Facility and Instrument PIDs in the Materials Research Data Domain

Got PID?

Is there a roadmap?

Are we optimistic or pessimistic?

David Elbert: elbert@jhu.edu

CDO PARADIM Materials Innovation Platform (MIP)
ARL HTMDEC Extreme Data PI
2 NSF DMREF Co-PI
DOE Supported Catalysis Co-PI
Materials Research Data Alliance (MaRDA)
MaRCN FAIROS-RCN PI

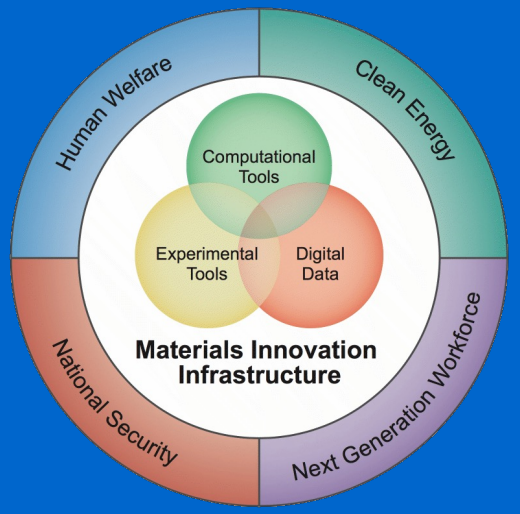
Illegitimi non carborundum

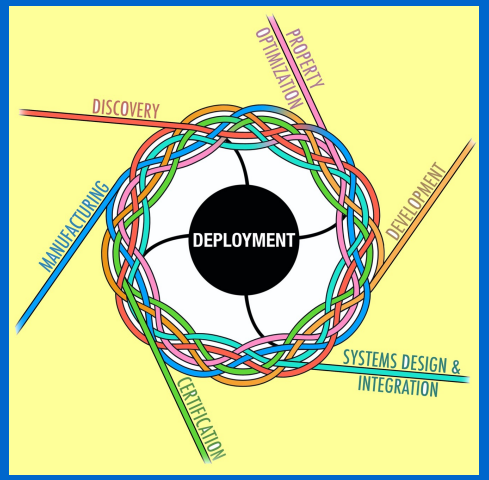
MaRDA Materials Research Data Alliance

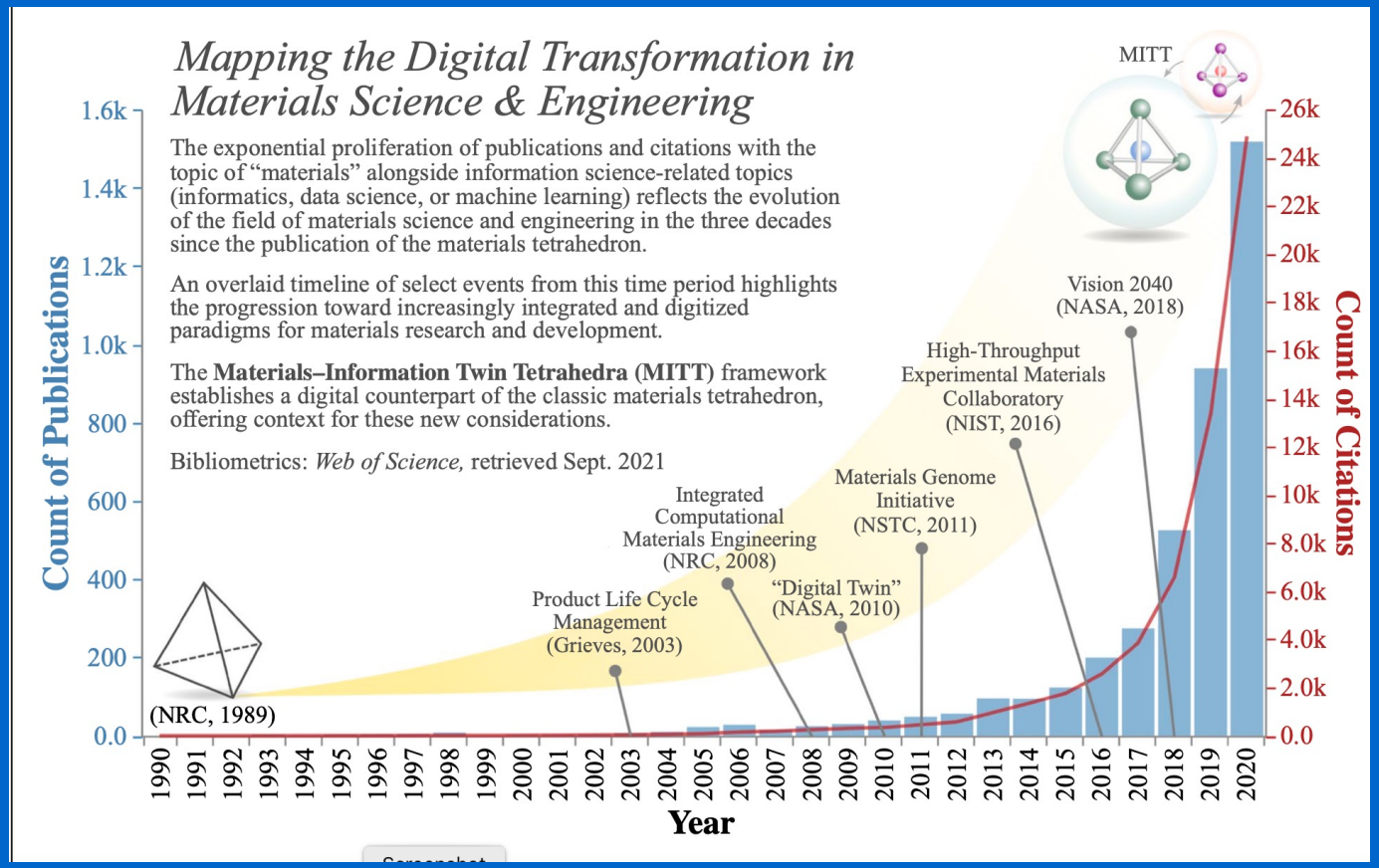
MaRCN

HEMI HOPKINS EXTREME MATERIALS INSTITUTE

Materials Genome Initiative (MGI) Driven Directions







Deagen et al, 2022 MRS Bulletin
doi:10.1557/s43577-021-00214-0

MaRDA Materials Research Data Alliance

MaRCN

HEMI HOPKINS EXTREME MATERIALS INSTITUTE

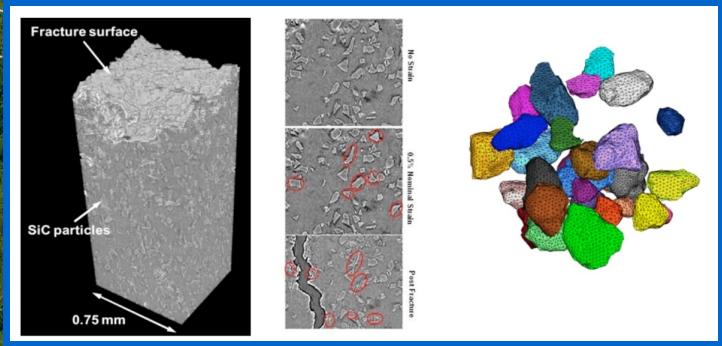
Materials Science Motivations

Firehose of Data

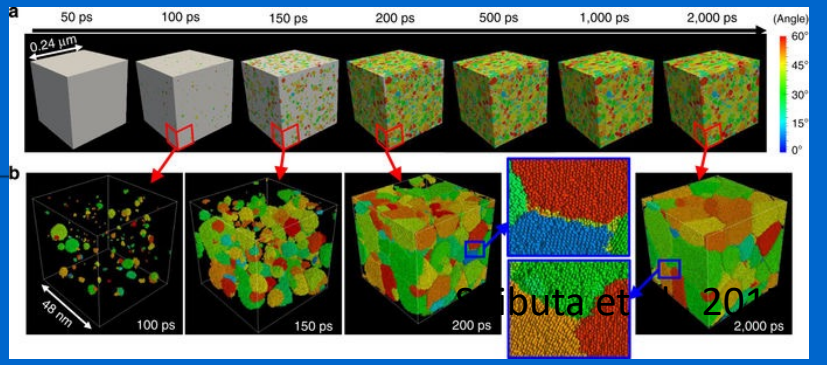
- Higher resolution
- Shorter time scales
- Higher dimensionality
- Dynamic experiments
- Larger simulations
- Tighter processing control

Diverse, Distributed Data

- Experimental
  - labs great and small
- Modeling
  - finite element to ab initio
  - desktops to supercomputers



De Carlo et al., 2012



Sharbuta et al., 201

MaRDA | Materials Research Data Alliance

MaRCN

MaRDA

Materials Research Data Alliance

MaRCN

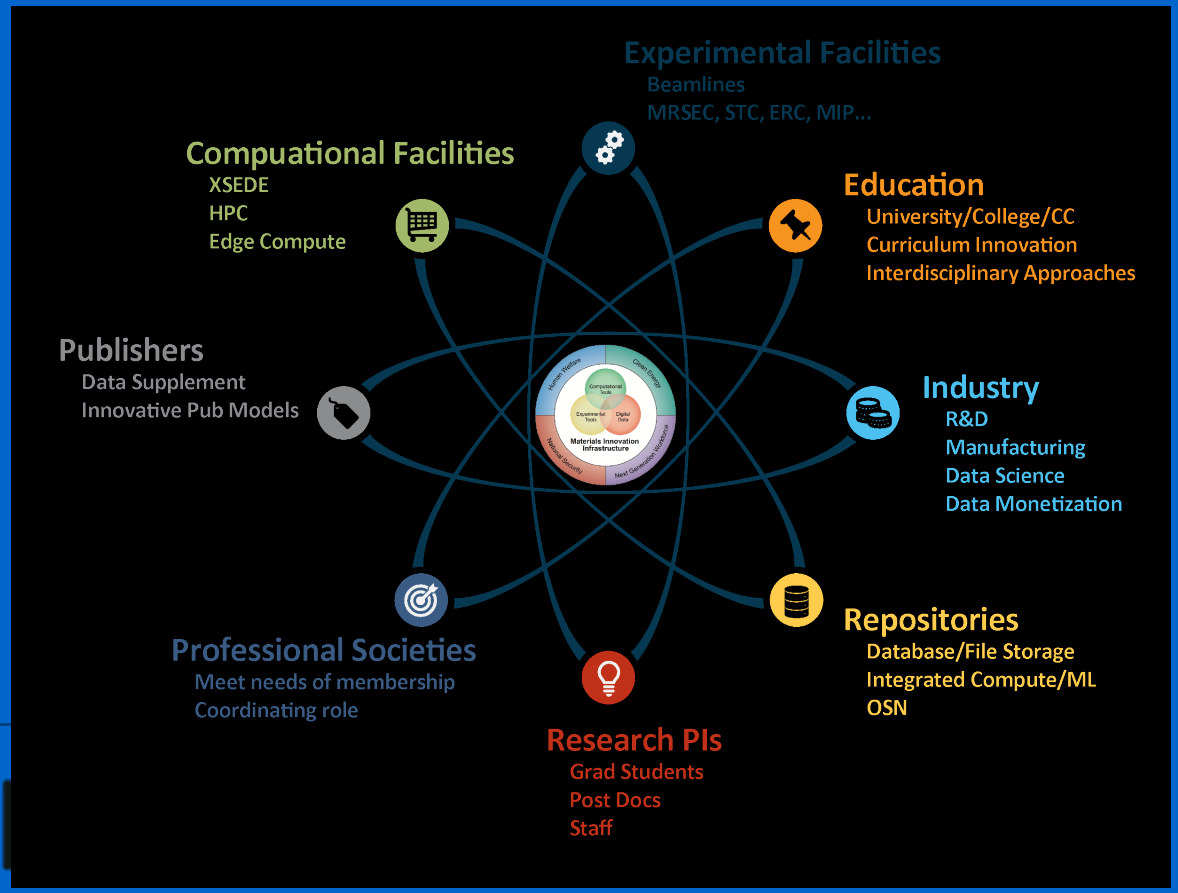
https://marda-alliance.org

MGI 2.0: *People are infrastructure, too!*

2021 Revised Strategic Plan Goal 1, Objective 2:

Unify the Materials Innovation Infrastructure

- Bridge, Build, and Bolster Elements of the MII.
- Establish a National Materials Data Network
- Accelerate through National Grand Challenges



MaRDA | Materials Research Data Alliance

MaRDA

Materials Research Data Alliance

MaRDA is a ***community network*** focused on developing the *open, accessible, and interoperable materials data* that fuels the Materials Genome Initiative (MGI). MaRDA is a **convergence of people and ideas working together** to connect materials data infrastructure to accelerate discovery, enable new insights into materials mechanisms, and lay a foundation for both human-centered and artificial intelligence-assisted approaches to materials design.

FAIR for Materials:

- FAIR is still poorly understood by most of our investigators
- They don't know that FAIR are principles not implementation

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 metdata are released with a clear and accessible data usage license

Adapted from Wilkinson et al., 2016
https://fairtoolkit.pistoiaalliance.org/fair-guiding-principles/

https://doi.org/10.1557/s43577-023-00498-4

MaRDA | Materials Research Data Alliance

MaRCN

HEMI | HOPKINS EXTREME MATERIALS INSTITUTE

FAIR for Materials

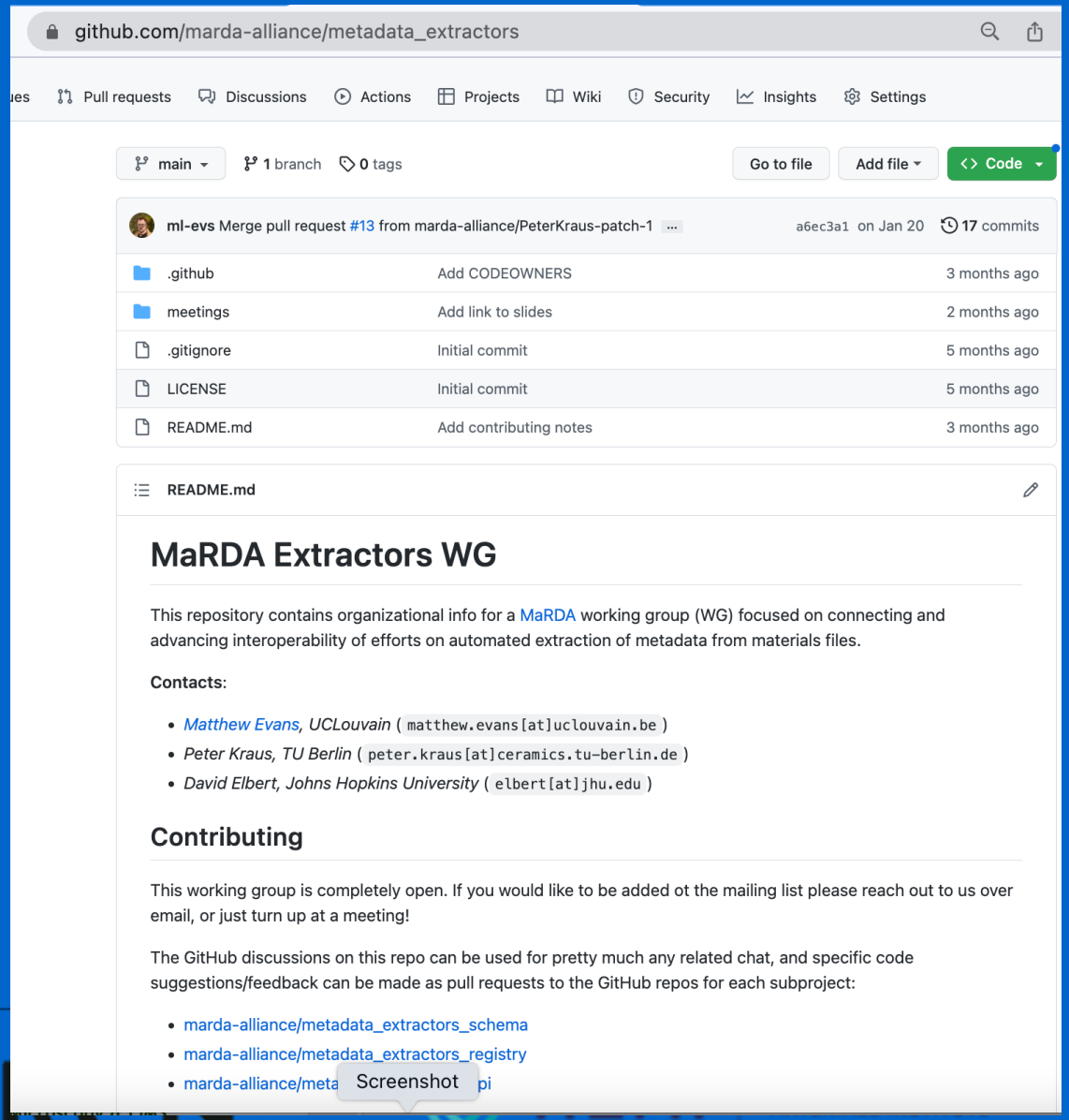
FAIR means metadata, too

https://doi.org/10.1557/s43577-023-00498-4

MaRDA | Materials Research Data Alliance

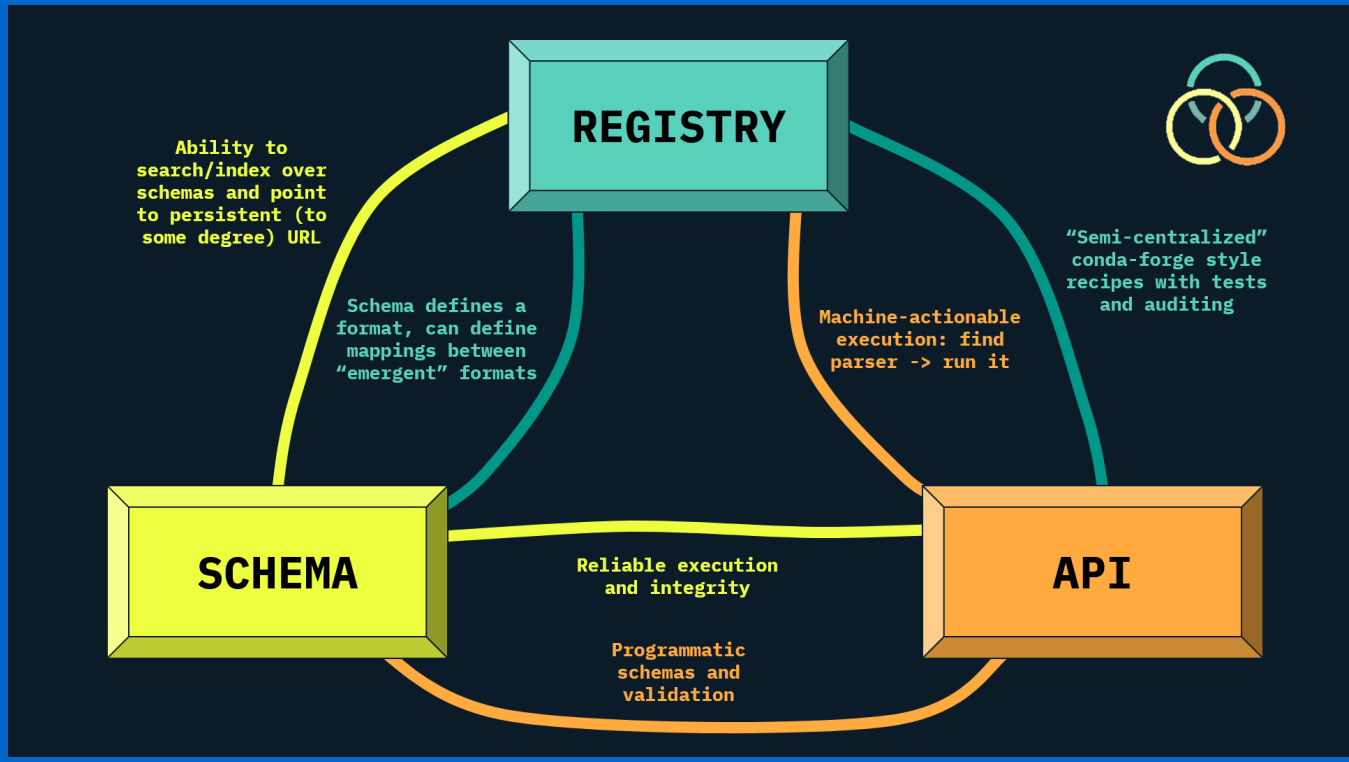
Example: Coordinated Development for Interoperability

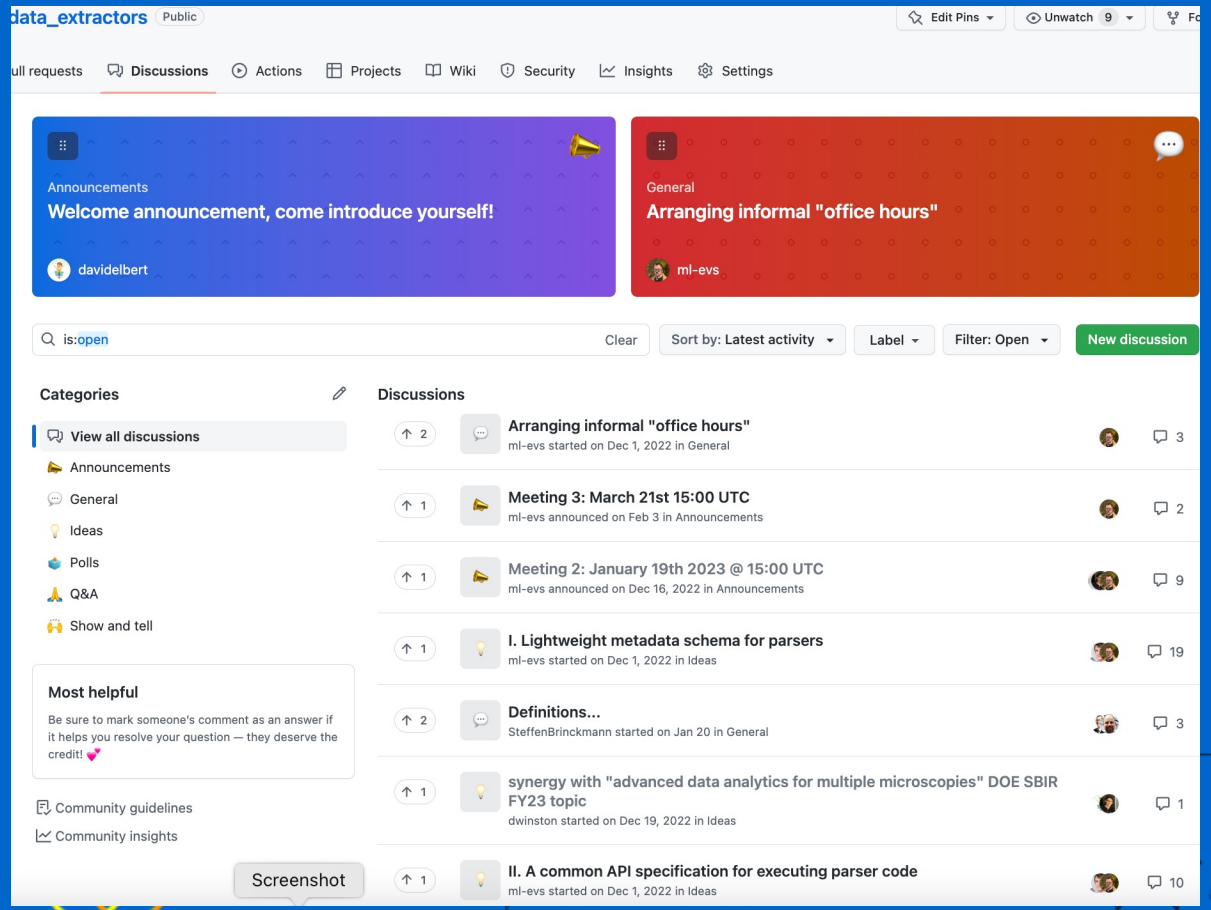
- Metadata Extractors
  - Matthew Evans, UC Louvain
  - Peter Kraus, TU Berlin
  - David Elbert, Johns Hopkins
- Interoperability Layer
- LinkML
  - Translate LinkML instance data to OWL (TBoxes and ABoxes)



MaRDA | Materials Research Data Alliance

- Interoperability Layer
- LinkML
  - Translate LinkML instance data to OWL (TBoxes and ABoxes)





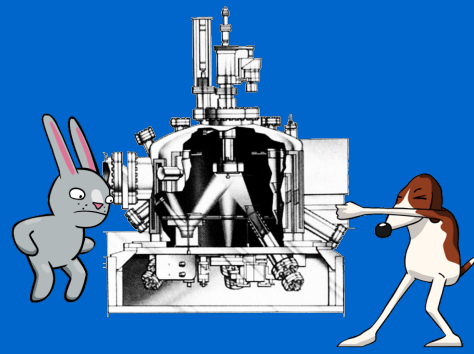
MaRCN

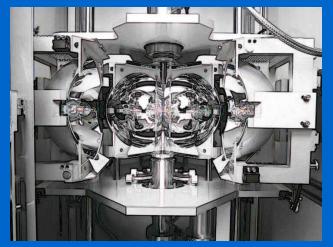
HEMI | HOPKINS EXTREME MATERIALS INSTITUTE

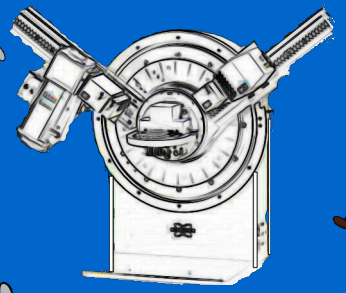
Scientists can be myopic

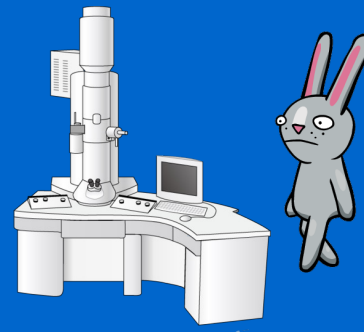
Many Moving Parts

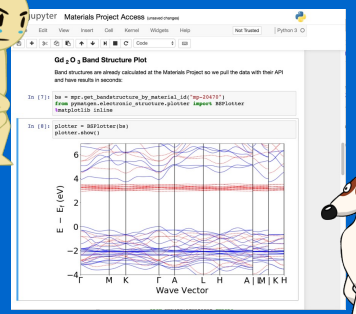
- Equipment
- People
- Ideas

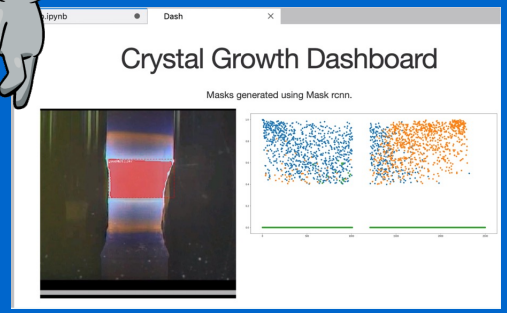


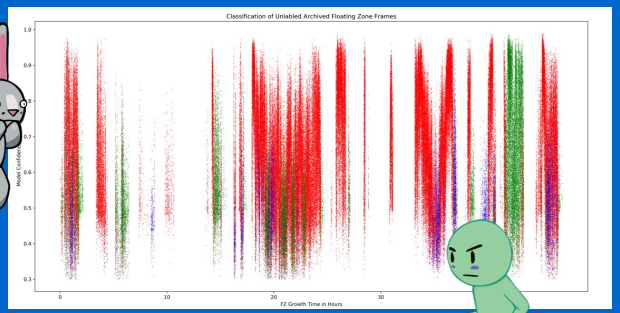












MaRDA | Materials Research Data Alliance

MaRCN

HEMI | HOPKINS EXTREME MATERIALS INSTITUTE

Somas "Data Landscape is Changing!"

- Automating Everything
- High Throughput
- Automation + Decisions = Autonomy
- Linked Data Is Foundational

Scientists are the worst judges of the reuse of their own data

Review Article | Published: 30 January 2023

The rise of self-driving labs in chemical and materials sciences

Milad Abolhasani & Eugenia Kumacheva

Nature Synthesis **2**, 483–492 (2023) | Cite this article

The high-throughput highway to computational materials design

Stefano Curtarolo, Gus L. W. Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito & Ohad Levy

Nature Materials **12**, 191–201 (2013) | Cite this article

43k Accesses | **1357** Citations | **34** Altmetric | Metrics

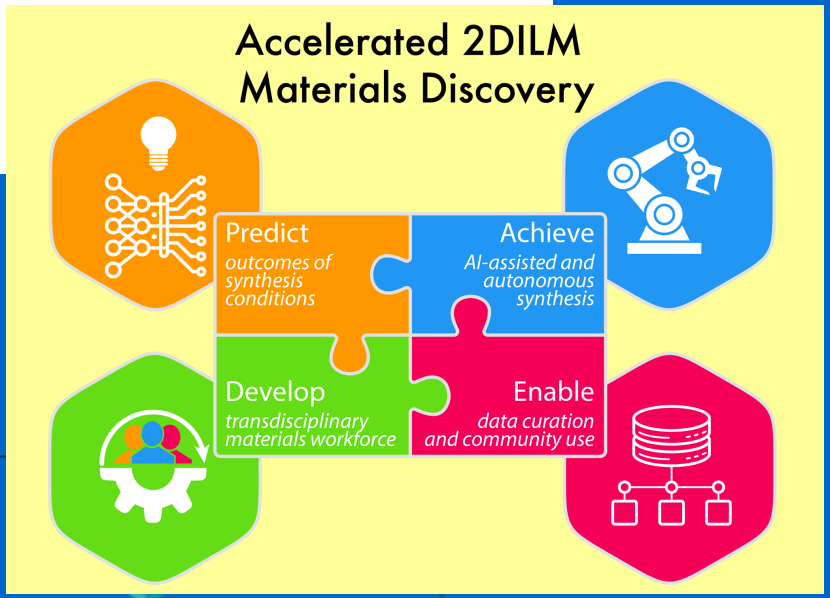
Review Article | Published: 09 March 2023

Combinatorial synthesis for AI-driven materials discovery

John M. Gregoire, Lan Zhou & Joel A. Haber

Nature Synthesis **2**, 493–504 (2023) | Cite this article

1258 Accesses | **5** Altmetric | Metrics



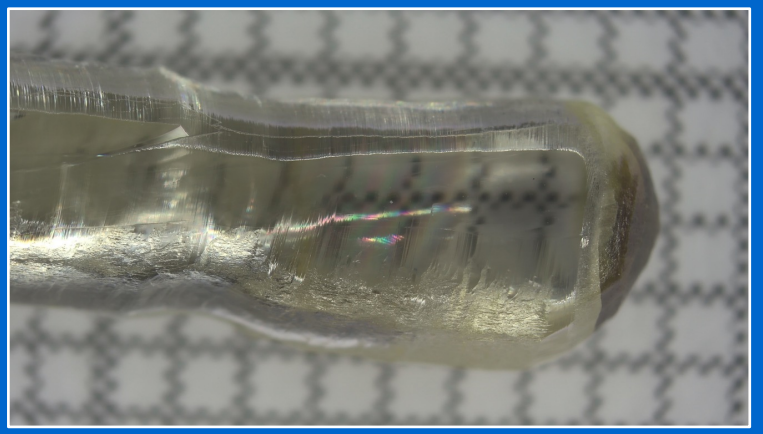
NSF Data-Driven Materials Discovery Future Workshop, 2022

MaRDA Materials Research Data Alliance

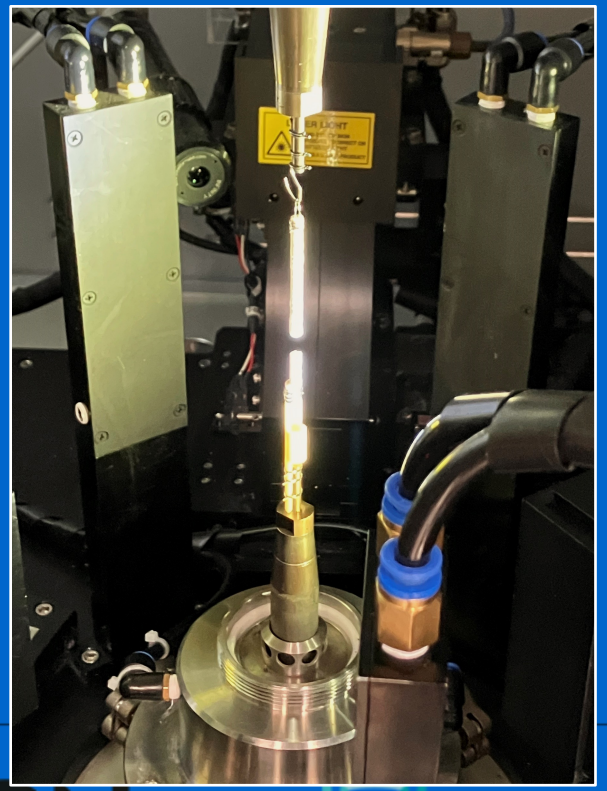
MaRCN

Linking Data GEMD Graphical Model

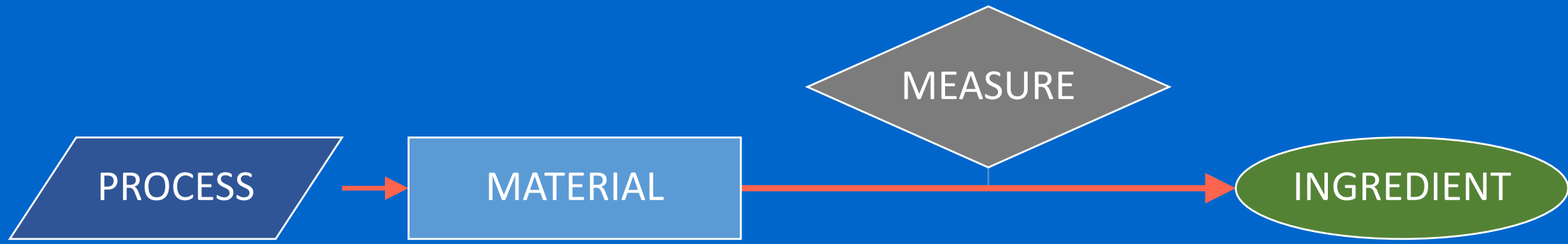
Yttrium Orthovanadate YVO_4



Gannon Murray PARADIM
REU from Earlham College



Graphical **E**xpression of **M**aterials **D**ata (GEMD)

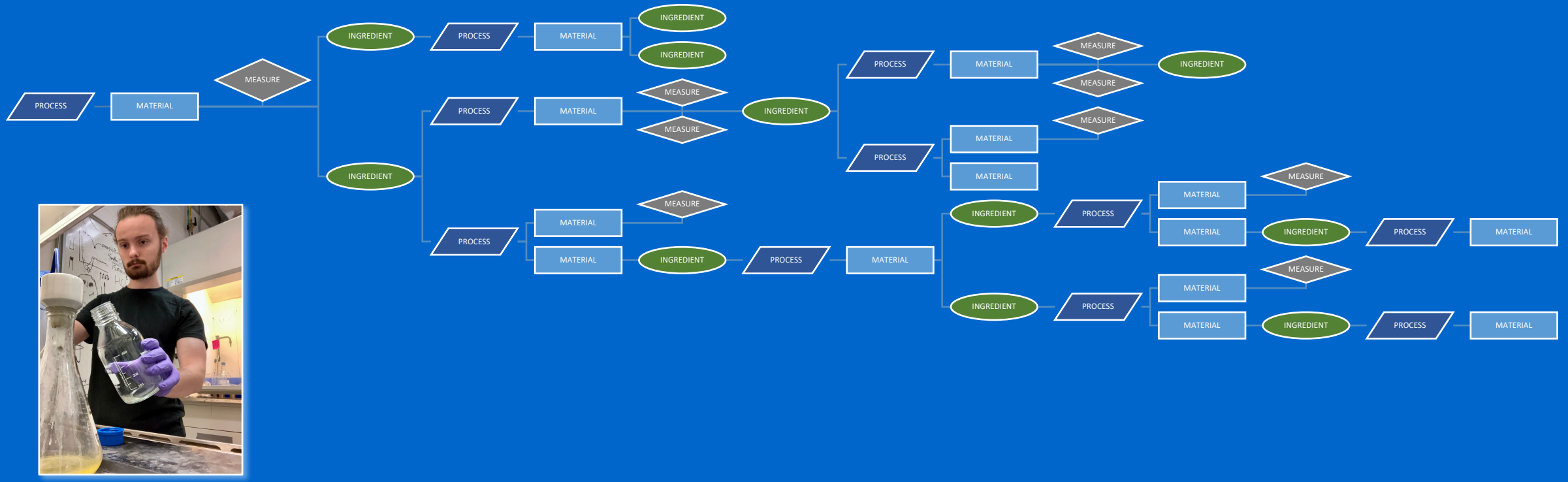


Data driven approaches are transformative

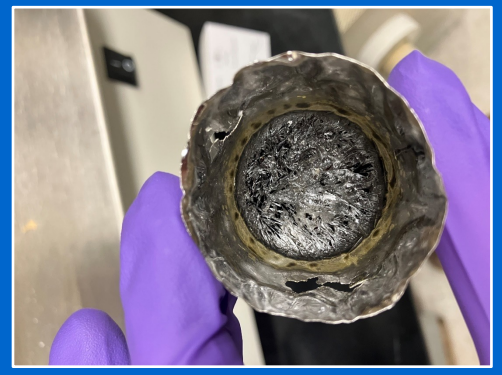
but don't underestimate that:

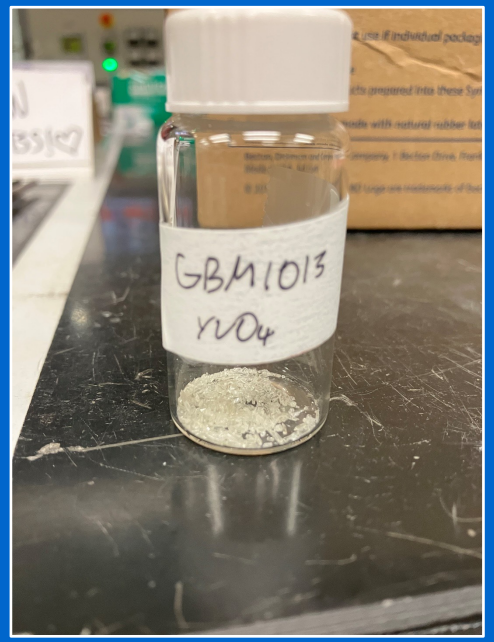
Connected data is unknown territory

Creating a Material History



Gannon Murray PARADIM
REU from Earlham College





GBM1013 Heating Process, fc2

GBM1013 Heated Material, 2ff

YVO4 LIVO3 Flux Ingredient, 694

GBM1013 Dissolving Process, ae5

GBM1013 Dissolved Material, 884

GBM1013 Ingredient, 12d

GBM1013 Filtering Process, f41

GBM1013 Filtered Material, d5c

GBM1013 YVO4 Ingredient, b74

GBM1013_YVO4 Heated Material, e8a

GBM1013_YVO4 Heating Process, abd

GBM1013_YVO4 Ingredient, 62a

GBM1013_YVO4 Ground Material, d90

GBM1013_YVO4 Grinding Process, dae

Y2O3 Ingredient, b4b

V2O5 Ingredient, 993

Y2O3 Acquired Material Base, 617

V2O5 Acquired Material Base

Y2O3 Acquisition Process, 120

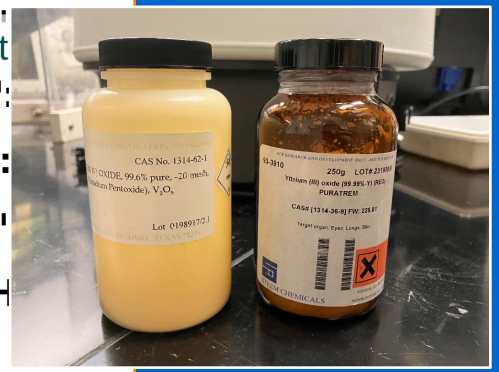
V2O5 Acquisition Process,

"Location, Hot Lab":"False"
"Step 1, {'quantities': {'Duration': 0.0,
'Temp': 0.0, 'Type': 'Init'}, 'type':
'nominal_composition'}":"False"
"Step 2, {'quantities': {'Duration': 9.0,
'Temp': 900
'nominal_c
"Step 3, {'quantities': {'Duration': 1.0,
'Temp': 900.0
'nominal_c
"Step 4, {'q
'Temp': 0.0, 'Type': 'Ramp'}, 'type':
'nominal_c
"Step 5, {'qu
'Temp': 0.0, 'Type': 'End'}, 'type':
'nominal_composition'}":"False"
label:"GBM1013_Flux_Medium H
Process, 5e1"

GBM1013_Flux_Medium Heated Material, 08a

GBM1013_Flux_Medium Heating Process, 5e1

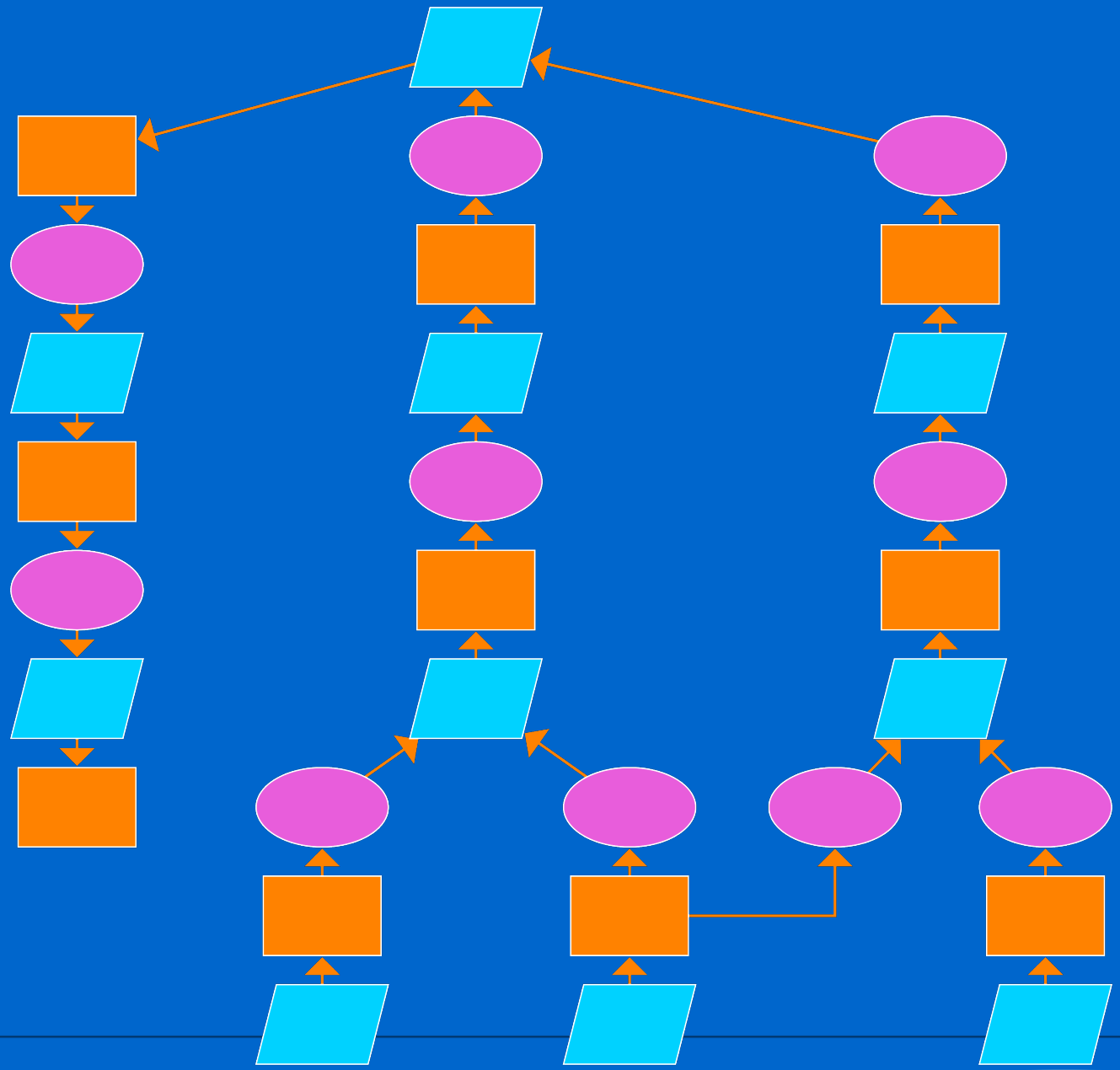
GBM1013_Flux_Medium Ingredient



MaRDA | Materials Research Data Alliance

MaRCN

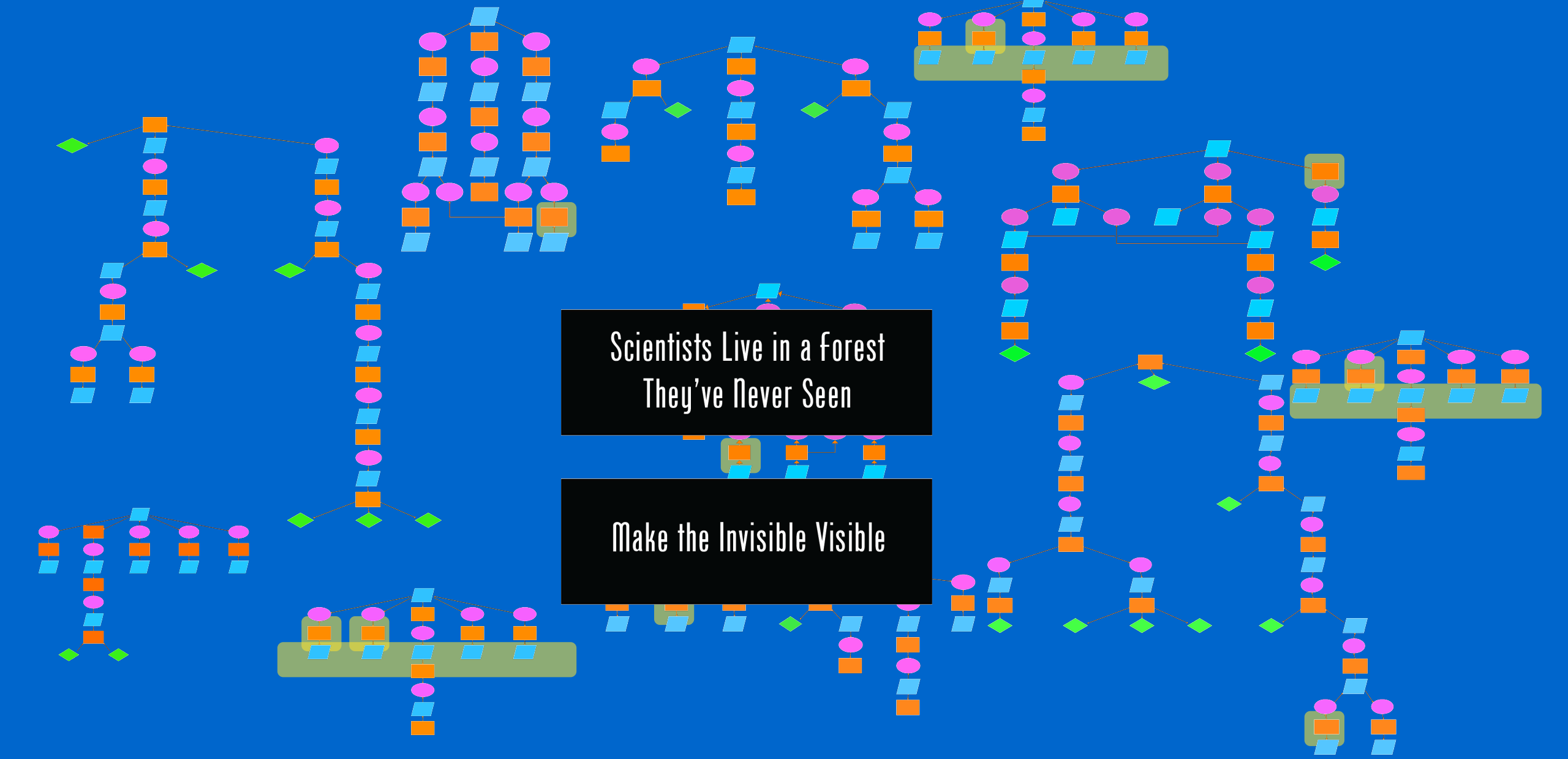
HEMI | HOPKINS EXTREME MATERIALS INSTITUTE



MaRDA | Materials Research Data Alliance

MaRCN

HEMI | HOPKINS EXTREME MATERIALS INSTITUTE



MaRDA | Materials Research Data Alliance

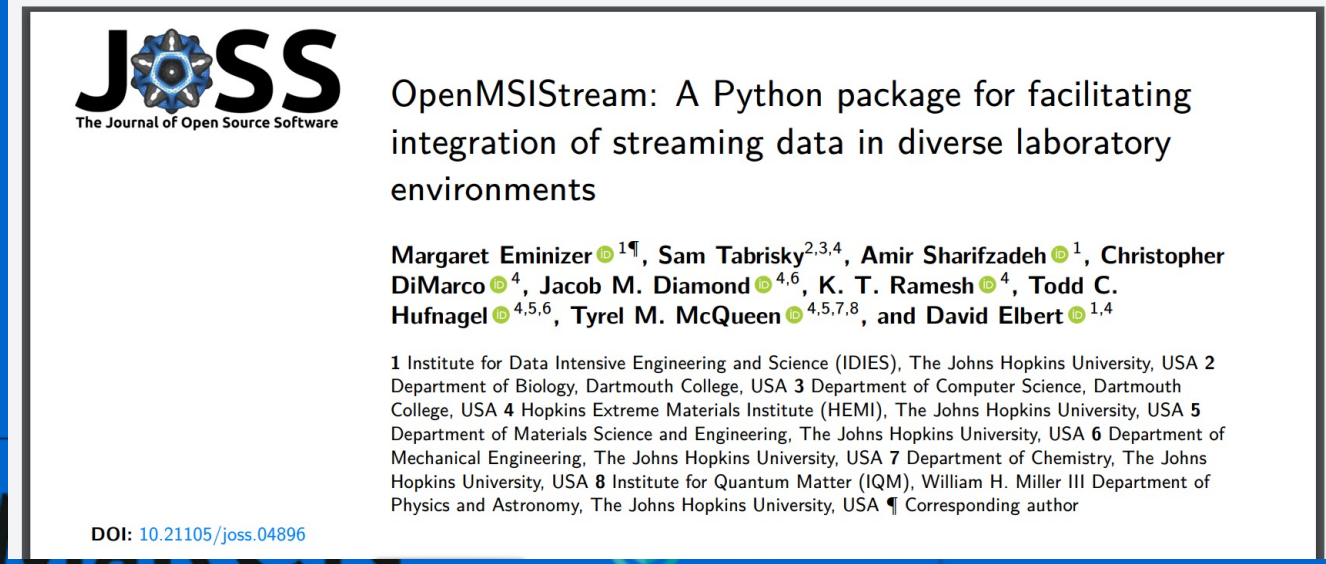
MaRCN

HEMI | HOPKINS EXTREME MATERIALS INSTITUTE

Event Driven Design

- Seamless
  - Dropbox-like entry point
- Automate Anything Repetitive
  - Curation
  - Reduction
  - Active Learning
  - Analysis Pipelines
- Path to Deployment
- Maintainable
- Flexible (Future Foundation)

https://openmsistream.readthedocs.io/



https://doi.org/10.21105/joss.04896

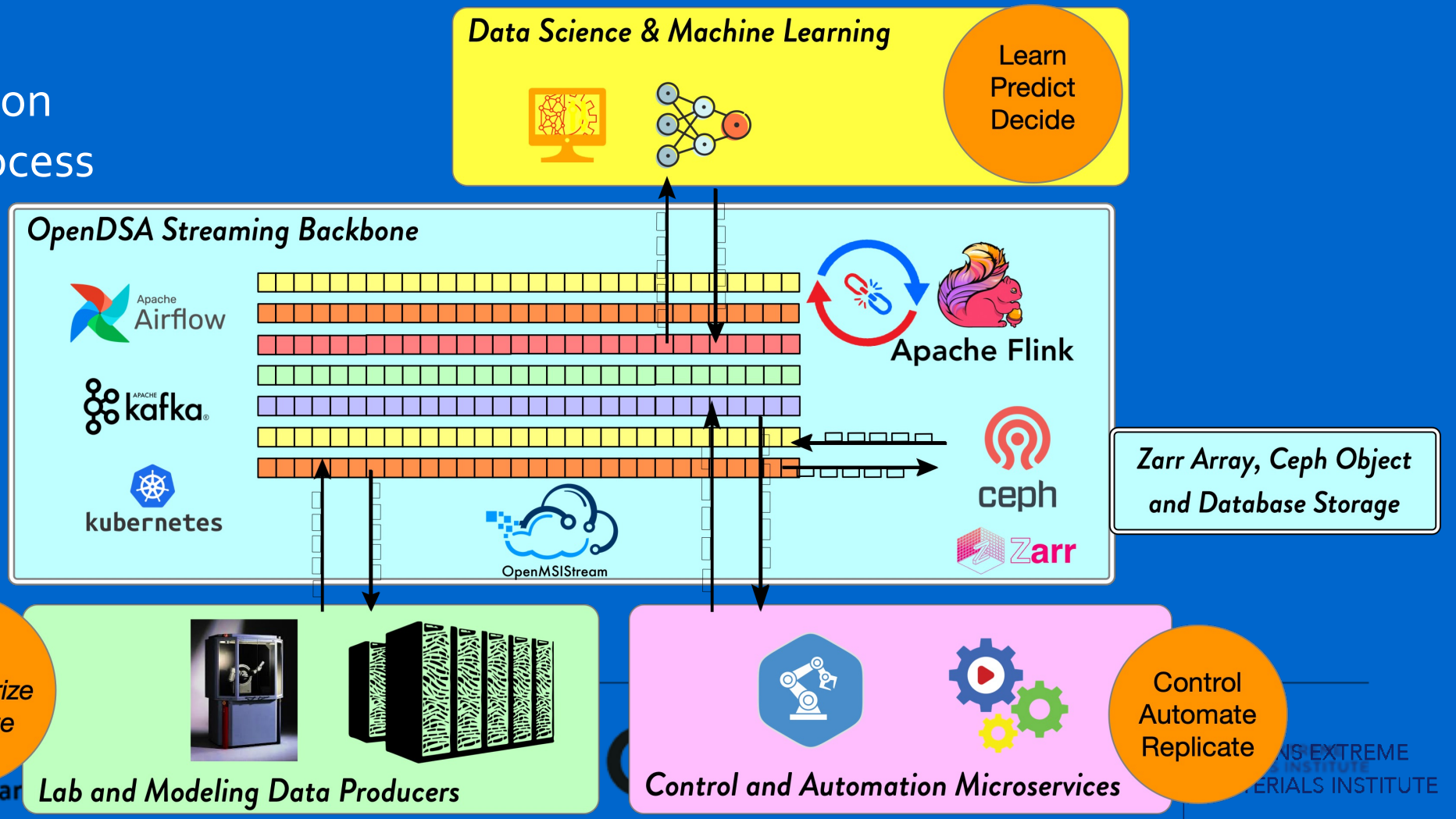
MaRDA | Materials Research Data Alliance

Event Driven Architecture makes Data the Unifying Thread that automates FAIR; accelerates analysis and AI/ML deployment; and empowers novel science and autonomy

Better Science and Production:

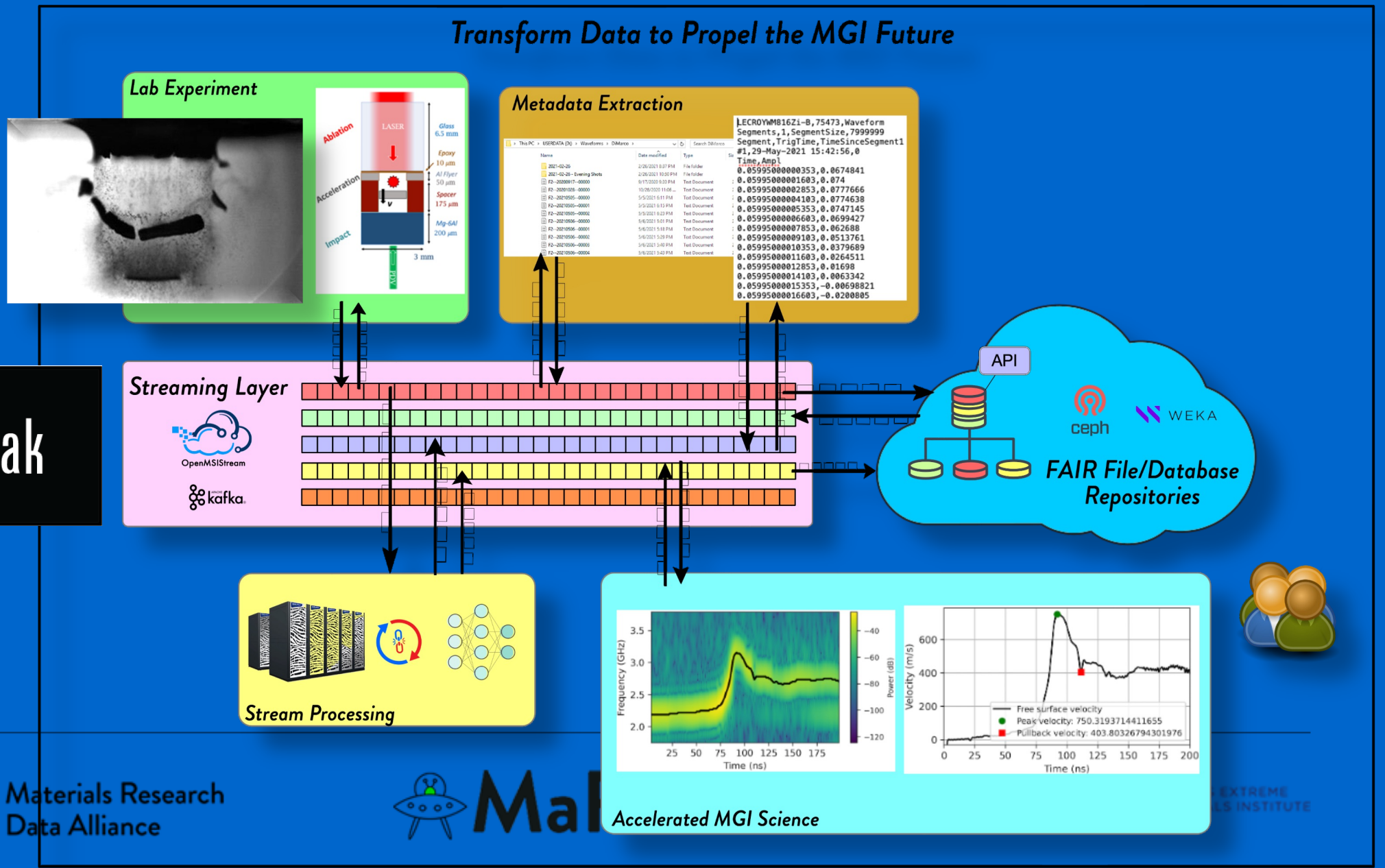
- Faster Discovery
- New Capabilities
- Quicker to production
- Optimization of process

- *High Throughput*
- *Automation*
- *Autonomy*



MaRDA

Example: DMREF Spall Resistant Aluminum OpenMSIStream



Seek the Squeak

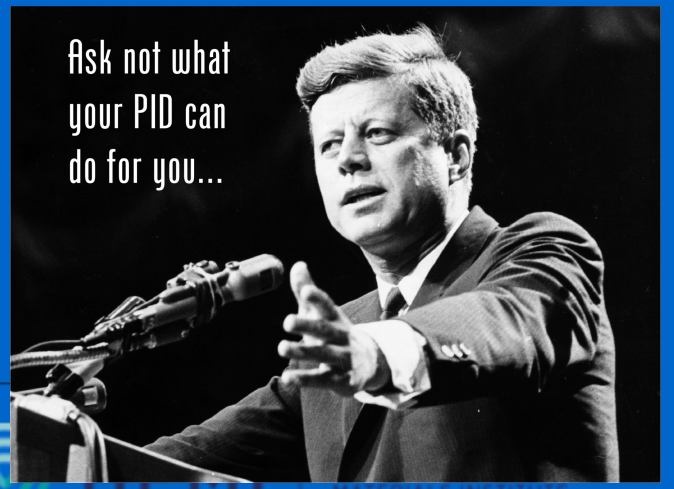
MaRDA | Materials Research Data Alliance

A Possible Path?

- PID Interoperability
  - empowers progress
  - perfect is the enemy of the good
- Outcomes Oriented PID as a Service
- Value Added Tools
- Seek the Squeak
- Lean Into Stakeholder Connections
  - Who controls the first mile?
  - Who controls the last mile?
  - Who gets the impact?
- MaRDA/MaRCN Connect Stakeholders

2023
2025 Do you know
2030 where your data are?
2035...

Got PID?



MaRDA | Materials Research Data Alliance

MaRCN

MATERIALS INSTITUTE