Foundations of TEM Data Management & Analysis

Why open science and reproducible workflows matter

CITEAM Workshop | May 20, 2026
University of Maryland | NSF Award #2519465, 2519466, and 2519467.

Why this workshop exists

Modern TEM work is not only imaging.

- files are larger and more varied than before
- analysis choices affect the final scientific claim
- data without context becomes hard to trust or reuse

The workshop is organized around three ideas:

Collect well — good files + basic metadata

Analyze transparently — repeatable steps + saved parameters

Share responsibly — reusable data + code + provenance

Goal: to scale from individual tasks to real research workflows

Open science in one minute

Open papers

publish so others can read

Open data

make datasets findable and reusable

Open code

save the exact steps you took

Open methods

record enough detail to repeat the work

FAIR is the simple idea underneath this: make research outputs Findable, Accessible, Interoperable, and Reusable.

Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” Scientific Data (2016): https://www.nature.com/articles/sdata201618

From individual files to a data ecosystem

Coordinated infrastructure across government, academia, and industry is needed so research data can actually be found, connected, and reused.

Campo et al., “Now Is the Time to Build a National Data Ecosystem for Materials Science and Chemistry Research Data,” ACS Omega (2022): https://pubs.acs.org/doi/10.1021/acsomega.2c00905

Why should you care? Three big wins

Replicability & trust

Sharing both data and analysis methods enables replicability (others can re-run the path from raw data to figure) and trust in your methods

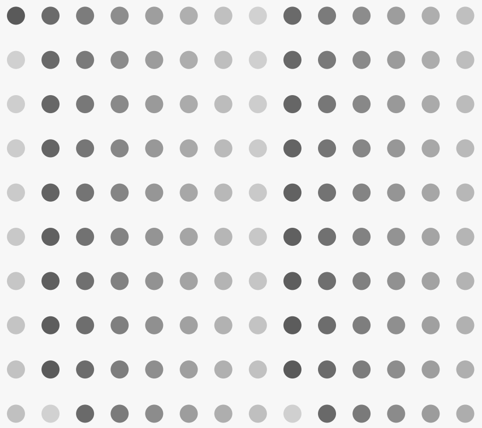
Reuse by others (and future you)

Well-structured data; consistent file names, metadata, and code turn a one-off experiment into a reusable asset.

Scales to teams & AI

Well-structured data is what lets labs collaborate, automate, and learn faster. It enables AI.

Why TEM data need context



Just an image

An image alone is often not enough to interpret later

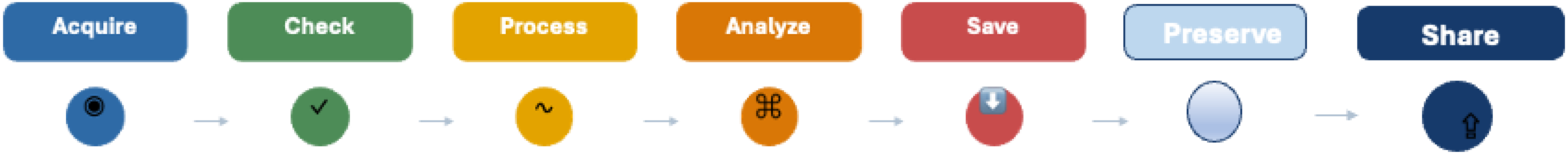
metadata and workflow context include:

- sample ID / composition
- why this experiment was conducted
- mode: TEM, STEM, HAADF, EDS...
- instrument settings and calibration
- acquisition date, operator, notes
- processing steps and saved parameters

The key point is that *data without context becomes hard to trust or reuse*

Taillon, J.S. et al, MaRDA recommendations on microscopy metadata and LIMS, l MRS Bulletin (2025) https://link.springer.com/article/10.1557/s43577-025-00882-2

Think in workflows, not “one and done”



A workflow is a repeatable path from instrument to result

If you can name the steps, save the parameters, and keep the files linked, you are on the path to robust science.

The “data trail” idea : keep a link between sample → experiment → settings → files → processing steps

R.J. Hanisch, et al, *NIST Research Data Framework (RDaF), Version 1.5* (NIST Special Publication NIST SP 1500-18r1, National Institute of Standards and Technology, Gaithersburg, May 2023). https://doi.org/10.6028/NIST.SP.1500-18r1; https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-18r1.pdf

Open code: own your analysis

Even a simple notebook helps because it records:

- what file you opened
- what parameter you changed
- what output you saved
- what someone else would need to repeat it

Notebook.ipynb

Think of open code as transparent lab notes you can rerun.

FAIR applies not only to data but also to the algorithms, tools, and workflows that led to the data: https://www.nature.com/articles/sdata201618
FAIR principles for research software overview: https://www.nature.com/articles/s41597-022-01710-x

What is metadata?

Examples of data

image

stack

spectrum

notebook

Metadata = the information that makes a file understandable.

- what sample it is
- how the data were acquired
- what the units / scale are
- what processing already happened
- what terms and identifiers were used

Minimum viable metadata for TEM

Sample ID	film_001
Technique	TEM / STEM / SAED
Voltage (kV)	200
Magnification / pixel size	30,000x / 0.2 nm
Detector(s)	BF, HAADF, EDS
Acquisition date	2026-05-14

Start small. Be consistent.

- You do not need every possible field on day one.
- You do need enough context for another person to understand the file.
- Simple file name conventions, augmented by links to metadata and associated units*

A short metadata template beats no template.

** Metadata recommendations grounded in MaRDA MRS Bulletin article: https://link.springer.com/article/10.1557/s43577-025-00882-2

* https://www.bipm.org/en/measurement-units

Standard terms

"Casual labels"

- "BF image"
- "bright field"
- "TEM BF"
- "bf1_final_reallyfinal"

→

Shared standard vocabulary

- bright-field TEM
- detector = BF
- technique = TEM
- units = nm, kV

MaRDA recommendations discuss comprehensive metadata, standardized vocabularies/identifiers: https://link.springer.com/article/10.1557/s43577-025-00882-2

Where do data and code go?

Working storage

local / lab copy
(active project)

Group or institution

shared backed-up copy
(team and/or public access)

Publication repository

DOI + citation
(long-term sharing)

Best habit: publish the data and the analysis code together whenever possible.

- General best-practice synthesis from FAIR data/software sources: https://www.nature.com/articles/sdata201618
- Research software FAIR principles overview: https://www.nature.com/articles/s41597-022-01710-x

Example repository: EMPIAR

What it is

- public archive for raw EM image data
- supports images underlying EMDB / cryo-EM and volume EM workflows
- entries are openly reusable under CC0

A repository is not just storage; it is discovery + citation + reuse.

1 Prepare metadata

2 Upload files

3 Publish and make the dataset discoverable

- EMPIAR homepage: https://www.ebi.ac.uk/empiar/
- EMPIAR about page / CC0 reuse: https://www.ebi.ac.uk/empiar/about/
- EMPIAR policies: https://www.ebi.ac.uk/empiar/policies/

Materials data in the bigger picture

Materials Data Facility (MDF)

- example of a platform for publishing and discovering materials datasets
- connects data with code, publications, and community schemas
- shows what “data infrastructure” looks like beyond a single lab

Bring it home for TEM

- TEM users benefit from the same ecosystem logic: repositories, standards, metadata, and tooling all reinforce one another.
- This is why workshop skills matter.

- Materials Data Facility site: https://www.materialsdatafacility.org/
- Campo et al., “Now Is the Time to Build a National Data Ecosystem for Materials Science and Chemistry Research Data,” ACS Omega (2022): https://pubs.acs.org/doi/10.1021/acsomega.2c00905

Electron microscopy for materials science - Antwerpen

Published May 10, 2024 | Version v1

Dataset Open

Three-dimensional Reconstructions and Quantitative Indicators for colloidal particles in Dry and Liquid Conditions in Scanning Transmission Electron Microscope (STEM)

Arenas Esteban, Daniel (Data collector)[1]; Wang, Da (Data collector)[1]; Kadu, Ajinkya (Researcher)[1]; Bals, Sara (Supervisor)[1]; Liz-Marzán, Luis (Supervisor)[2]

Hide affiliations

Graphene and ISO

1. University of Antwerp
2. University of Vigo

This dataset accompanies the research presented in the paper:

Esteban, D.A., Wang, D., Kadu, A., Olluyn, N., Iglesias, A.S., Perez, A.G., Casablanca, J.G., Nicolopoulos, S., Liz-Marzán, L.M. and Bals, S., 2023. Liquid phase fast electron tomography unravels the true 3D structure of colloidal assemblies. *arXiv preprint arXiv:2311.05309.* [link]
It provides a comprehensive collection of three-dimensional reconstructions and quantitative descriptors for small colloidal particles. These gold nanoparticles are arranged in tetrahedral and other intricate geometries under both dry and liquid conditions. The dataset contains 3D reconstructions and quantitative indicators such as centroids, volumes, surface areas, solidity measures, and principal axis lengths for assemblies with 4, 5, and 6 particles.

The dataset includes: `N4_dry_dart.rec` and `N4_liquid_dart.rec` for the 3D reconstructions of an assembly with 4 particles in dry and liquid conditions respectively; `N4_quant_descriptors_dry.mat` and `N4_quant_descriptors_liquid.mat` providing quantitative descriptors for these conditions. Similar files are provided for assemblies with 5 and 6 particles, such as `N5_dry_dart.rec`, `N5_liquid_dart.rec`, `N5_quant_descriptors_dry.mat`, `N5_quant_descriptors_liquid.mat`, and the corresponding files for N6.

This dataset can be used to study the structural dynamics of nanoparticle assemblies and studies in colloidal chemistry, materials science, and nanotechnology. The `.rec` files can be visualized using volume rendering software (e.g. Amira or Avizo), while the `.mat` files contain structured data for analysis in MATLAB. The supporting code and scripts for this dataset are available on the GitHub repository: https://github.com/ajinkyakadu/LiquidET_NatComm2024.

Files

Liquid_Nat_Comm_datasets.zip

Liquid_Nat_Comm_datasets.zip

datasets

N4_dry_dart.rec	537.0 MB
N4_liquid_dart.rec	537.0 MB
N4_quant_descriptors_dry.mat	1.2 kB
N4_quant_descriptors_liquid.mat	1.2 kB
N5_dry_dart.rec	537.0 MB
N5_liquid_dart.rec	537.0 MB
N5_quant_descriptors_dry.mat	1.3 kB
N5_quant_descriptors_liquid.mat	1.3 kB
N6_dry_dart.rec	537.0 MB
N6_liquid_dart.rec	537.0 MB
N6_quant_descriptors_dry.mat	1.4 kB
N6_quant_descriptors_liquid.mat	1.4 kB

Versions

Version v1 May 10, 2024
10.5281/zenodo.11175299

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.11175298. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

OpenAIRE

Communities

Electron microscopy for materials science - Antwerpen

Keywords and subjects

EuroSciVoc

Computed tomography | Electron microscopy | Physical chemistry | Nano-materials

Details

DOI
DOI 10.5281/zenodo.11175299

Resource type
Dataset

Publisher
Zenodo

Languages
English

https://zenodo.org/records/11175299

Materials Data Resources

GENERAL

- **Materials Resource Registry** allows for the registration of materials resources, bridging the gap between existing resources, software and repositories and end users. The Materials Resource Registry functions as a federated service, making the registered information from multiple institutions available for research to the materials community. https://materials.registry.nist.gov/
- **NIST MaterialsData Repository** is an effort to establish data exchange protocols and mechanisms to foster data sharing and reuse across a wide community of researchers, with the goal of enhancing the quality of materials data and models. Dataresent on this system are varied and may originate from within NIST or from the worldwide materials community and may or may not be critically reviewed. https://materialsdata.nist.gov/
- **NIST Schema Repository and Registry** is a service operated to improve discovery, access and reuse of materials-related schemas, data models, ontologies and more. https://schemas.nist.gov/

PHASE-BASED

- **TRC Alloy Data** is a collection of published experimental thermophysical and thermochemical property data, including full provenance and critically evaluated uncertainties. Data collection efforts focus on unary, binary, and ternary metallic systems. API access is available upon request. http://trc.nist.gov/metals_data

What you should leave with today

- A practical TEM workflow you can repeat
- A minimum metadata template to describe your files
- Basic notebooks you can run and modify
- A clearer path to sharing code and data responsibly
- Appreciation for the value of open science

Open, well-managed data = better science, improved efficiency, better return on investment- given potential for re-use.

Let the workflow carry some of the cognitive load so you can focus on the science.