



DATA SCIENCE &  
SCIENTIFIC COMPUTING



**UNIVERSITÀ  
DEGLI STUDI  
DI TRIESTE**

# Designing and deploying a FAIR-by-design data pipeline and platform for electron microscopy laboratories

Research thesis in: Data Management

Supervisor

Dott. Federica Bazzocchi

Candidate

Nicola Perin

University of Trieste

19 settembre 2025

# Agenda

## Context & Goals

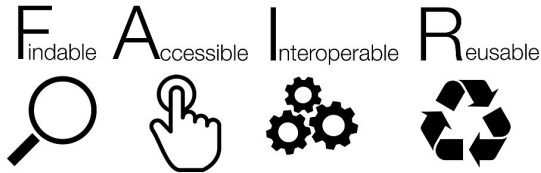


# Motivation & Problem

- ▶ Electron microscopy (EM) labs produce very large and mixed datasets (images, diffraction patterns, spectra).
- ▶ Proprietary file formats and little metadata make data hard to share and reuse.
- ▶ Informal methods (like file names or personal notes) do not work well in collaborations.

**Challenge:** How can we make this data easier to share, reuse, and preserve?

# FAIR Principles



Answer: apply the FAIR principles — make microscopy data **Findable, Accessible, Interoperable, Reusable**. Emphasis on metadata, standardized formats, and machine-actionable records.

# Standards & Formats

`img/diagrams/nexus_nxem.png`

- ▶ **HDF5**: hierarchical container for large arrays + attributes.
- ▶ **NeXus**: scientific conventions (NXinstrument, NXsample).
- ▶ **NXem**: EM definition (images, diffraction, EDS/EELS, 4D-STEM).

# Institutional Context

Area Science Park → RIT  
institute with labs:

- ▶ LADE (data engineering)
- ▶ LAGE (genomics)
- ▶ LAME (electron microscopy)

ORFEO data center provides  
HPC, Ceph storage, and identity  
services. This project targets

LAME workflows, but is scalable  
across NFFA-DI.

img/diagrams/area\_science\_park.png