

## Modernizing Microscopy Data Infrastructure: Data and Metadata Curation

Rachel F. Devers<sup>1</sup>, June W. Lau<sup>2</sup>, Gretchen Greene<sup>2</sup>, Todd Nicholson<sup>3</sup>, Phuong Nguyen<sup>3</sup> and Steve Konstanty<sup>3</sup>

<sup>1</sup>. Department of Electrical and Computer Engineering, University of Maryland, College Park, MD.

<sup>2</sup>. Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg MD.

<sup>3</sup>. Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL.

The limitations of current electron microscopy data curation practices are felt whenever a scientist wishes to share and revisit data. As soon as raw instrument data is written to a file, determining the contents of each file (with and without proprietary software) tends to be a serial, time-consuming task. In some cases, an image thumbnail may be available, but the thumbnail alone usually lacks readily accessible contextual information which makes it valuable. This causes scientists to comb through files sequentially, sometimes requiring instrument or detector-specific proprietary software to view data and metadata. This method of data examination limits the significance of each file to a combination of the researcher's notes or memory, OS-generated metadata (file size and time stamp), and perhaps file naming convention. This is not a tractable premise for the hundreds of images for a given sample, thousands of images that may have contributed to publications, and hard drives full of project data contributed by multiple researchers over the span of a project. Ultimately, individually managed data and metadata based on project discipline, chronological order, or some other arbitrary user preference is fundamentally lacking in transparency, longevity and reusability.

With an increasing demand of reproducibility of research produced by public funds, we are shifting towards thoughtful data coordination and curation. We summarized that the ideal laboratory information management system (LIMS) would be centered on the FAIR (Findable, Archivable, Interoperable, Reusable) data principle. We are presenting a pilot use-case based on the T2C2 4CeeD curation and coordination framework [1] for a share-used electron microscopy facility within NIST. Metadata from proprietary data formats are curated using HyperSpy[2]-based extractors. For this presentation, we will share the progress we made and the lessons we learned in tackling this ambitious but essential step in modernizing electron microscopy data infrastructure [3].

### References:

[1] T2C2 is supported by NSF Award Number: 1443013, "CIF21 DIBBs: T2C2: Timely and Trusted Curator and Coordinator Data Building Blocks", PI: Nahrstedt

[2] HyperSpy 1.3 DOI: 10.5281/zenodo.583693.

[3] The authors gratefully acknowledge the assistance from Drs. Michael Katz and Karl Schliep for unraveling the hidden metadata structure and for code testing. The authors of this paper are grateful for the stimulating conversations and assistance provided by Dr. Zachary Trautt at NIST.

**Figure 1.**  
T2C2  
screenshot.

The screenshot displays a web interface for a dataset titled "Magnetic Tunnel Junction". The interface includes a navigation bar with "4CeeD" branding and a search bar. Below the dataset title, there is a list of files. A red arrow points from the first file in the list to a large image of a magnetic tunnel junction. Below the image, there is a "Comments" section and a "Metadata" section. The metadata section contains detailed technical information about the file, including acquisition parameters and microscope settings.

**Dataset: Magnetic Tunnel Junction**  
 Created by Michael Katz  
 Created on Jan 18, 2018  
 Access: Public (Space Default)

**Files:**

- 12 - 790k.dm3**  
 digitalmicrograph/raw  
 Jan 18, 2018  
 17.4 MB
- 13.18.39 Scanning Acquire\_2.ser**  
 digitalmicrograph/raw  
 Jan 18, 2018  
 2.1 MB
- 20 - 115k.dm3**  
 digitalmicrograph/raw  
 Jan 18, 2018  
 17.4 MB
- HAADF.dm3**  
 digitalmicrograph/raw  
 Jan 18, 2018  
 311.8 kB

**Metadata:**

- Extracted by <http://cloudner.nca.illinois.edu/extractors/deprecatedapi> on Jan 18, 2018
- Name:** 12 - 790k
- AcquisitionDate:** 7/6/2017
- AcquisitionTime:** 11:09:58 AM
- ExposureNumber:** 415985
- ExposureTime(s):** 1.0
- MicroscopeInfo:**
- ImageData:**
- ImageTags:**
  - DeviceName:** EM-UltraScan
  - Tecna:**
    - Microscope Titan 300 kV D3188 SuperTwin
    - User MBK1
    - Gun FEG HT 300 Extr volt 4400 V Gun Lens 3 Emission 117.0uA
    - Mode TEM up Mh Zoom Image Defocus (um) 0.064 Magn 790000x
    - Spot 2
    - C2 42.132%
    - C3 63.031%
    - Cty 87.589%
    - Df 89.809%
    - Image shift 0.000/0.000um
    - Stage -137.311 um, -35.733 um, -138.619 um, 3.96 deg, 1.70 deg
    - C1 Aperture: 2000 um
    - C2 Aperture: 150 um
    - CEJ Aperture: retracted
    - SA Aperture: retracted
    - Filter related settings:
      - Mode: Spectroscopy
      - Selected dispersion: 0.10[eV/Channel]
      - Selected aperture: 3mm
      - Prism shift: -0.00[eV]
      - Drift tube: 0.00[eV]
      - Total energy loss: 0.00[eV]
  - Device:**
  - Transform:**
  - Intensity:**
    - Bias (counts): 1000.0
    - Dark Current (counts/s): 0.0
    - Dark Level (counts): 1000.0
    - Maximum Value (counts): 65535.0
    - Minimum Value (counts): 0.0
    - Saturation Level (counts): 50000.0
  - ReferenceImages:**
  - Sequence:**
    - Detector Readout Continuous: 1
    - ModelName: Imaging
  - HighLevel:**
  - Private:**
  - AnnotationGroupList:**
  - ImageDisplayInfo:**

**Extractions:**  
 List extraction events