PARADIM Data Collective—A Collaborative Discovery Platform

David Elbert, Nicholas Carey & Tyrel M. McQueen, Johns Hopkins University

Tillmann Kubis & Daniel Mejia, NanoHub at Purdue University

Goals:

- Collaborative data investigation platform
- Integrated data-centric, workforce development
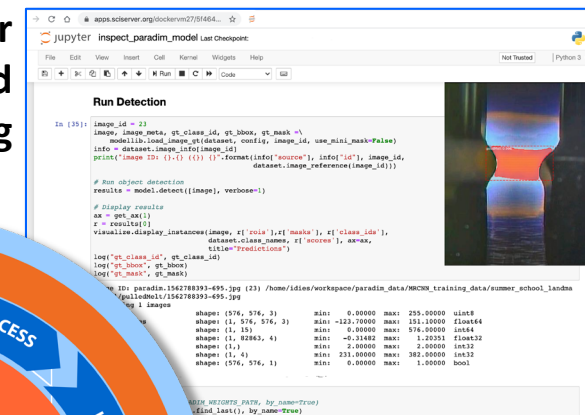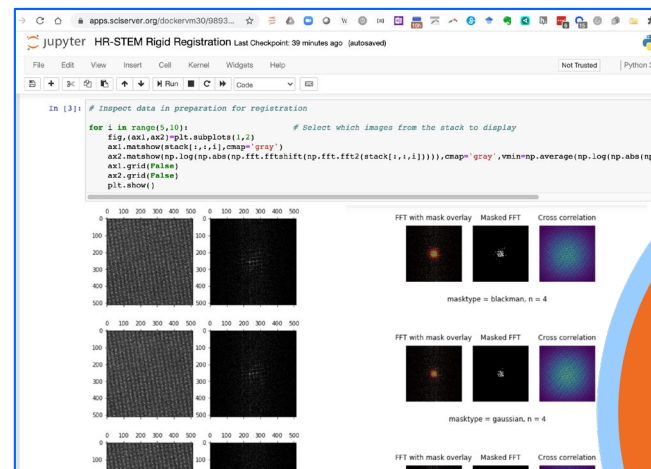- *"Bring the analysis to the MIP"*

Built On:

- SciServer Platform (NSF DIBB)
  - Containerized Compute
  - Integrated SQL Server
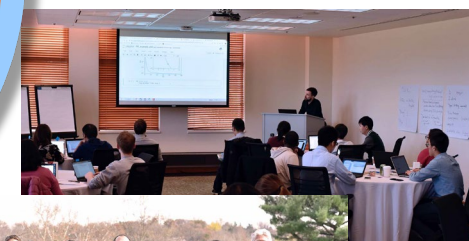- Custom Python environments
- NanoHub Remote

Results:

- Python-based data wrangling, visualization, and analysis
- Machine Learning development and deployment
- Jupyter notebook training materials
- Secure, automated data ingress
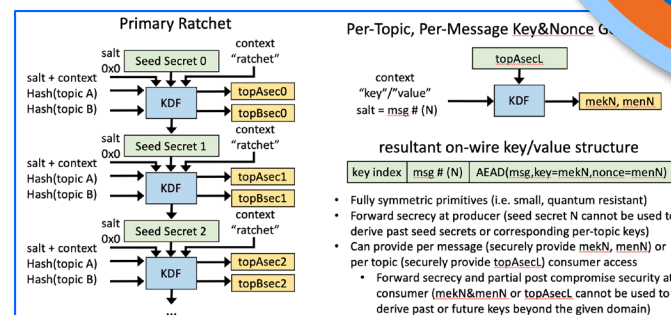
https://www.paradim.org/publications/data_sets

https://www.paradim.org/toolbox/datatools

Python notebooks for training for data visualization, analysis, and machine learning

End-to-end secure data streaming with KafkaCrypto

PARADIM Data Infrastructure

Actionable Information Fast and FAIR

ACQUIRE · PROCESS · LABEL · ANALYZE · STORE · SHARE · REUSE

Training Workshop and Summer Schools

PARADIM Data Collective—FAIR Data for Materials Innovation

David Elbert & Tyrel M. McQueen, Johns Hopkins University, Tillmann Kubis, NanoHub at Purdue University

Goals:
- Create FAIR Materials Data infrastructure
- Provide MIP with FAIR compliance for their data
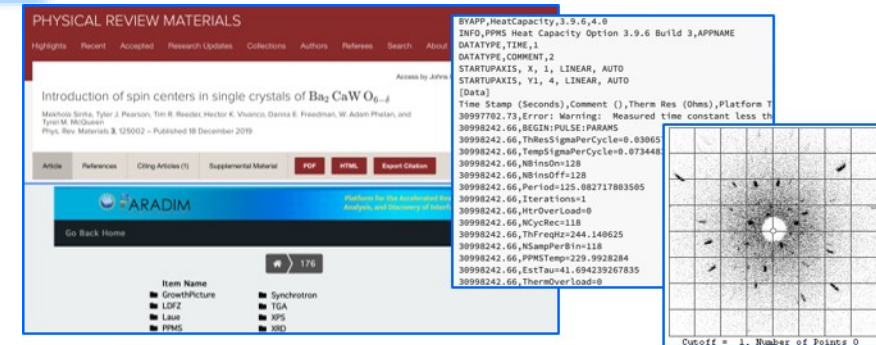
Results:
- **FAIR** components implemented
  - **F**indable:
    - Digital Object Identifier (DOI)
    - DataCite Metadata
    - www.paradim.org browsing
  - **A**ccessible:
    - Permanent landing page
    - Data retrievable by DOI
  - **I**nteroperable:
    - Open file formats
    - Instrument standard formats
  - **R**eusable:
    - Explicit license (CC4.0-BY-NC-ND)
- Three levels of compliance:
  1. Public
  2. Public and citable
  3. Public, citable, and highly curated

Future Work: Richer Metadata, Deeper Curation, Improved Browsability, More Data
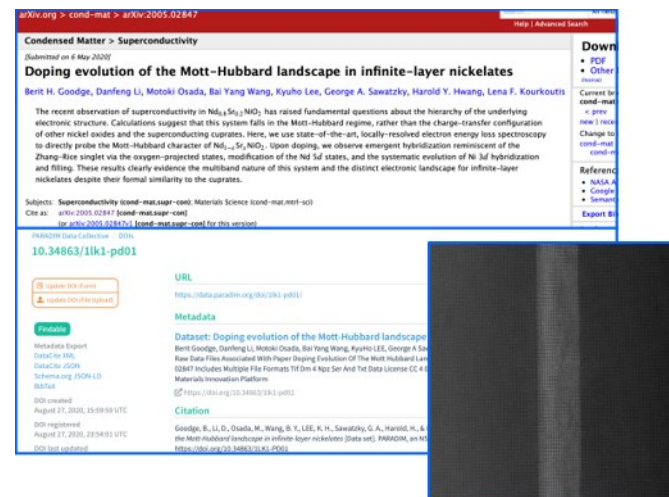
Developed Levels of FAIR Data

1. Publicly browsable data

2. Publicly browsable and DOI citable

3. Publicly browsable, DOI citable, and curated with associated analysis codes

Where Materials Begin and Society Benefits