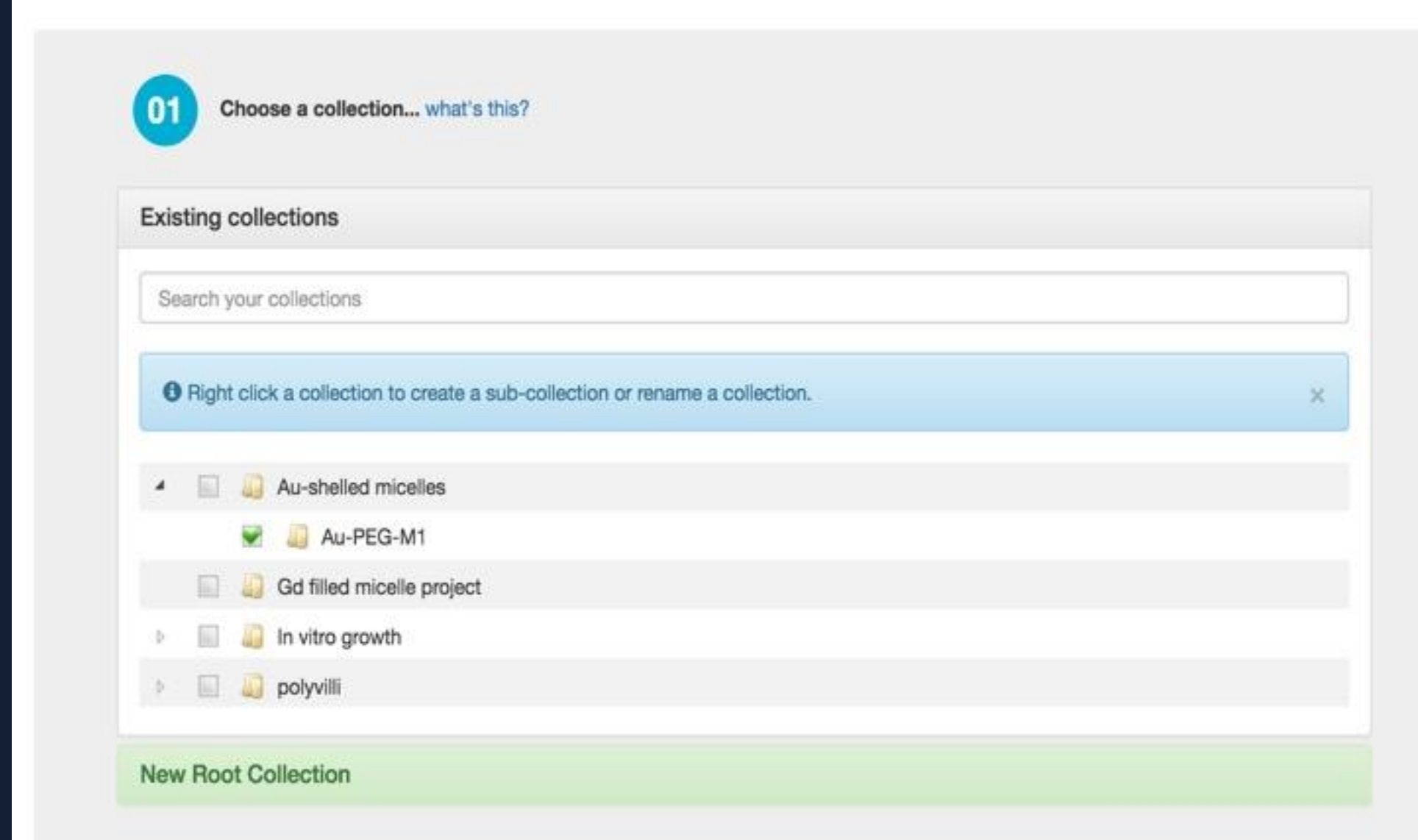
2) Users can create, share, and reuse metadata templates to rapidly and accurately describe their metadata.

*Datasets contain metadata and aggregate related files.*

3) Users can browse or drag and drop their files to be stored in the 4CeeD Cloud.

Supported file types will have a previewer, and the extracted metadata is correlated with the original file, available as a single download.

4) Users can view previews of recently uploaded files, and user and system generated metadata.

CURATOR DESIGN

- Nested collections represent hierarchical, sequential experiment structure.

- Datasets and collections can be downloaded in Library of Congress **BagIt** format.

- Extractors are available for **TEM**, **SEM**, **AFM**, and **X-ray** instruments.

- Extractors provide image previews and metadata which was often lost previously due to file conversion to more common image formats.

- Sharing is possible through role-based managed spaces.

- We enable elastic search plugin.
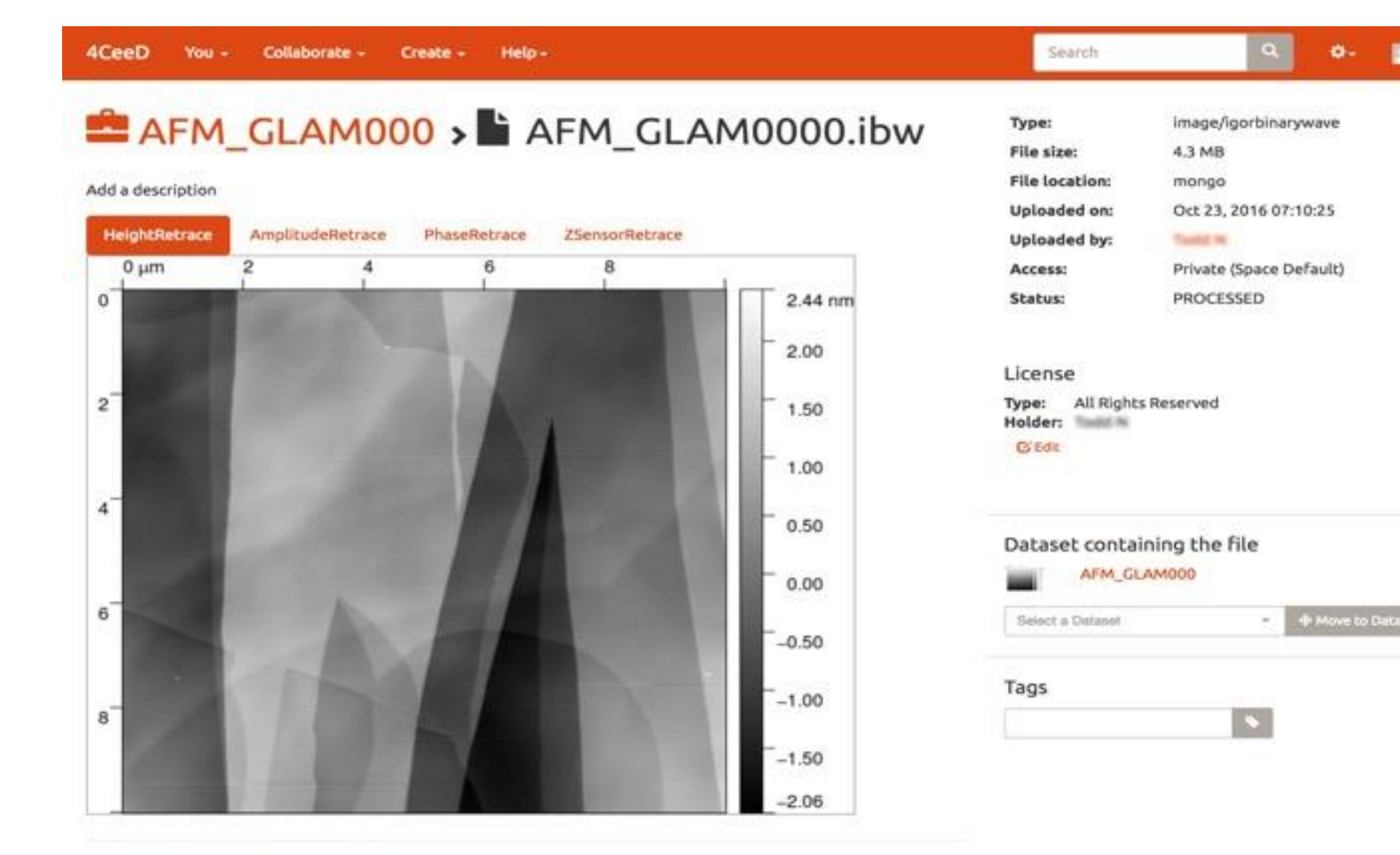

Example shows an uploaded AFM image.

- Multiple previews are provided for various channels

- Extracted metadata provided by the 4CeeD extractors are visible in the metadata panel.

- Further annotation is possible through comments, tags, and notes.

- All files in datasets can be viewed as a list.

- User-defined metadata from the uploader templates are visible.

- Users can easily share their data by creating and adding datasets to a space.
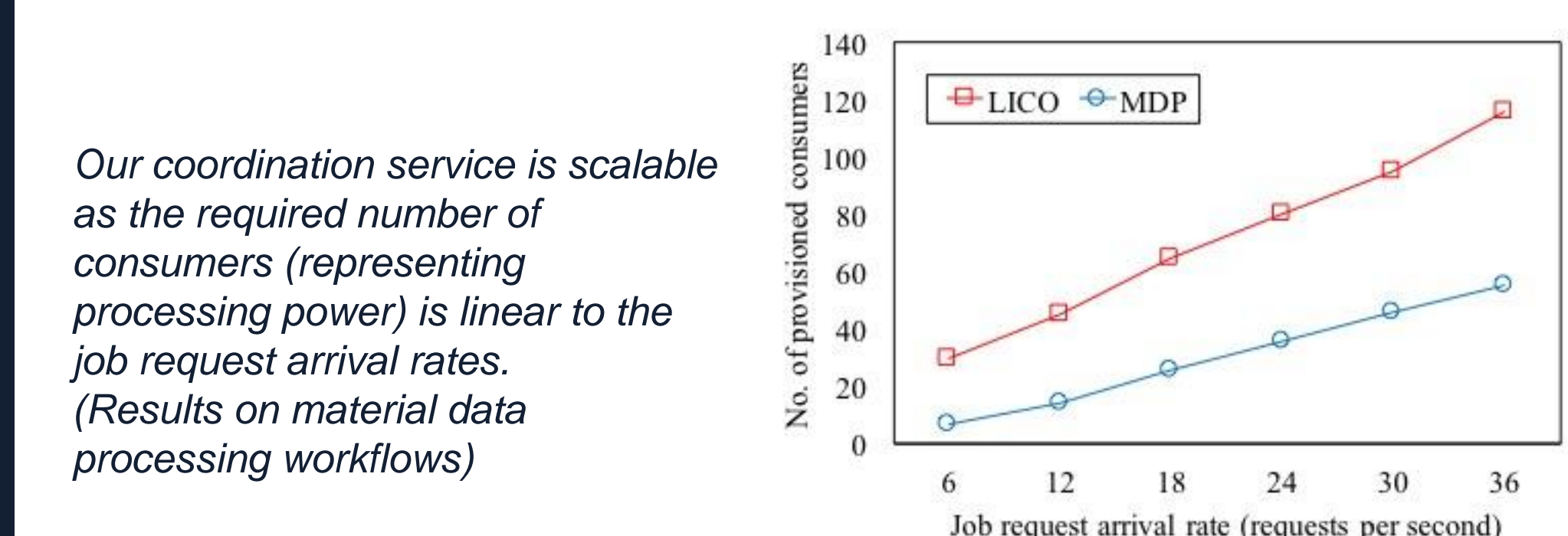
COORDINATION SERVICE
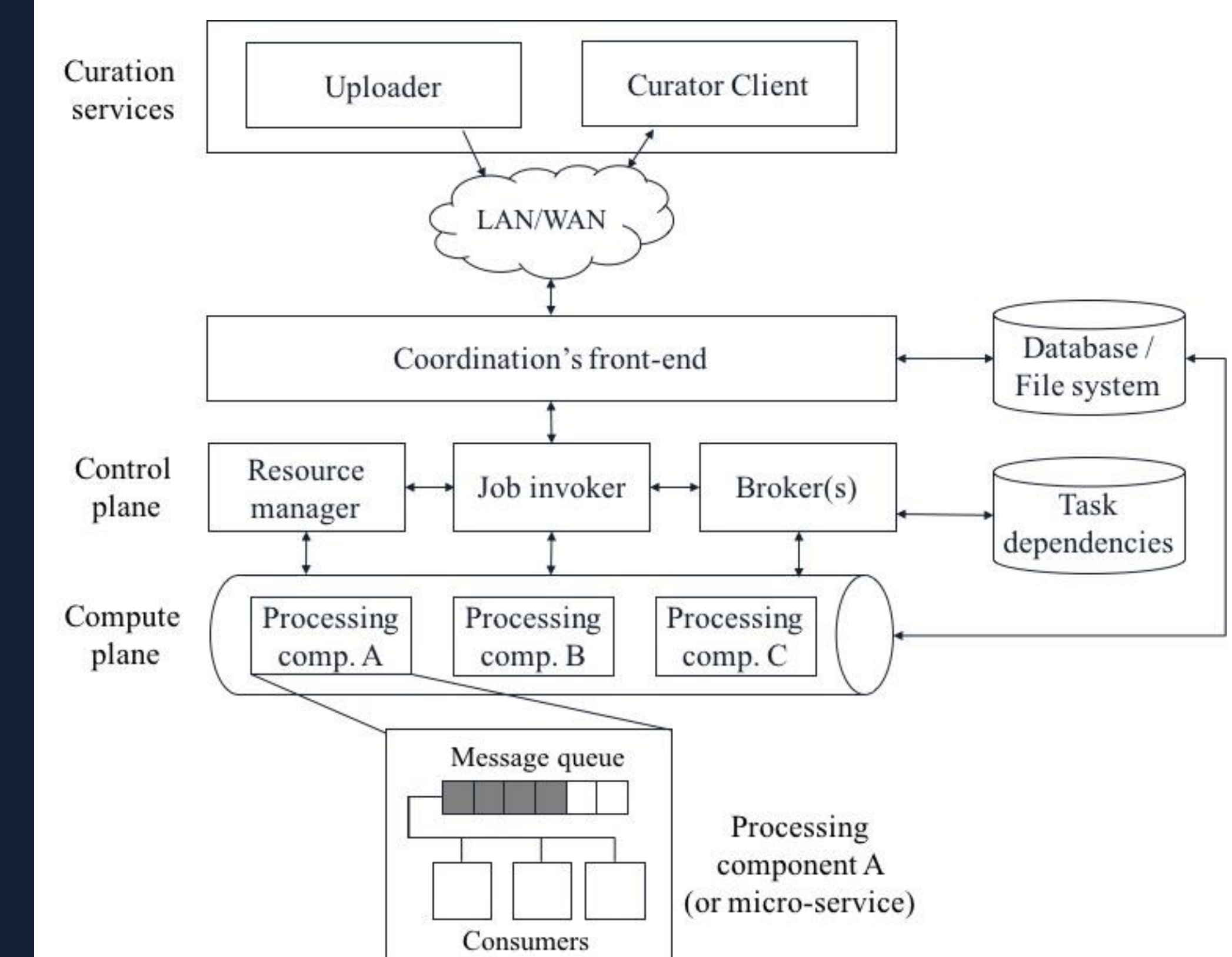
- Micro-service based architecture supports heterogeneous types of workflow-based data processing jobs
- 4CeeD Cloud leverages publish-subscribe middleware to connect curation and coordination services
- Robust dynamic resource management is based on explicit performance modeling of the micro-service based infrastructure

Coordination service's resource management mechanism (GRESMAN) is effective in dealing with variable and bursty workload situation.
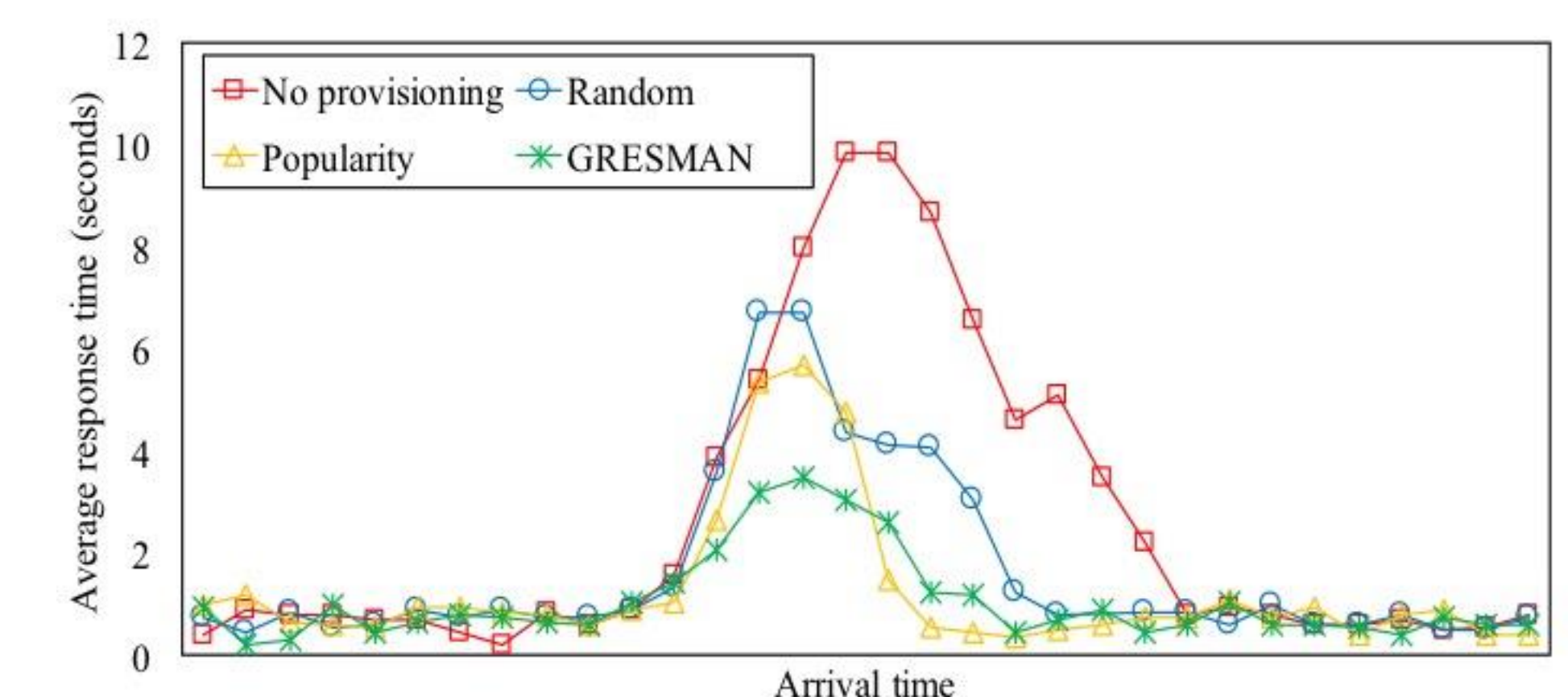
Material data processing workflows

Our coordination service is scalable as the required number of consumers (representing processing power) is linear to the job request arrival rates. (Results on material data processing workflows)

Future Directions:
(1) New instruments and new extractors; (2) new upload and curation functions due to diverse user base; (3) correlation algorithms; (4) increase user base and dissemination beyond UIUC; (5) 4CeeD training materials; (6) sustainability efforts.

ACKNOWLEDGEMENTS