

# Getting to the Future Faster: the FAIR Trade

*Spoiler Alert: It's not just about the data!*



David Elbert

Hopkins Extreme Materials Institute &  
CDO, PARADIM Materials Innovation Platform

elbert@jhu.edu

## The FAIR Guiding Principles

### Findable:

- F1 Data and metadata are assigned a globally unique and persistent identifier
- F2 Data are described with rich metadata (defined by R1 below)
- F3 Metadata clearly and explicitly include the identifier of the data it describes
- F4 Data and metadata are registered or indexed in a searchable resource

### Accessible:

- A1 Data and metadata are retrievable by their identifier using a standardized communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- A2 Metadata are accessible, even when the data are no longer available

### Interoperable:

- I1 Data and metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2 Data and metadata use vocabularies that follow FAIR principles
- I3 Data and metadata include qualified references to other (meta)data

### Reusable:

- R1 Data and metadata are richly described with a plurality of accurate and relevant attributes
  - R1.1 Data and metadata are released with a clear and accessible data usage license
  - R1.2 Data and metadata are associated with detailed provenance
  - R1.3 Data and metadata meet domain-relevant community standards

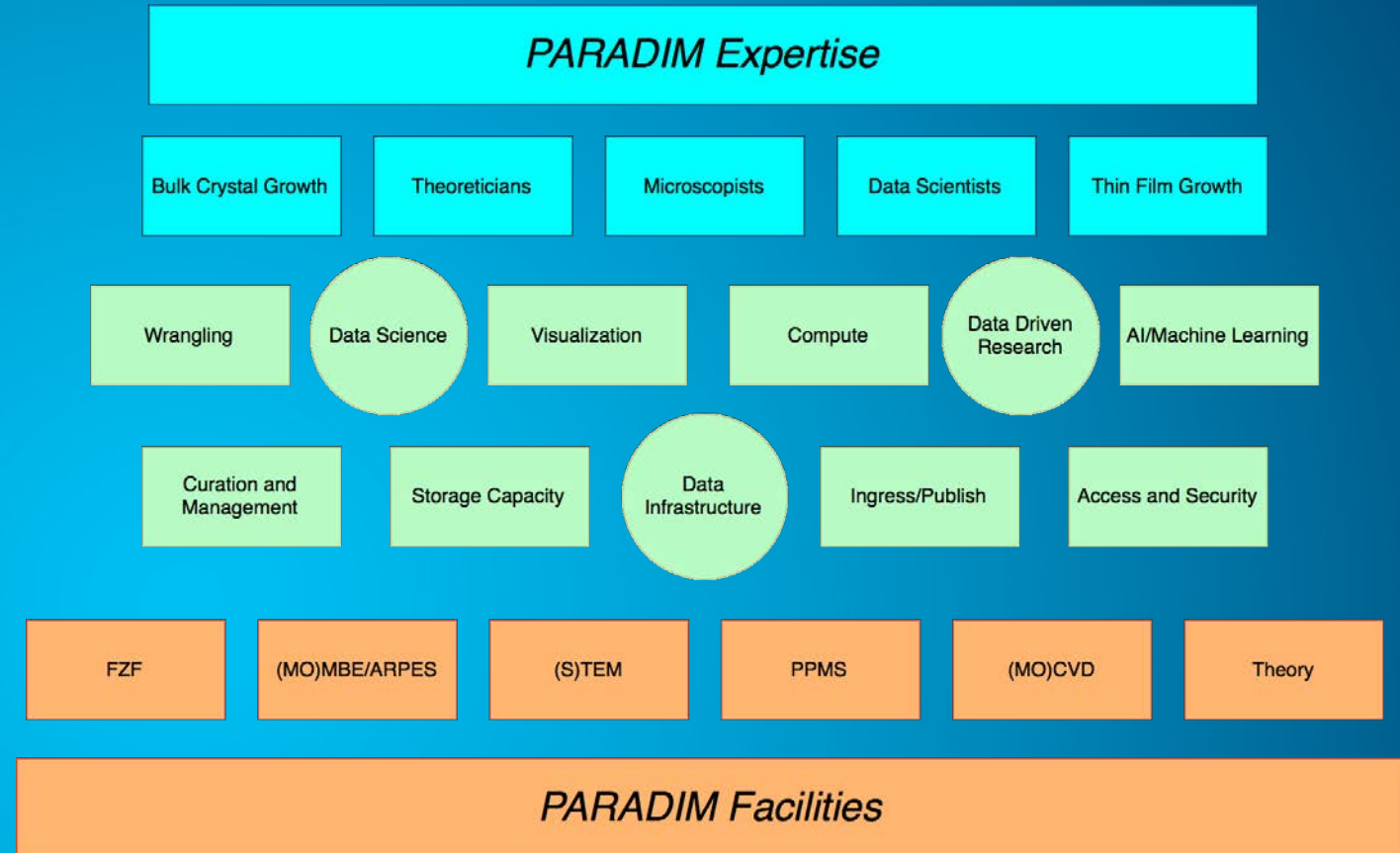
# PARADIM

## User Facility:

- Crystal Growth Facility (Hopkins)
- PPMS and Characterization (Hopkins)
- Thin Film Deposition (Cornell)
- Electron Microscopy (Cornell)
- Theory/Modeling (Cornell/Clark Atlanta)

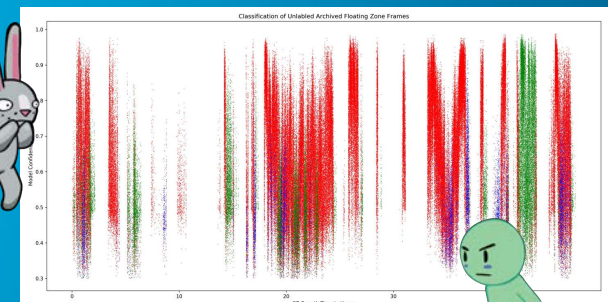
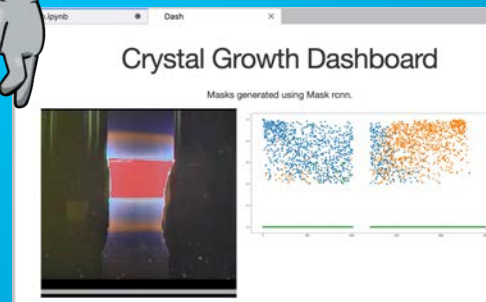
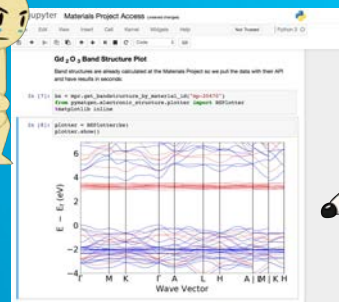
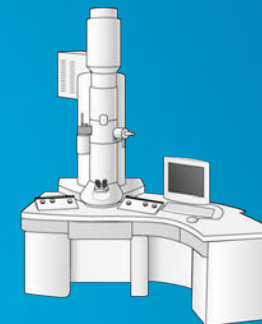
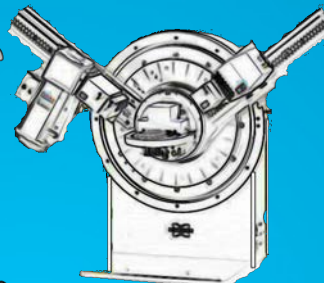
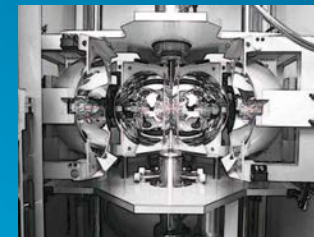
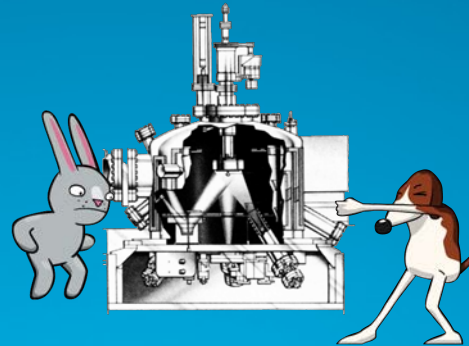
## Data Challenges:

- User access/Security/Privacy
- Large Volume/Dispersed Activity
- Scalable visualization/analysis
- Education for broad user base
- Changing/Evolving infrastructure
- Time value of data



# Many Moving Parts

- Equipment
- People
- Ideas





# Materials Genome Initiative (MGI)

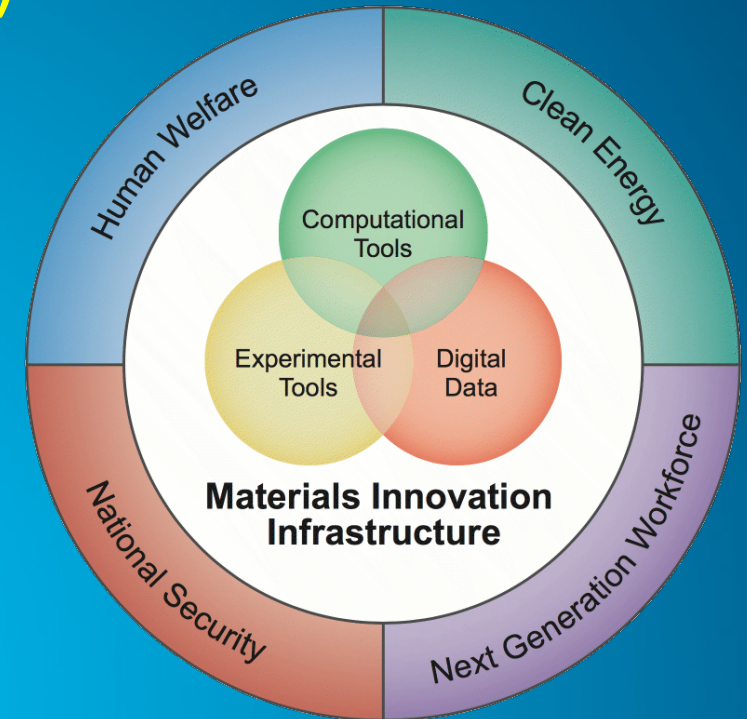
*- Discover, Develop, and Deploy Twice as Fast*

## Strategic Goals:

- Facilitate Access to Materials Data
- Equip the Next-Generation Materials Workforce
- Integrate Experiments, Computation, and Theory
- Enable a Paradigm Shift in Materials Development

## Cross Cutting Themes:

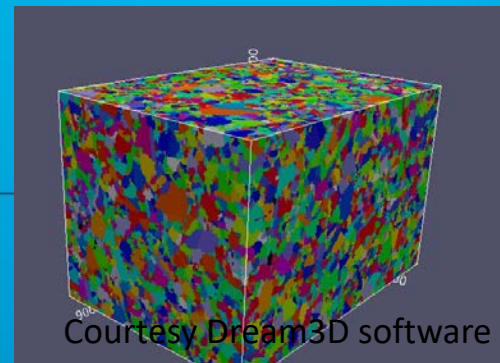
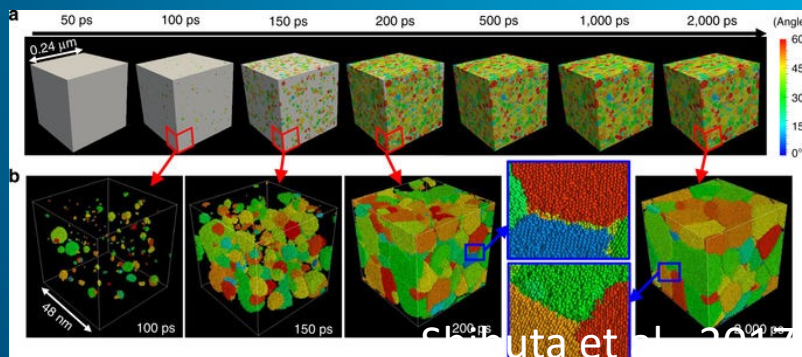
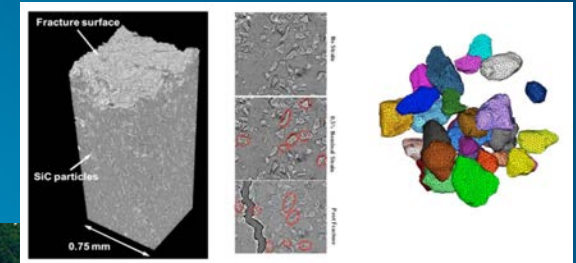
- Incentivizing open data and access of tools
- Structuring public-private partnerships
- Driving innovation across computation, data informatics, and experimentation
- Moving the community to a different cultural norm



Refs: <https://www.mgi.gov/content/mgi-infographic> and [https://www.mgi.gov/sites/default/files/documents/wadia\\_mgi\\_talk.pdf](https://www.mgi.gov/sites/default/files/documents/wadia_mgi_talk.pdf)

# Materials Data is Exploding

- Higher resolution
- Shorter time scales
- Higher dimensionality
- Dynamic experiments
- Larger simulations
- Tighter processing control



Courtesy Dream3D software

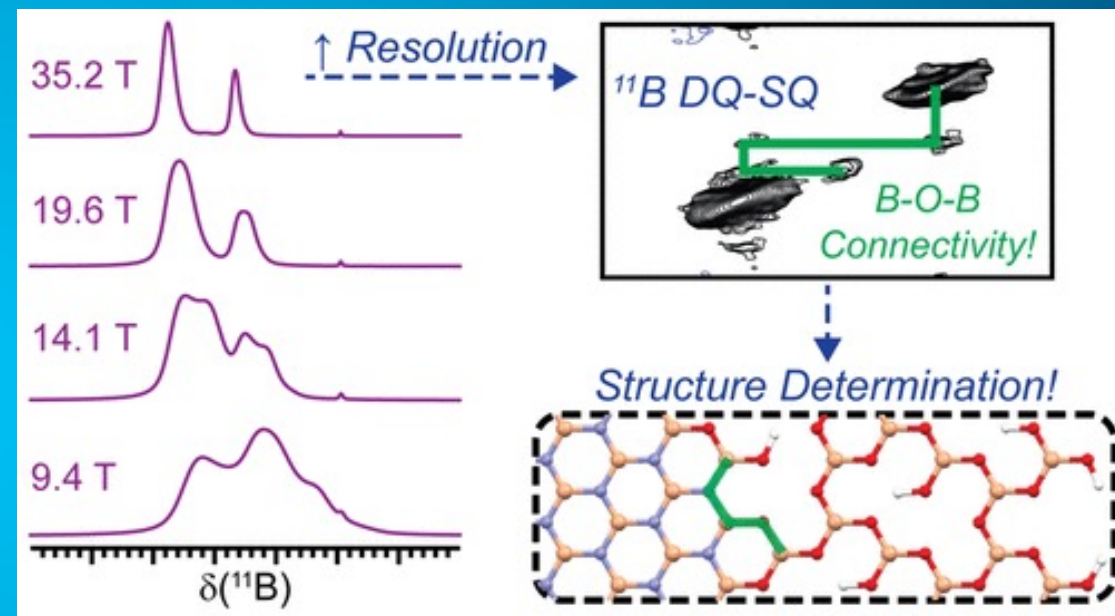




# What's the Value of Data?

## *Superior Resolution Solves Structure Dilemma*

- Oxidative Dehydrogenation Catalyst
- $^{11}\text{B}$  with  $B_0 = 35.2$  T
- Superior resolution
- Distinguish active B species
- Determine the critical structure



Dorn et al, 2021 ACS DOI [10.1021/acscatal.0c03762](https://doi.org/10.1021/acscatal.0c03762)

# Disruptive Change

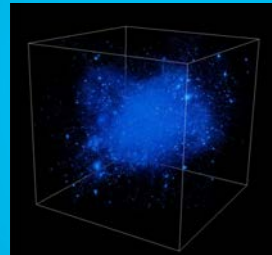
**1000 Years:** empirical descriptions of natural phenomena



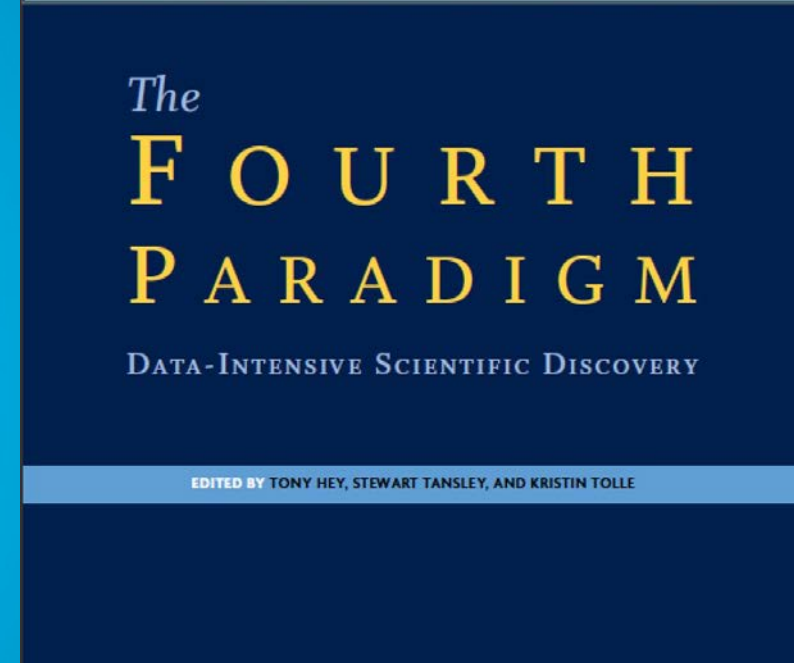
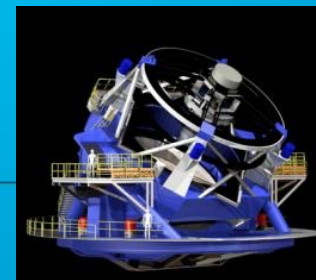
**100's Years:** theoretical branch using models, generalizations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$

**10's Years:** computational branch simulation of complex phenomena



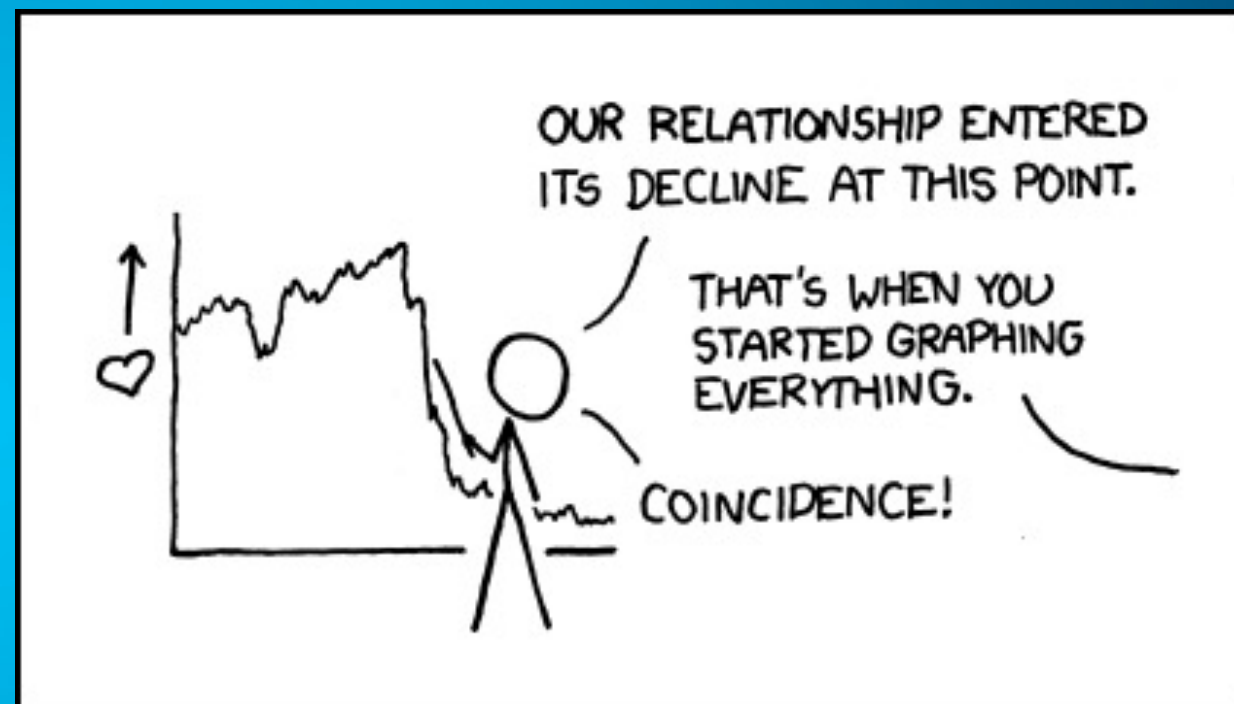
**Now:** data intensive science, **discovery directly from the data.** Synthesizes theory, experiment and computation with statistics



# What's the Value of Data?

*Provides Proof of What We Assert – the Heart of Scientific Knowledge*

- Property measurements
- Probes structures and processes
- Prove/Eliminate hypotheses
- Enable validation



<https://xkcd.com/523>



# What's the Cost of Data?


*Data is Dearly Won!*

- Instrumentation
- Scientist Time
- Expertise
- Validation
- Curation
  - Annotation
  - Storage
  - Maintenance



# The Value of Data Changes Over Time

*Data regains value over time if it can be reused*



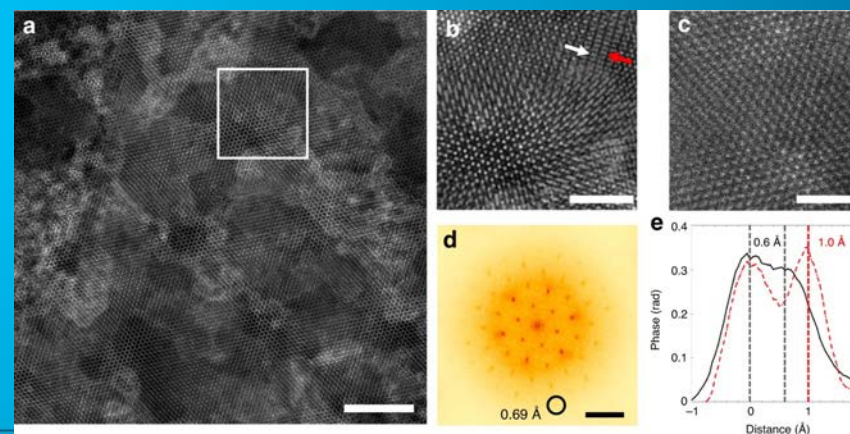
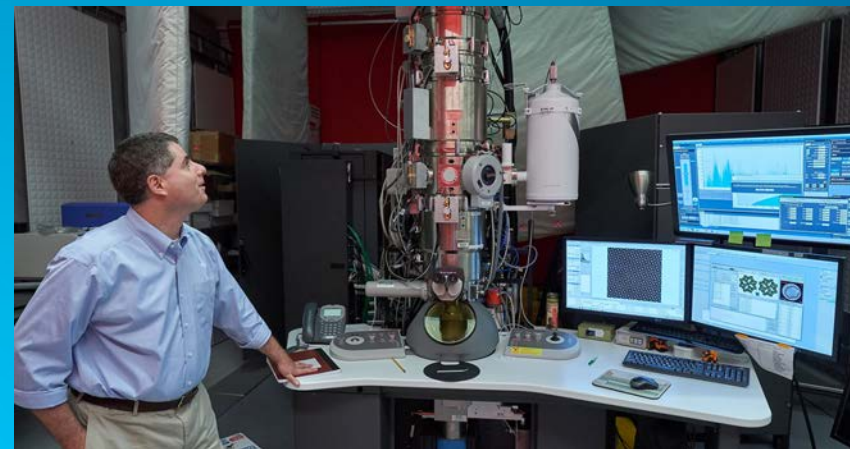
ARTICLE Check for updates

<https://doi.org/10.1038/s41467-020-16688-6> OPEN

## Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose

Zhen Chen<sup>1</sup>, Michal Odstrcil<sup>2,6</sup>, Yi Jiang<sup>3</sup>, Yimo Han<sup>1,7</sup>, Ming-Hui Chiu<sup>4</sup>, Lain-Jong Li<sup>4,8</sup> & David A. Muller<sup>1,5</sup>✉

Both high resolution and high precision are required to quantitatively determine the atomic structure of complex nanostructured materials. However, for conventional imaging methods in scanning transmission electron microscopy (STEM), atomic resolution with picometer precision cannot usually be achieved for weakly-scattering samples or radiation-sensitive materials, such as 2D materials. Here, we demonstrate low-dose, sub-angstrom resolution imaging with picometer precision using mixed-state electron ptychography. We show that correctly accounting for the partial coherence of the electron beam is a prerequisite for high-quality structural reconstructions due to the intrinsic partial coherence of the electron beam. The mixed-state reconstruction gains importance especially when simultaneously pursuing high resolution, high precision and large field-of-view imaging. Compared with conventional atomic-resolution STEM imaging techniques, the mixed-state ptychographic approach simultaneously provides a four-times-faster acquisition, with double the information limit at the same dose, or up to a fifty-fold reduction in dose at the same resolution.





# The Value of Data Changes Over Time

*Data regains value over time if it can be reused*



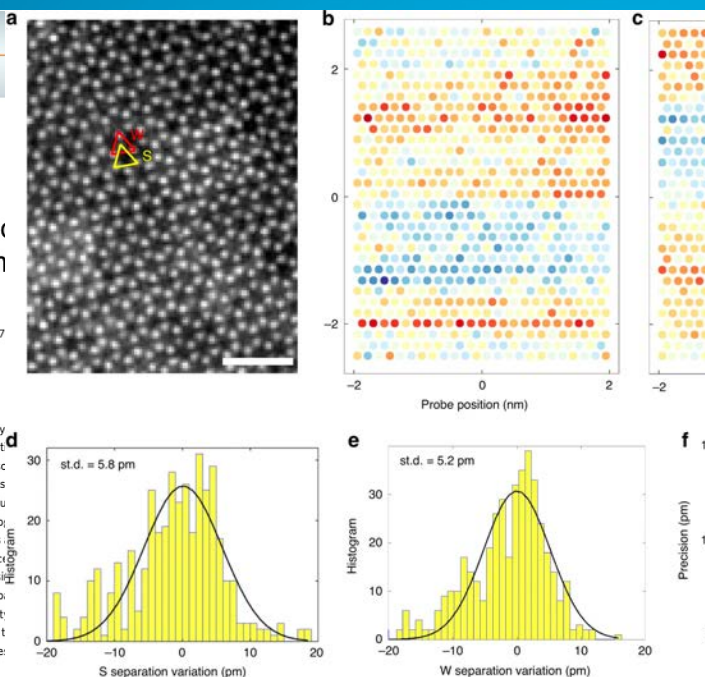
ARTICLE

<https://doi.org/10.1038/s41467-020-16688-6> OPEN

## Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose

Zhen Chen<sup>1</sup>, Michal Odstrcil<sup>2,6</sup>, Yi Jiang<sup>3</sup>, Yimo Han<sup>1,7</sup>, David A. Muller<sup>1,5,8\*</sup>

Both high resolution and high precision are required to quantitatively structure of complex nanostructured materials. However, for conventional scanning transmission electron microscopy (STEM), atomic resolution cannot usually be achieved for weakly-scattering samples materials, such as 2D materials. Here, we demonstrate low-dose, sub-angstrom resolution imaging with picometer precision using mixed-state electron ptychography. Correctly accounting for the partial coherence of the electron beam is quality structural reconstructions due to the intrinsic partial coherence. The mixed-state reconstruction gains importance especially when sub-angstrom resolution, high precision and large field-of-view imaging. Compared to conventional STEM imaging techniques, the mixed-state ptychography simultaneously provides a four-times-faster acquisition, with double the same dose, or up to a fifty-fold reduction in dose at the same resolution.



## Data set: Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose

Zhen Chen, Michal Odstrcil, Yi Jiang, Yimo Han, Ming-Hui Chiu, Lain-Jong Li, David A. Muller

Item	Link
Input Data (Matlab mat)	<a href="#">rawdata_1x_crop.mat</a>
Output Result (png)	<a href="#">pty_crop_phase.png</a>
Analysis Code 1	<a href="https://github.com/muller-group-cornell/ptychography">https://github.com/muller-group-cornell/ptychography</a>
Analysis Code 2	<a href="https://www.psi.ch/en/sls/csaxs">https://www.psi.ch/en/sls/csaxs</a>



<http://doi.org/10.34863/g4wa-0j57>



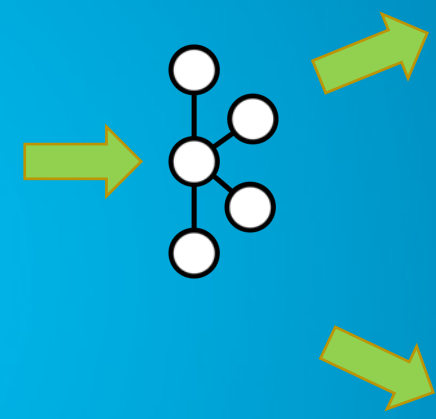
# Seamless to User

*Data regains value over time if it can be reused*

- Streaming Data
- Asynchronous
- Loosely Coupled
- Header metadata to DB
- PID

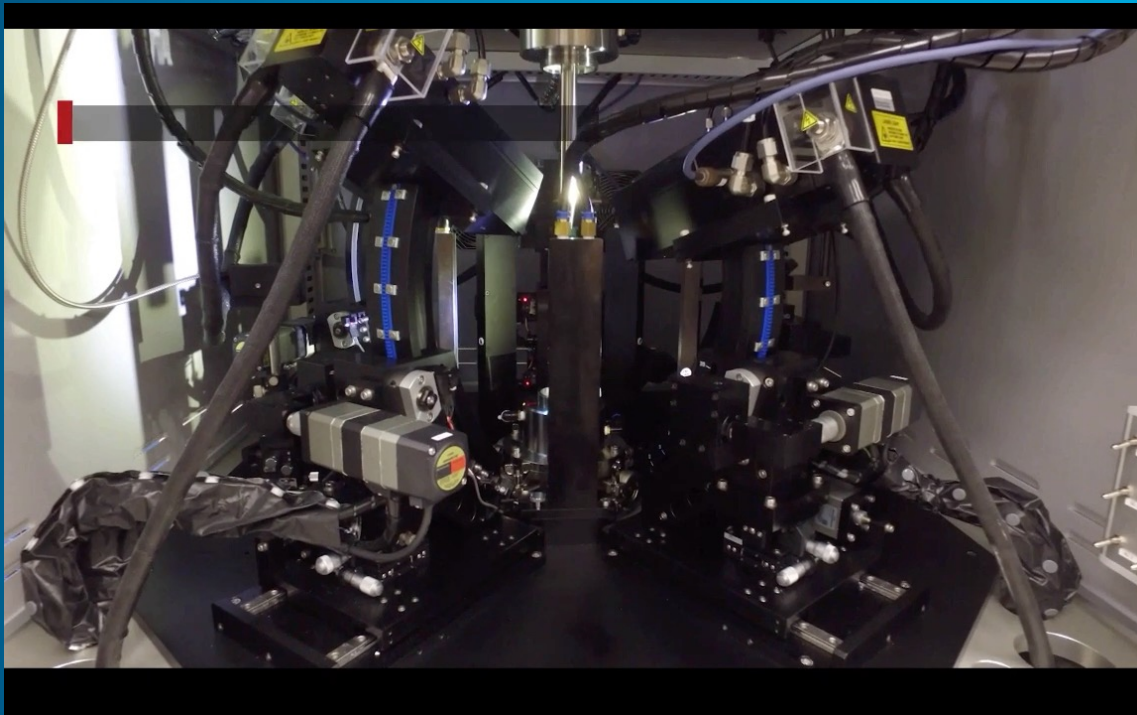


```
# Data processing loop
threads = []
try:
    client_ = conn.accept()
except:
    client_socket = None
    cmd_part = ""
    metadata = {}
    while cmd_part != None:
        threads = ListOfList(items=[a.is_alive(), threads])
        if len(threads) == 0:
            logging.warning("Currently %s files being prepared for upload.", len(threads))
            upload_threads = ListOfList(items=[a.is_alive(), upload_threads])
            if not upload_queue.empty():
                logging.warning("Currently %s chunks and metadata left to upload by %s workers.", upload_queue.qsize(), len(threads))
            if len(threads) == 0 and upload_queue.empty():
                logging.warning("No current uploads.")
            while cmd_part != None and cmd.find("r") == -1 and cmd.find("u") == -1:
                try:
                    with tempfile.TemporaryFile() as f:
                        for req in upload_queue:
                            client_socket.send(req)
                            client_socket.recv(4096)
                            cmd_part = None
                            except IOError:
                                cmd_part = None
                                if cmd_part != None:
                                    lines = cmd.splitlines()
                                    if not cmd.endswith("r") and not cmd.endswith("u") and len(lines) == 0:
                                        # Last entry is a partial line, remove it and save for next time
                                        cmd = lines.pop()
                                    else:
                                        cmd = ""
                                for i in range(1, len(lines)):
                                    if all(len(l) >= len(upload) and len(l) >= 7):
                                        file = all(l)
                                        logging.info("Uploading %s", file)
                                        # Determine how many to send
                                        at = Thread(target=upload_file, args=(file, metadata))
                                        at.daemon = True
                                        at.start()
                                        threads.append(at)
                                        metadata = {}
                                    elif all(len(l) >= len(upload) and all(len(l) >= 7)):
                                        bp = all(l)
                                        if bp and not bp.endswith("\n"):
                                            bp = all(bp)
                                        if len(bp) == 0:
                                            if len(bp) == 0:
                                                metadata[bp] = bp
                                            else:
                                                metadata.append(bp)
                                        else:
                                            logging.warning("Invalid metadata key in command %s", all)
                                        at.join()
                                        logging.warning("Invalid metadata command %s", all)
                                client_socket.close()
finally:
    conn.close()
```



# Data Reuse Can Bring Unexpected Value

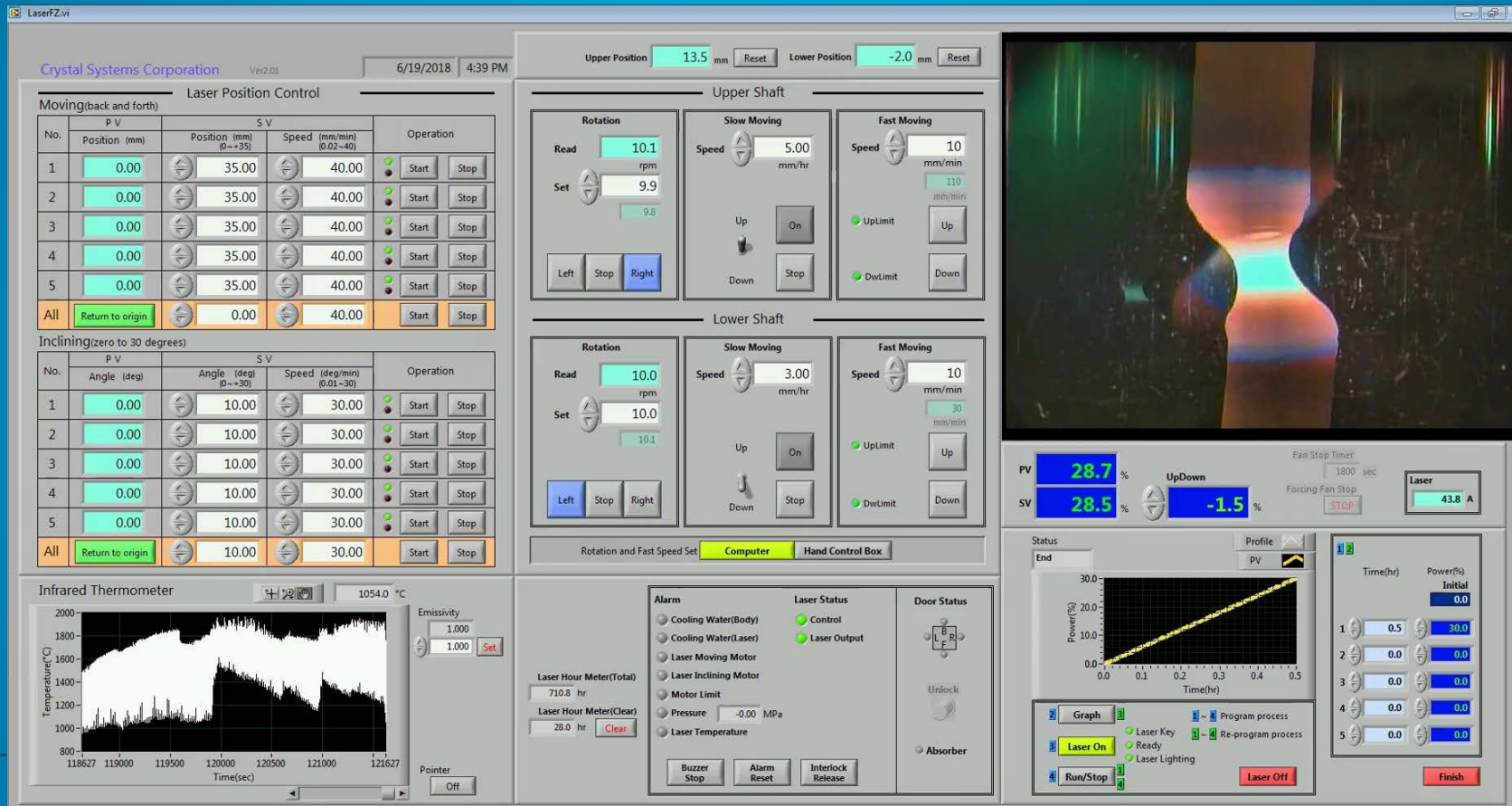
*Data regains value over time if it can be reused*





# Data Reuse Can Bring Unexpected Value

*Data regains value over time if it can be reused*





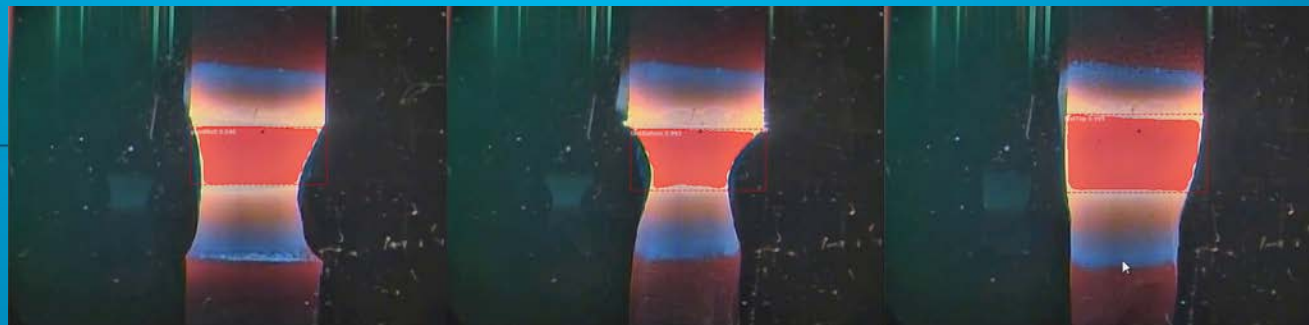
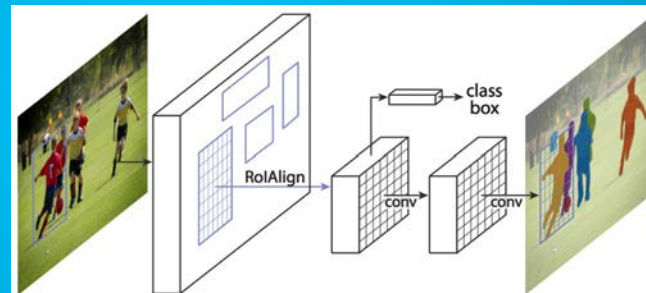


# Data Reuse Can Bring Unexpected Value

*Data regains value over time if it can be reused*

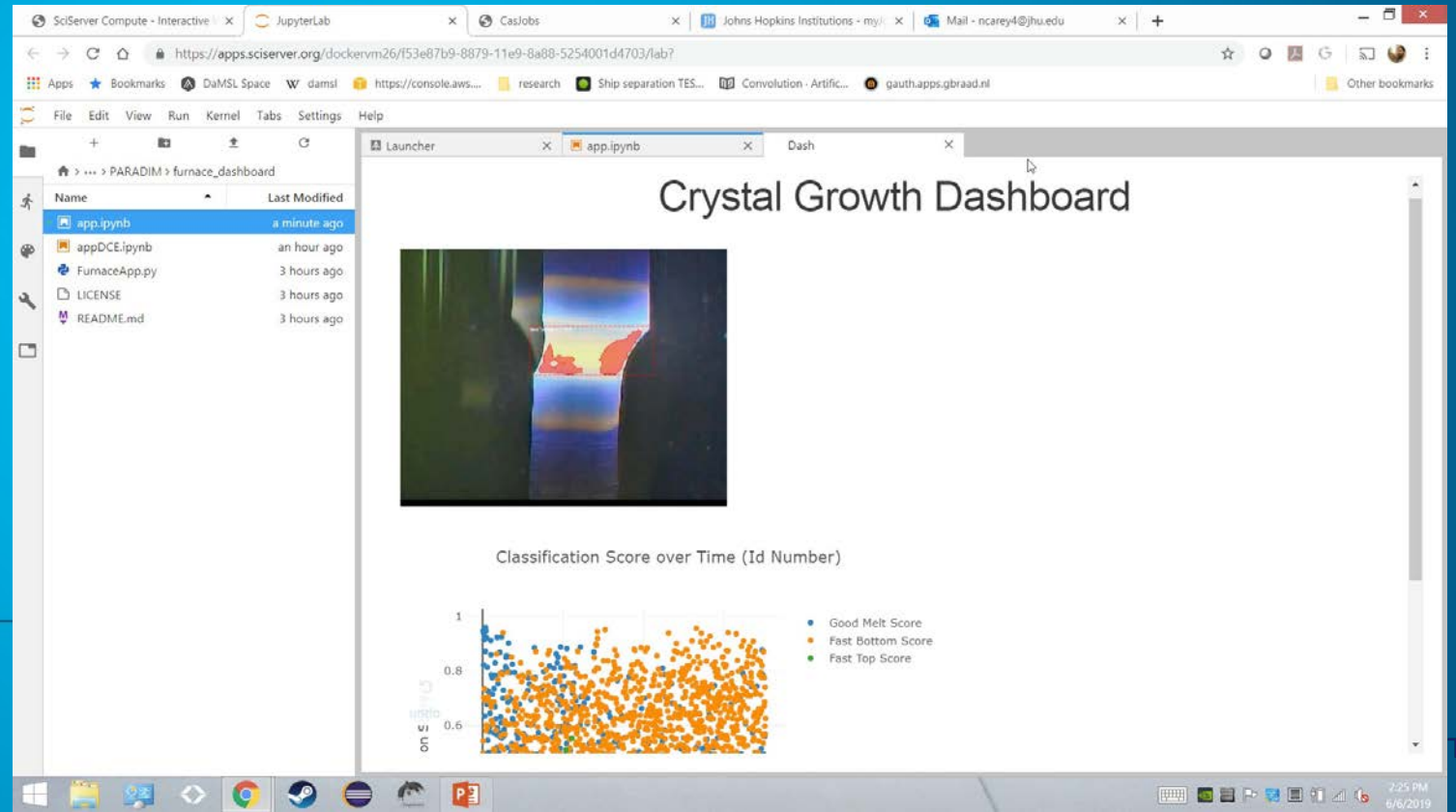


Mask R-CNN  
He et al, 2017



# Data Reuse in Real Time Deployment

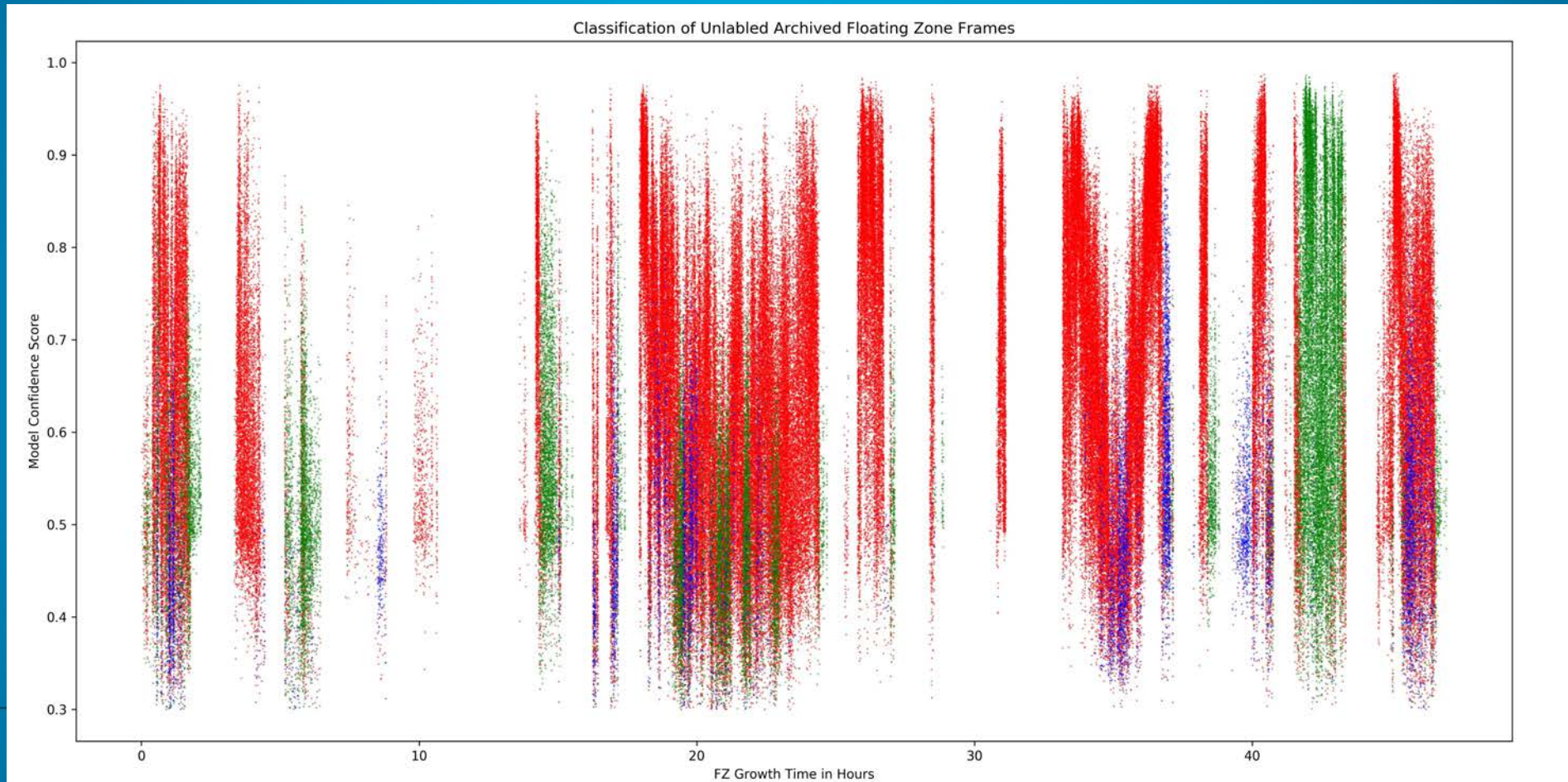
*Data regains value over time if it can be reused*





# Data Reuse in Real Time Deployment

*Data regains value over time if it can be reused*



# Power, Practice, Leadership

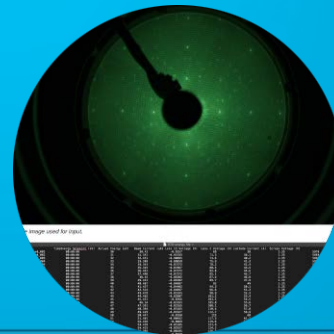
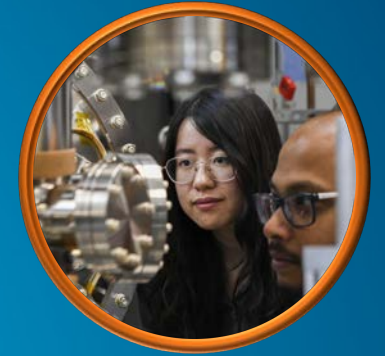
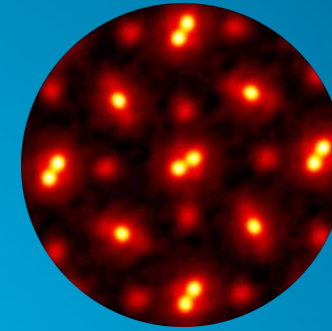
- FAIR is a critical enabler of the MGI  
the currency of Materials Data Infrastructure





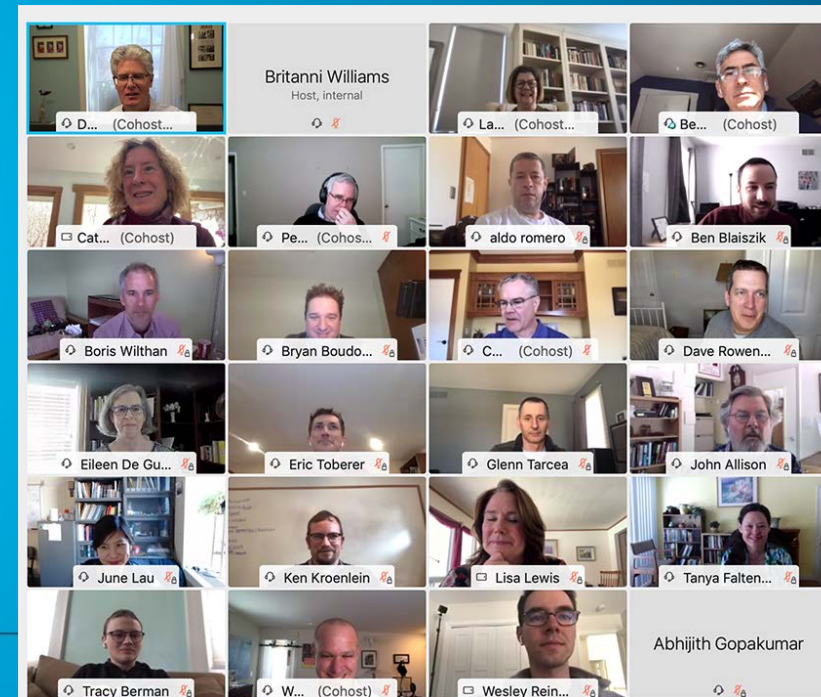
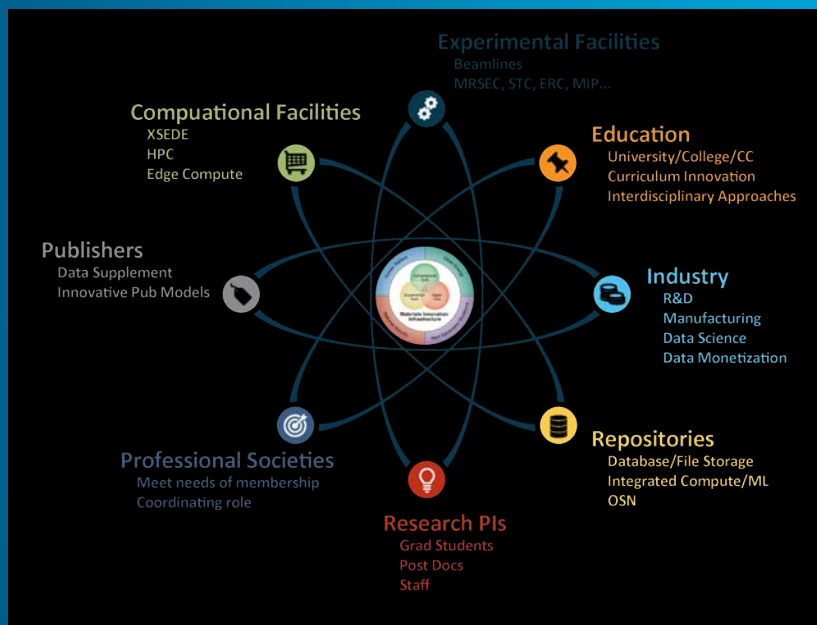
# Power, Practice, Leadership

- PARADIM is strategic for FAIR production
  - center of data production
  - microcosm of domain
  - many users
  - embedded training
  - in-house program



# Power, Practice, Leadership

- FAIR needs community





# FAIR-ification: All Data is not Equal

## PARADIM Data Domains:

- **physical samples** – digital twin/digital thread
- **synthesis recipes** – intent and realization
- **characterization** – structure and properties
- **codes/methods** – software, notebooks, infrastructure
- **reports** – publications and presentations
- **training materials** – lectures, problems, infrastructure

FAIR is a set of principles, but in practice depends on the type of data and the needs or desires of those sharing the data.

# FAIR Today Fairer Tomorrow

## F: Searchable Metadata

- DataCite Search
- Google Data Search (schema.org)
- Web/Pub Links
- Digital Object Id (DOI)

## A: DataCite API

- Year Embargo
- Individual Files

## I: Community Standards

- File Formats
- Open Tools

## R: License and Relevant Metadata

- CC-4.0-BY-NC-ND
- in development

FAIR principle	FAIR Maturity Indicator
F1	Data and metadata identifiers are unique and persistent
F2	Metadata are structured (weak) or grounded in shared vocabularies (strong)
F3	Data and metadata identifiers are included explicitly in metadata
F4	Searchable in web-based search engines
A1.1	Uses open free protocol for data & metadata retrieval
A1.2	Data and metadata authentication and authorization
A2	Metadata persistence
I1	Data and metadata knowledge representation language (weak or strong)
I2	Metadata uses FAIR vocabularies or ontologies (weak or strong)
I3	Metadata contains qualified outward references
R1.1	Metadata includes a license for data usage (weak or strong)
R1.2	Metadata includes provenance (weak)
R1.3	Metadata contains community standards (weak)

weak - readable only by humans; strong - readable by machines

FAIR Data Maturity Model Specifications and Guidelines  
<https://doi.org/10.15497/rda00050>



# FAIR Today Fairer Tomorrow

## F: Searchable Metadata

- DataCite Search
- Google Data Search (schema.org)
- Web/Pub Links
- Digital Object Id (DOI)

## A: DataCite API

- Landing Page
- Year Embargo
- Individual Files

FAIR principle	FAIR Maturity Indicator
F1	Data and metadata identifiers are unique and persistent
F2	Metadata are structured (weak) or grounded in shared vocabularies (strong)
F3	Data and metadata identifiers are included explicitly in metadata
F4	Searchable in web-based search engines
A1.1	Uses open free protocol for data & metadata retrieval
A1.2	Data and metadata authentication and authorization
A2	Metadata persistence
I1	Data and metadata knowledge representation language (weak or strong)
I2	Metadata uses FAIR vocabularies or ontologies (weak or strong)
I3	Metadata contains qualified outward references
R1.1	Metadata includes a license for data usage (weak or strong)
R1.2	Metadata includes provenance (weak)
R1.3	Metadata contains community standards (weak)

weak - readable only by humans; strong - readable by machines

FAIR Data Maturity Model Specifications and Guidelines  
<https://doi.org/10.15497/rda00050>

# FAIR Today Fairer Tomorrow

## F: Searchable Metadata

- DataCite Search
- Google Data Search (schema.org)
- Web/Pub Links
- Digital Object Id (DOI)

## A: DataCite API


- Landing Page
- Year Embargo
- Individual Files

data.paradim.org/doi/pvfm-0y37/

### Epitaxial stannate pyrochlore thin films: Limitations of cation stoichiometry and electron doping

Felix V. E. Hensling, Diana Dahliah, Prabin Dulal, Patrick Singleton, Jiaxin Sun, Jürgen Schubert, Hanjoong Paik, Indra Subedi, Biwas Subedi, Gian-Marco Rignanese, Nikolas J. Podraza, Geoffroy Hautier, and Darrell G. Schlom

We have studied the growth of epitaxial films of stannate pyrochlores with a general formula  $A_2Sn_2O_7$  ( $A = La$  and  $Y$ ) and find that it is possible to incorporate 25% excess of the A-site constituent; in contrast, any tin excess is expelled. We unravel the defect chemistry, allowing for the incorporation of excess A-site species and the mechanism behind the tin expulsion. An A-site surplus is manifested by a shift in the film diffraction peaks, and the expulsion of tin is apparent from the surface morphology of the film. In an attempt to increase  $La_2Sn_2O_7$  conductivity through n-type doping, substantial quantities of tin have been substituted by antimony while maintaining good film quality. The sample remained insulating as explained by first-principles computations, showing that both the oxygen vacancy and antimony to tin substitutional defects are deep. Similar conclusions are drawn on  $Y_2Sn_2O_7$ . An alternative n-type dopant, fluorine on oxygen, is shallow according to computations and more likely to lead to electrical conductivity. The bandgaps of stoichiometric  $La_2Sn_2O_7$  and  $Y_2Sn_2O_7$  films were determined by spectroscopic ellipsometry to be 4.2 eV and 4.48 eV, respectively.



Growth Data		
Item	Type	File
MBE Raw Growth Files - 48 samples	zipped folder	<a href="#">GrowthData.zip</a>

La2Sn2-xSbxO7 RHEED Data		
Item	Type	File
Sample 4	tif	<a href="#">FH004_end_green.tif</a>
Sample 5	png	<a href="#">FH05_right_angle.png</a>
Sample 5	png	<a href="#">FH005end.png</a>
Sample 5	tif	<a href="#">Image13.tif</a>
Sample 6	tif	<a href="#">FH006_end.tif</a>
Sample 7	tif	<a href="#">FH007_end_green.tif</a>
Sample 13	tif	<a href="#">FH088_end_green.tif</a>
Sample 13	tif	<a href="#">FH088_end.tif</a>
Sample 14	bmp	<a href="#">endFH089.bmp</a>
Sample 14	png	<a href="#">endFH089.png</a>
Sample 15	png	<a href="#">End.png</a>
Sample 16	tif	<a href="#">FH091_end.tif</a>
Sample 29	bmp	<a href="#">FH104.bmp</a>

La2Sn2-xSbxO7 XRD Data		
Item	Type	File
Sample 3	csv	<a href="#">2ThetaOmega27_33.csv</a>
Sample 3	csv	<a href="#">5_75_XRD.csv</a>
Sample 3	csv	<a href="#">FH003_2ThetaOmega_57_67.csv</a>
Sample 9	csv	<a href="#">FH022_2Theta_Omega_5_75.csv</a>
Sample 9	csv	<a href="#">FH022_2Theta_Omega_25_35.csv</a>
Sample 9	csv	<a href="#">FH022_2Theta_Omega_57_67.csv</a>
Sample 9	PNG	<a href="#">41_85.PNG</a>
Sample 13	csv	<a href="#">FH088_rocking_better.csv</a>
Sample 13	csv	<a href="#">FH088_rocking_final.csv</a>
Sample 13	csv	<a href="#">FH088_rocking_new.csv</a>
Sample 13	csv	<a href="#">FH088_rocking.csv</a>
Sample 14	PNG	<a href="#">44_2nm.PNG</a>
Sample 14	csv	<a href="#">FH089_2ThetaOmega_5_75.csv</a>
Sample 14	csv	<a href="#">Rocking_222_0_0997.csv</a>
Sample 14	csv	<a href="#">XRR_Quick_Pixel_29.csv</a>
Sample 15	csv	<a href="#">FH090_2ThetaOmega_5_75.csv</a>
Sample 16	csv	<a href="#">FH091_2ThetaOmega_10_70.csv</a>
Sample 22	csv	<a href="#">FH097_2ThetaOmega_24_60.csv</a>
Sample 29	csv	<a href="#">FH104_2ThetaOmega_10_70.csv</a>
Sample 31	csv	<a href="#">FH106_2ThetaOmega_10_70.csv</a>

Y2Sn2-xSbxO7 RHEED Data		
-------------------------	--	--



# FAIR Today Fairer Tomorrow

## I: Community Standards

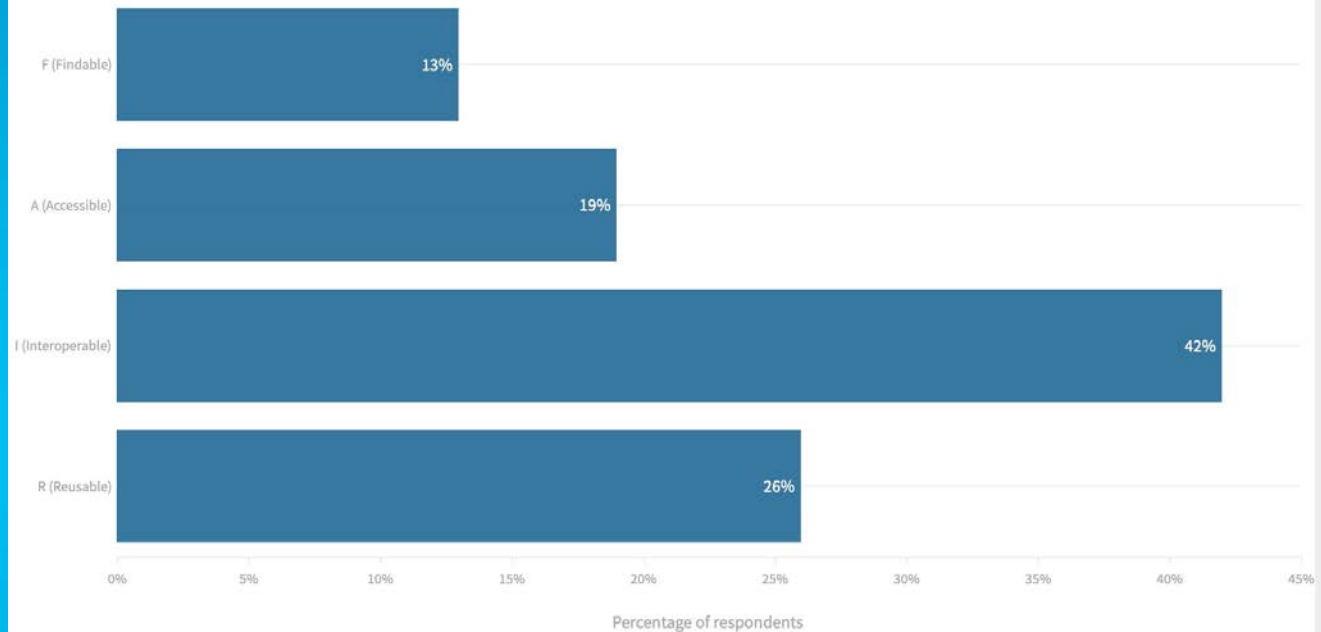
- Standard File Formats
- Open Tools

## R: License and Relevant Metadata

- CC-4.0-BY-NC-ND
- in development

### Which of the FAIR principles do you think most needs better definition?

Interoperability is the least understood FAIR principle. Some 42% of the 187 respondents who answered this question felt that it needed further clarification.



Source: [State of Open Data](#)

source: State of Open Data

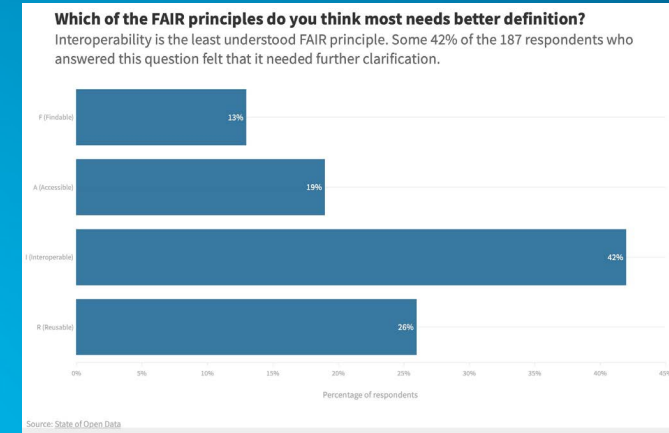
# FAIR Today Fairer Tomorrow

## I: Community Standards

- Standard File Formats
- Open Tools

## R: License and Relevant Metadata

- CC-4.0-BY-NC-ND
- in development



**Attribution-NonCommercial-NoDerivatives 4.0  
International**

# FAIR Today Fairer Tomorrow

## 1. Publication linked

- Characterization
- Reproducibility
- Persistent Identifiers (DOI)

## 2. How do we improve?

### Dataset: Charge order textures induced by non-linear couplings in a half-doped manganite

#### Dataset

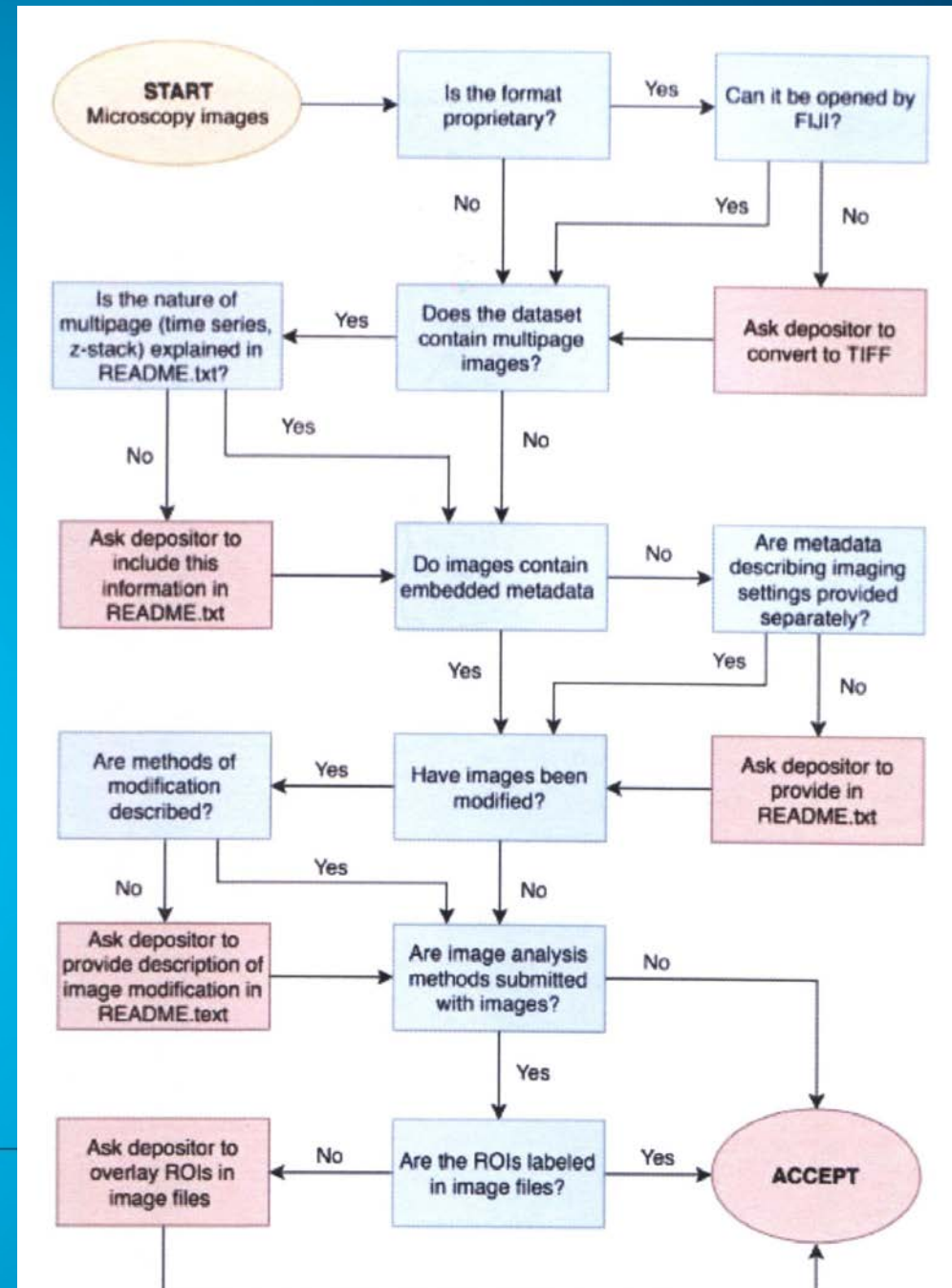
Ismail El Baggari, David Baek, Michael Zachman, Di Lu, Yasuyuki Hikita, Harold Hwang, Elizabeth Nowadnick & Lena Kourkoutis  
Dataset published 2021 via PARADIM, an NSF Materials Innovation Platform

Raw data associated with publication. The self-organization of strongly interacting electrons into superlattice structures underlies the properties of many quantum materials. How these electrons arrange within the superlattice dictates what symmetries are broken and what ground states are stabilized. Here we show that cryogenic scanning transmission electron microscopy (cryo-STEM) enables direct mapping of local symmetries and order at the intra-unit-cell level in the model charge-ordered system  $\text{Nd}_{1/2}\text{Sr}_{1/2}\text{MnO}_3$ . In addition to imaging the prototypical site-centered charge order, we discover the nanoscale coexistence of an exotic intermediate state which mixes site and bond order and breaks inversion symmetry. We further show that nonlinear coupling of distinct lattice modes controls the selection between competing ground states. The results demonstrate the importance of lattice coupling for understanding and manipulating the character of electronic self-organization and that cryo-STEM can reveal local order in strongly correlated systems at the atomic scale. keywords:

Created May 27, 2021, 21:22:23 UTC. [Findable](#)



<https://doi.org/10.34863/bg5n-4s68>





# FAIR Physical Samples

## Digital Object Proxy

- digital twin
- digital thread
- entry point to re-creation



### Sample Tracking and Data Management

Version: 1.0.2

Instrumentation in the laboratory is being moved to a more automated data ingestion and processing system, with corresponding sample tracking. In order for this new system to function as designed, samples must be named with a universal scheme. The uniform format for a new sample is:

AAA\_BBB\_DDMMYYYY\_C\_III\_S\_(QQQQQQQQQ)-EE

you \*MUST\* then use this sample ID as the base filename for all electronic files generated about that material. You can arbitrarily add additional items to the filename, but only \*AFTER\* the sample id.

Item	Definition																				
AAA	Lab Identifier: <table border="1"> <tr> <td>ML</td> <td>McQueen Laboratory</td> </tr> <tr> <td>IQM</td> <td>Institute for Quantum Matter</td> </tr> <tr> <td>PDC</td> <td>PARADIM</td> </tr> </table>	ML	McQueen Laboratory	IQM	Institute for Quantum Matter	PDC	PARADIM														
ML	McQueen Laboratory																				
IQM	Institute for Quantum Matter																				
PDC	PARADIM																				
BBB	Synthesis tool identifier. This is an organic list that evolves over time, and defines the furnace or equipment used to carry out the reaction. For named furnaces/ovens, this is the name of the furnace/oven without spaces (e.g. GobletOfFire or ThinMan). Other synthesis tools currently on the list: <table border="1"> <tr> <td>IDF1</td> <td>PARADIM Induction Furnace</td> </tr> <tr> <td>LDFZ</td> <td>PARADIM Laser Diode Floating Zone Furnace</td> </tr> <tr> <td>HPFZ</td> <td>PARADIM High Pressure Floating Zone Furnace</td> </tr> <tr> <td>CVT1</td> <td>PARADIM CVT Furnace</td> </tr> <tr> <td>XEN1</td> <td>IQM Xenon FZ Furnace</td> </tr> <tr> <td>HALO</td> <td>IQM Halogen FZ Furnace</td> </tr> <tr> <td>MARCC</td> <td>MARCC</td> </tr> <tr> <td>HPCC</td> <td>HPCC</td> </tr> <tr> <td>PPMS</td> <td>McQueen Lab PPMS</td> </tr> <tr> <td>IQMPPMS</td> <td>IQM PPMS</td> </tr> </table> <p>If you use something that is not named and not listed above, let me know and we will either name it or add it to the codes list. Yes, computations (when not associated with existing samples) also get their own identifiers. So too does instrument data when not associated with a specific sample (e.g. a calibration or an addenda measurement).</p>	IDF1	PARADIM Induction Furnace	LDFZ	PARADIM Laser Diode Floating Zone Furnace	HPFZ	PARADIM High Pressure Floating Zone Furnace	CVT1	PARADIM CVT Furnace	XEN1	IQM Xenon FZ Furnace	HALO	IQM Halogen FZ Furnace	MARCC	MARCC	HPCC	HPCC	PPMS	McQueen Lab PPMS	IQMPPMS	IQM PPMS
IDF1	PARADIM Induction Furnace																				
LDFZ	PARADIM Laser Diode Floating Zone Furnace																				
HPFZ	PARADIM High Pressure Floating Zone Furnace																				
CVT1	PARADIM CVT Furnace																				
XEN1	IQM Xenon FZ Furnace																				
HALO	IQM Halogen FZ Furnace																				
MARCC	MARCC																				
HPCC	HPCC																				
PPMS	McQueen Lab PPMS																				
IQMPPMS	IQM PPMS																				
DDMMYYYY	Day, month, year in numerical format, e.g. 26022019 is February 26th, 2019.																				
C	Alphanumeric identifier indicating which sample it is within a given group, tool and day tuple. It runs 1-9, then A-Z. So the 1st sample is "1", the second sample is "2", and the 11th is "B". Should you do more than 35 samples in a single day on a given group, tool, and day tuple, ask me (this is unlikely).																				
III	Provenance identifier. If AAA is ML or IQM, this is your initials (e.g. I would use TMM). If AAA is PDC, this is the proposal																				

# FAIR Physical Samples

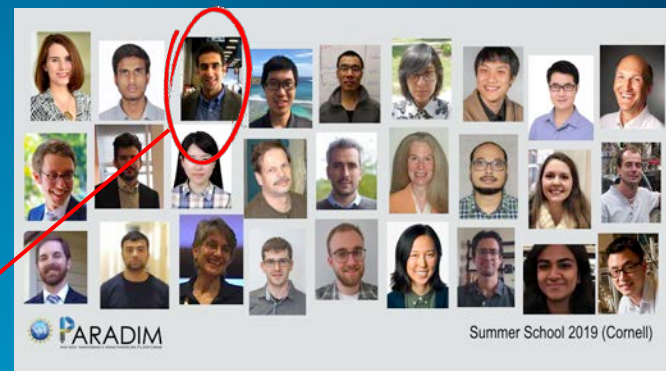
## Digital Object Proxy

- digital twin
- digital thread
- link to location
- entry point to recreation

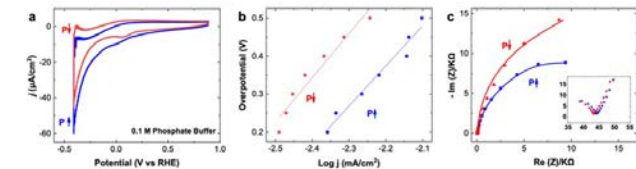
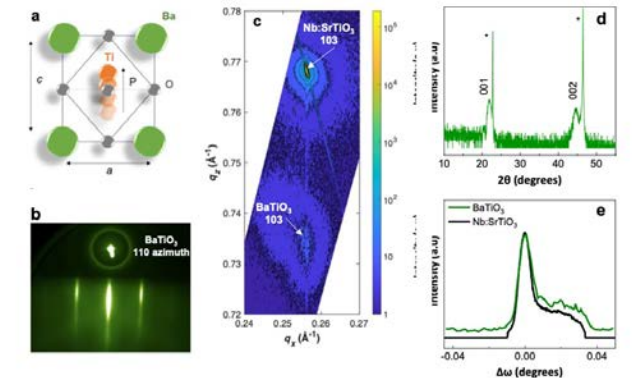
## Proposal 223 – Dynamic Electrocatalysis on Ferroelectric BaTiO<sub>3</sub> Films

- David Fenning, UCSD Nanoengineering Group
- Pedram Abbasi (PhD Student)
- 2019 Summer School

“MBE+ARPES: Customizing Oxides...”



Proposal 223: Dynamic Electrocatalysis on Ferroelectric BaTiO<sub>3</sub> Films  
David Fenning, UCSD



Proposal 223: Abassi et al., submitted Nature Catalysis



# Training for FAIR

## FAIR for Training

### 10 Rules for Making Training Materials FAIR

1. Plan
2. Fully Describe
- F 3. Identifier
4. Register
- A 5. Define Access Rules
- I 6. Interoperable Format
7. Reusable for Trainers (license)
- R 8. Reusable for Trainees (metadata)
9. Contribution Friendly
10. Up-to-date



Garcia et al, PLOS Comp Bio, 2020,  
(<https://doi.org/10.1371/journal.pcbi.1007854>)



# Community <https://marda-alliance.org>

*The Materials Research Data Alliance (MaRDA) is a community-led network focused **on connecting and integrating U.S. materials research data infrastructure to realize the promise of open, accessible, and interoperable materials data.** Each of these elements are aligned with the goals of the Materials Genome Initiative (MGI). MaRDA provides a platform that **promotes the convergence of ideas, people, data, and tools** to accelerate discovery, enable new insights into materials mechanisms, and lay the foundation for both human-centered and artificial intelligence-assisted approaches to materials design. MaRDA is governed by an elected council, **MaRDAC**, that promotes the interests of materials data researchers nationally and internationally, and coordinates the efforts of **MaRDA.***

# Community <https://marda-alliance.org>

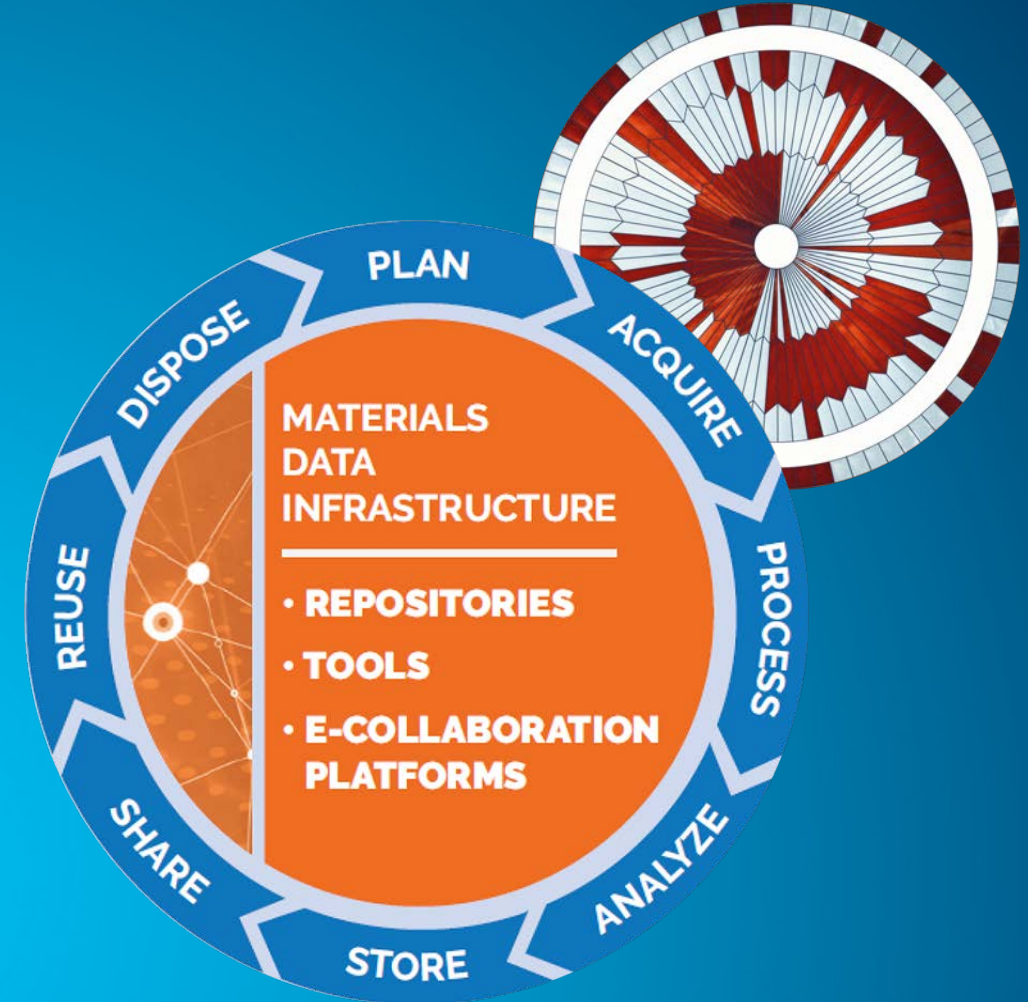
## MaRDA: A Year of Building

- Founding Council
- Clarification of Mission
- Bylaws
- Website



# Forecasting the FAIR Future

- FAIR is about data use and reuse
  - Data type and purpose are key
  - What do we want to empower?
- Can't be sustained by investigators
  - Expectation yes! Burden? no!!
  - Scaffold to infrastructure solution
- Materials research is always evolving
  - FAIR cannot be viewed as static
  - Community effort must be sustained
- Reject lock-in
  - Reward connection
  - Incentivize joint success (national/international)
- FAIR is a Community Effort
  - MaRDA (marda-alliance.org)



TMS 2017 Study  
[https://doi.org/10.7449/mdistudy\\_1](https://doi.org/10.7449/mdistudy_1)

How will a PARADIM user 15 years from now leverage the data the platform creates today?