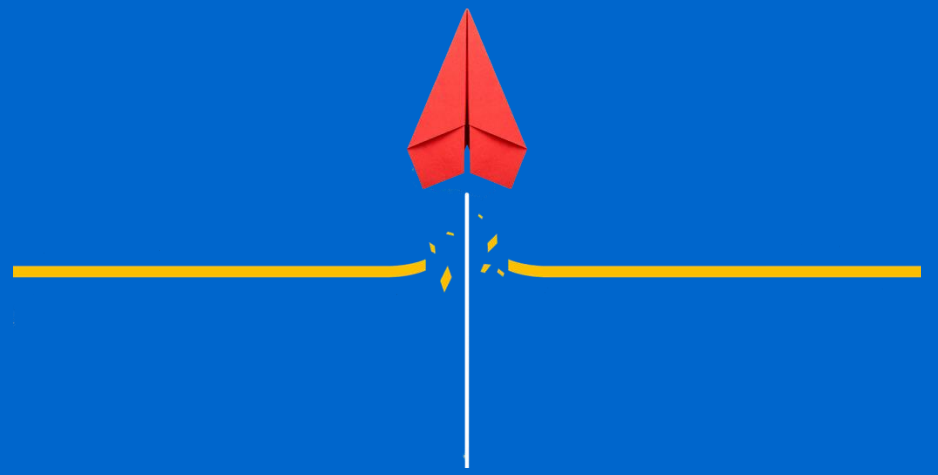
Facility and Instrument PIDs in the Materials Research Data Domain

how do we break the data barrier?



David Elbert: elbert@jhu.edu

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

Illegitimi non carborundum

MaRDA | Materials Research Data Alliance

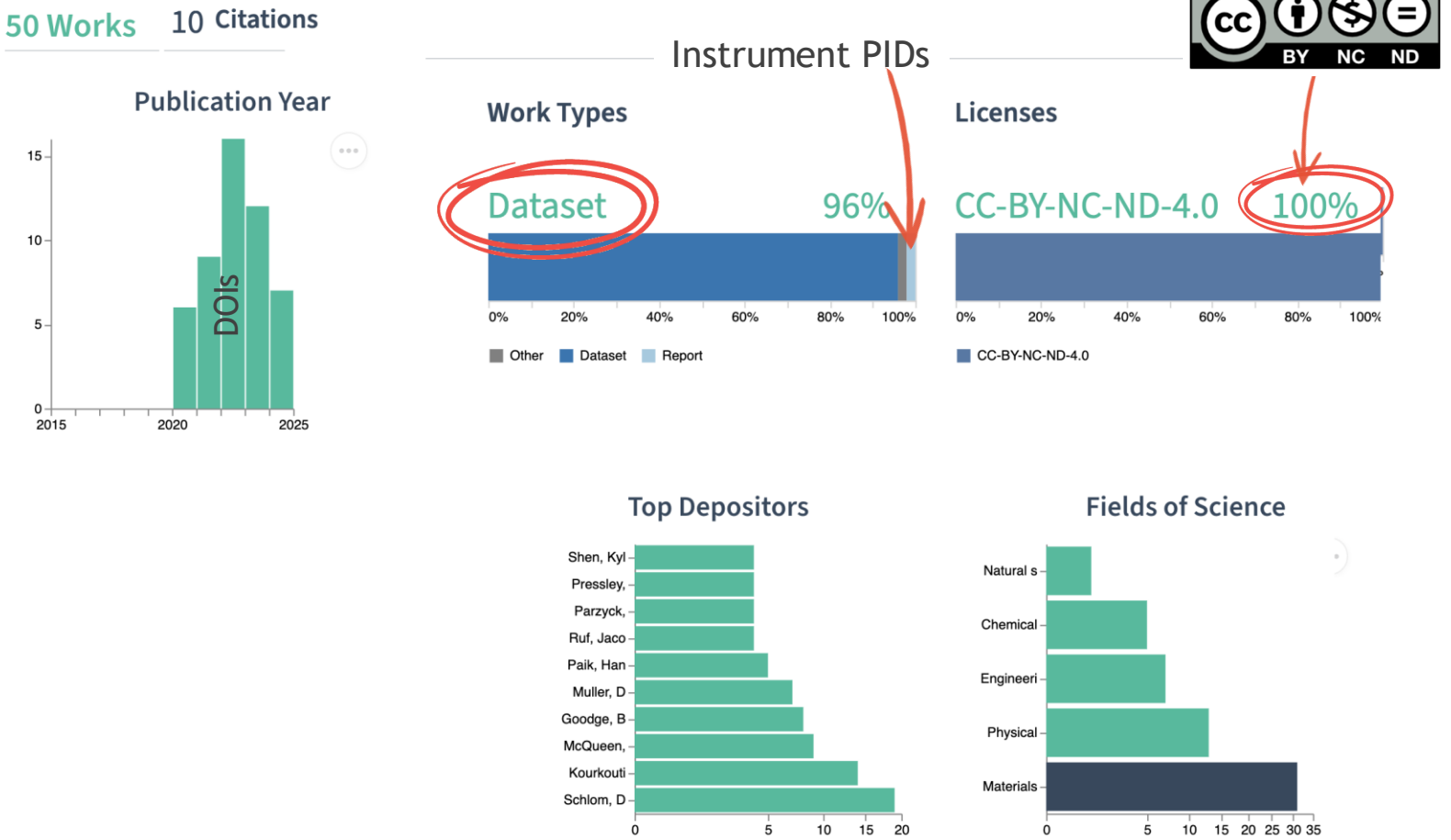
MaRCN

HEMI | HOPKINS EXTREME MATERIALS INSTITUTE

Facility and Instrument PIDs in the Materials Research Data Domain

- Driving Motivations in Materials Science and Engineering
  - PIDs Serve PIDs – How can they serve science and society?
- PID Motivators
  - Value Added Tools
- Challenges:
  - Links Not Labels – Dynamic?
  - Can We Go Faster? – backend value, schema changes?
  - Gaps – what is a gap analysis in a rapidly changing landscape?

Recap: 79 Published Datasets



50 Curated PARADIM Publication Datasets
- **33 PARADIM DOIs**
- **17 Other Publishers**
  - 7 Zenodo
  - 2 Phys Rev Materials
  - 2 Jour App Phys
  - 1 Fighare
  - 1 Materials Project
  - 1 MDF
  - 1 ICSD (CCDC FIZ Karlsruhe)
  - 1 GT Library
  - 1 PDC

25 "Raw" PARADIM Publication Datasets
- **11 PARADIM DOIs**
- **14 Other Publishers or No DOI**

4 Collaboratively Curated Datasets

2 Reports

R: License and Relevant Metadata
- CC-4.0-BY-NC-ND
- GEMD Knowledge Graphs

Our PID Experiment

	A	B	C	D	E	F	G	H	
1	Identifier	Name of Instrument	URL	Date	Owner	Manufacturer	Model	Description	Instrument Typ
2		IT700 SEM		2024	MCP	JEOL	JSM-IT700HR	Its new electron gun with spatial resolution of 1 nr	Scanning Electr
3		F200		2024	MCP	JEOL	JEM-F200	The JEM-F200 is equipped with a Cold Field Emiss	Multi-purpose E
4		GrandARM		2024	MCP	JEOL	JEM-ARM300F2 GRAND ARM™2	It enables observation at ultrahigh spatial resoluti	Transmission El
5		FIBSEM		2024	MCP	ThermoFisher Scientific	Helios 5 UC DualBeam	The Thermo Scientific™ Helios™ 5 UC DualBeam is	Focused Ion Be
6		MicroCT		2024	MCP	RX Solutions	Easytom	EasyTom S is a highly configurable CT system. A co	Computed Tom
7		XRD - Bruker		2024	MCP	Bruker	D8 ADVANCE Plus	It represents the ultimate X-ray platform for multi-	X-Ray Diffractic
8		XRD - Powder		2024	MCP	Malvern Panalytical	Aeris Powder X-ray diffractometer	The Malvern PANalytical Aeris research edition po	powder x-ray di
9		DED		2024	MCP	FormAlloy	L2	FormAlloy's award-winning directed energy depos	Directed Energy
10		PIPS		2024	MCP	Gatan	PIPS II	X,Y stage permits alignment of argon beams to reg	Precision Ion Pc
11		ICP-OES		2024	MCP	Agilent	OES	It is a powerful method for analyzing the concentr	Inductively Cou
12		Bioprinter		2024	MCP	RegenHu	R-GEN 200	The R-GEN 200 bioprinter embodies this innovatic	Bioprinter
13									

- 11 Lab Instruments
- Curator and Lab Scientist
- Spreadsheet of PIDInst Schema
- Notebook to mint

Next Steps:
- Derived webpage
- Derived Facilities Doc
- IGSN

```
    # Construct the data payload for each record if not a
    data_payload = {
        "data": {
            "type": "dois",
            "attributes": {
                "prefix": "10.34863",
                "identifiers": [
                    {
                        "identifier": "1234567",
                        "identifierType": "SerialNumber"
                    }
                ],
                "creators": [
                    {
                        "nameIdentifiers": [
                            {
                                "schemeUri": "https://orc
                                "nameIdentifier": "https:
    }

    # Convert the payload to JSON
    data = json.dumps(data_payload)

    # Send the POST request
    response = requests.post(api_endpoint, auth=(username
    if response.status_code == 201:
        doi = response.json()['data']['id']
        timestamp = datetime.datetime.now().isoformat()
        print(f"DOI {doi} minted for {row['Name of Instru
        # Log success to CSV
        with open('doilog.csv', 'a', newline='') as csvfi
            logwriter = csv.writer(csvfile)
            logwriter.writerow([timestamp, row['Owner'],
    else:
        print(f"Failed to mint DOI for {row['Name of Inst
```

```
Bulk minting DOIs for files in TEST.xlsx
DOI already minted for F200. Skipping...
DOI already minted for GrandARM. Skipping...
DOI 10.34863/rgn4-3j98 minted for FIBSEM
```

README Apache-2.0 license

jhu_pidinst

Pilot for Instrument PIDs in JHU Labs

Getting Started:

1. Upload the data Excel sheet using the format of "PIDInst Excel Template" - Identifier and URL columns should stay empty.
2. Edit datacite-api-config.json
3. Open datacite-api-notebook, change path = 'your_excel_name.xlsx', and run!

Completed:

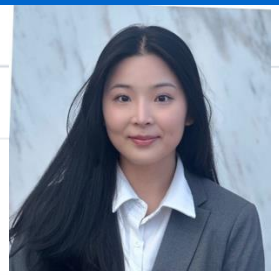
Jupyter Notebook to automate the process of minting DOIs for records in an Excel file, logging each successful operation to a CSV file.

Expecting:

1. Table of instrument metadata for a suite of instruments in PARADIM, MCP, and the Malone SEM (including DOI once minted in DataCite)
2. Code to:
3. i. use DataCite API to access DOI and retrieve JSON of metadata
   ii. Jupyter Notebook to generate Facilities Document mock-up by listing DOIs of instruments
   iii. Jupyter Notebook to generate static html page for instrument from DOI
4. Thoughts/Commentary/Iteration?
5. Once that is working can we envision a way to keep the instrument metadata list here and have a script automatically create DOIs for newly added instruments and modify metadata for changes to existing instruments (maybe with DOI version increment?)
6. What is the best way to trigger the automatic step envisioned? Part of CI/CD or a different runner that watches changes in this repo?

Screenshot

Christine Park
JHU undergrad



Motivators: Grand Challenges – Advanced Materials

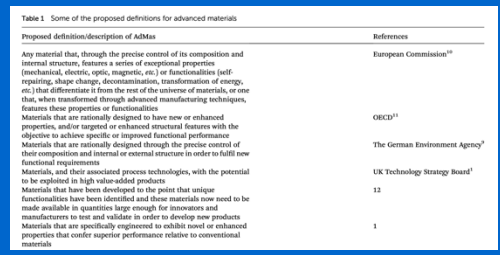
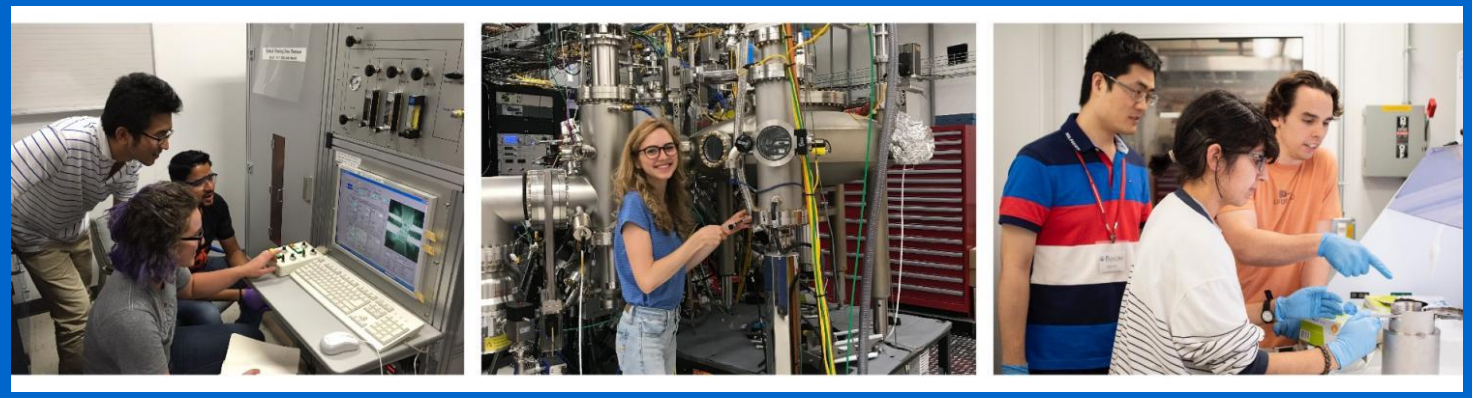


Table 1 Some of the proposed definitions for advanced materials

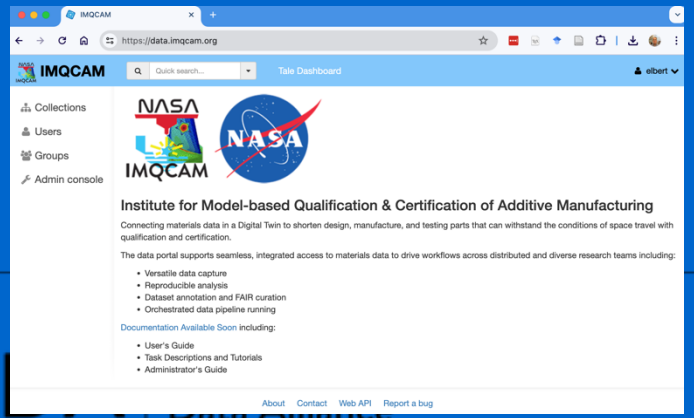
Proposed definition/description of AdMas	References
Any material that, through the precise control of its composition and internal structure, features a series of exceptional properties (mechanical, electric, optic, magnetic, *etc.*) or functionalities (self-repairing, shape change, decontamination, transformation of energy, *etc.*) that differentiate it from the rest of the universe of materials, or one that, when transformed through advanced manufacturing techniques, features these properties or functionalities	European Commission[10]
Materials that are rationally designed to have new or enhanced properties, and/or targeted or enhanced structural features with the objective to achieve specific or improved functional performance	OECD[11]
Materials that are rationally designed through the precise control of their composition and internal or external structure in order to fulfil new functional requirements	The German Environment Agency[9]
Materials, and their associated process technologies, with the potential to be exploited in high value-added products	UK Technology Strategy Board[1]
Materials that have been developed to the point that unique functionalities have been identified and these materials now need to be made available in quantities large enough for innovators and manufacturers to test and validate in order to develop new products	12
Materials that are specifically engineered to exhibit novel or enhanced properties that confer superior performance relative to conventional materials	1

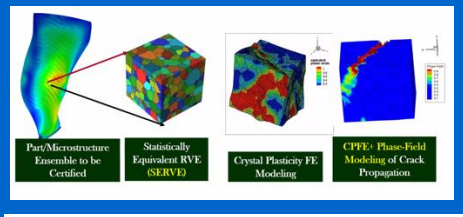
DOI: 10.1039/d2va00128d

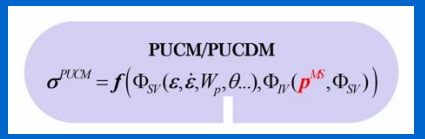
PARADIM Quantum Materials

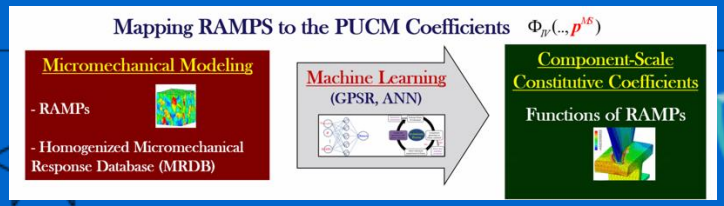


IMQCAM Certified Metal-Additive Parts Manufactured in Space







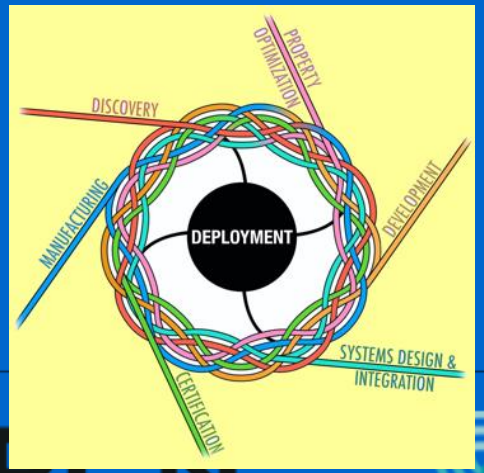


HEMI
HOPKINS EXTREME MATERIALS INSTITUTE

Motivators: Societal Grand Challenges

Materials Genome Initiative (MGI)

- 2011 Interagency Initiative (Obama Administration)
  - Materials Discovery/Deployment Accelerator
  - OSTP Subcommittee Explication and Coordination
  - Funding: NSF, DOE, DOD, NIST (NASA, NIH)
- 2022 MGI Strategic Plan "2.0"
  - Unify the Materials Innovation Infrastructure (MII)
    - computational/experimental/integrated platforms/data infrastructure
    - national materials data network (MaRDA)
    - unify/incentivize through Grand Challenges
  - Harness the Power of Materials Data
    - AI-Ready Data and AI-Driven R&D
  - Educate, Train, Connect R&D Workforce



MaRDA | Materials Research Data Alliance

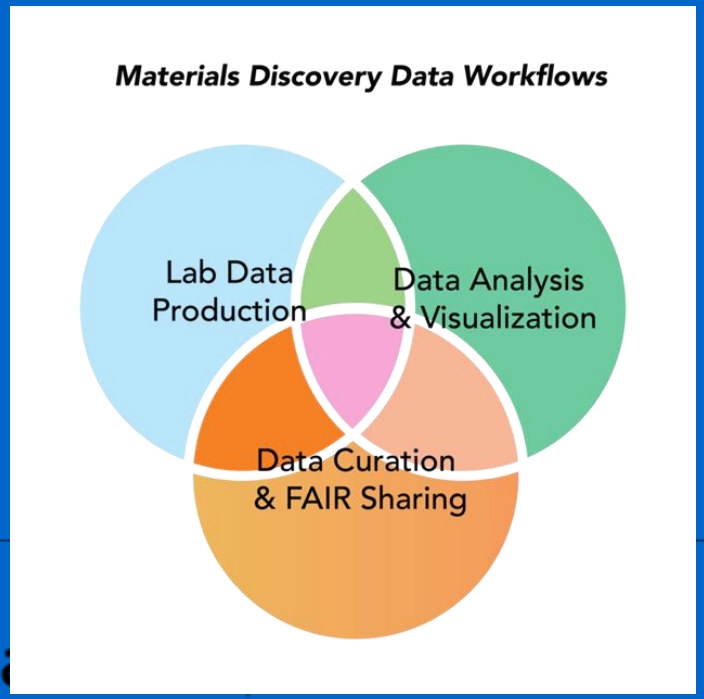
MaRCN

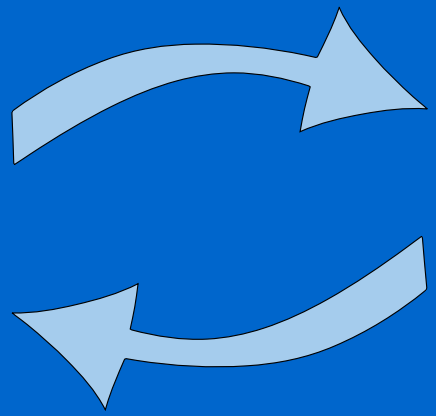
HEMI | HOPKINS EXTREME MATERIALS INSTITUTE

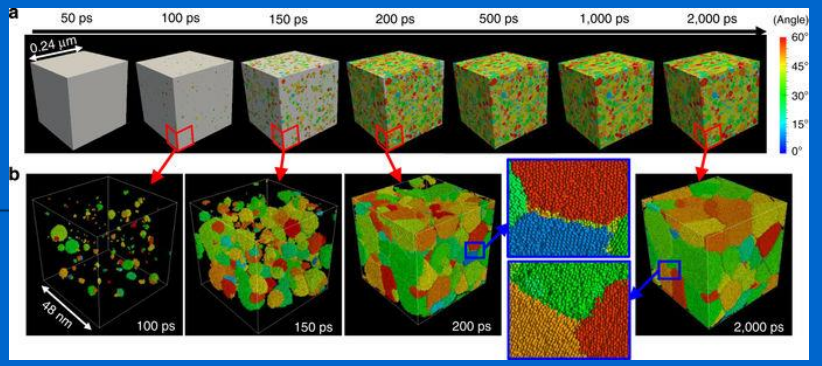
Motivators: Data and Workflows

Fundamental Materials Problems

- Understand how a material works
- Reveal structure-property-synthesis relationships
- Provide path to make designed materials
- Expand knowledge of parameter and design space
- Develop AI/ML to accelerate solutions







Shibuta et al., 2017

Landscape is Changing

- Automation + Decisions = Autonomy

Accelerated Materials Experimentation Enabled by the Autonomous Materials Innovation Infrastructure (AMII) A Workshop Report

Date: June 10th-11th, 2024
Time: 8.00 AM – 5.00 PM
Venue: National Science Foundation (NSF)
Room 2020/2030
2415 Eisenhower Ave
Alexandria, VA 22314

MAPs: Accelerating materials research and development to meet urgent societal challenges

Simon P. Stier[1]*, Christoph Kreisbeck[2], Holger Ihssen[3], Matthias Albert Popp[1], Jens Hauch[4], Kourosh Malek[5], Marine Reynaud[6], Fedor Goumans[7], Johan Carlsson[8], Ilian Todorov[9], Lukas Gold[1], Andreas Räder[1], Wolfgang Wenzel[10], Shahbaz Tareq Bandesha[1], Philippe Jacques[11], Francisco Garcia-Moreno[12], Oier Arcelus[6], Pascal Friederich[10], Simon Clark[13], Mario Maglione[14], Anssi Laukkanen[15], Ivano Eligio Castelli[16], Montserrat Casas Cabanas[6], Javier Carrasco[6], Helge Sören Stein[17], Ozlem Ozcan[18], David Elbert[19], Tejs Vegge[16], Sawako Nakamae[20], Monica Fabrizio[21], Mark Kozdras[22]

Affiliations:

[1]Department Digital Transformation, Fraunhofer Institute for Silicate Research ISC; Neunerplatz 2, 97082 Würzburg, Germany

[2]Aixelo Inc.; Cambridge, MA 02141, US

[3]Helmholtz Association; Rue du Trône 98, 1050 Bruxelles, Belgien

[4]Forschungszentrum Jülich GmbH, Helmholtz-Institut Erlangen-Nürnberg for Renewable Energy (HI ERN), Institute of Materials for Electronics and Energy Technology (i-MEET); 91058 Erlangen, Germany

[5]Forschungszentrum Jülich GmbH, Theory and Computation of Energy Materials (IEK-13), Institute of Energy and Climate Research (IEK); 52428 Jülich, Germany

[6]Centro de Investigación Cooperativa de Energías Alternativas (CIC energiGUNE), Basque Research and Technology Alliance (BRTA); Parque Tecnológico de Álava, Albert Einstein 48, 01510 Vitoria-Gasteiz, Spain

[7]Software for Chemistry & Materials BV; De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

[8]Dassault Systemes Deutschland GmbH; Cologne, Germany

[9]Scientific Computing Department, Science and Technology Facilities Council, Daresbury Laboratory; Warrington, UK

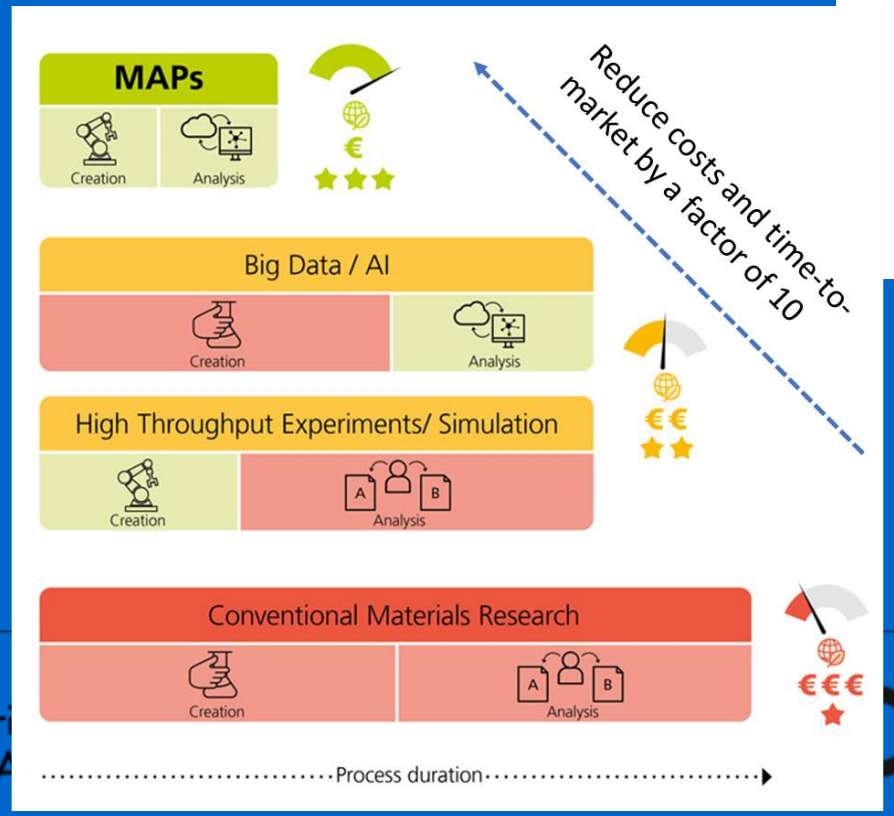
[10]Institute of Nanotechnology (INT), Karlsruhe Institute of Technology; Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[11]EMIRI AISBL; Rue de Ransbeek 310, B-1120 Brussels, Belgium

[12]Institute of Applied Materials, Helmholtz-Zentrum Berlin für Materialien und Energie; Hahn-Meitner-Platz 1, Berlin 14109, Germany

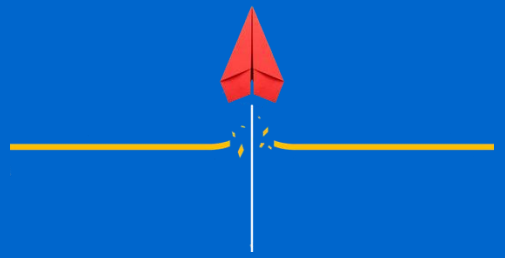
[13]SINTEF Industry, New Energy Solutions; Sem Sælands vei 12, Trondheim, 7034 Norway

[14]Institut de [illegible] Matière Condensée de Bordeaux (ICMCB)-UMR 5026, CNRS, Université [illegible] Avenue du Docteur Schweitzer, F-33608 Pessac, France

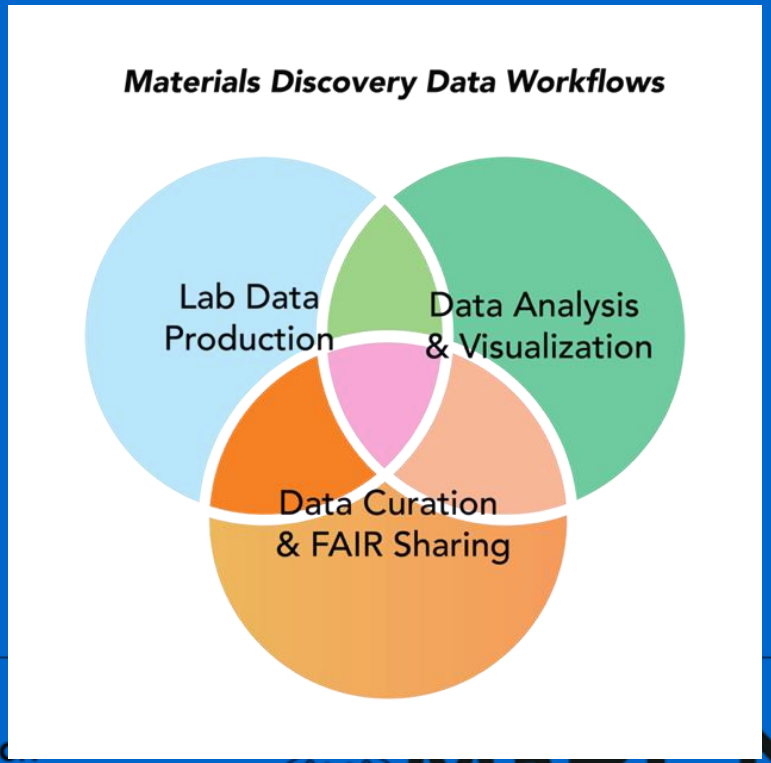


HEMI
HOPKINS EXTREME MATERIALS INSTITUTE

Motivators: Breaking the Data Barrier

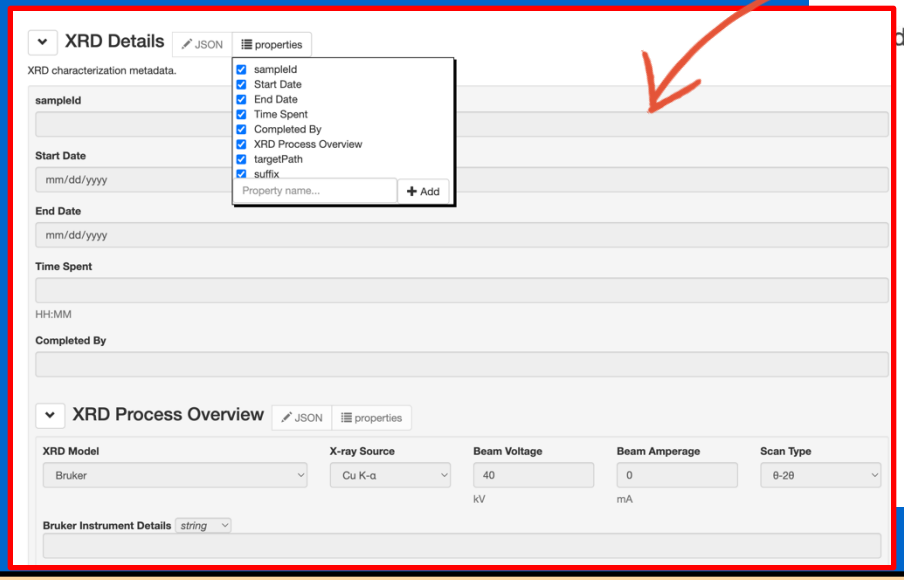


How can we aggregate data in meaningful ways across complex scientific workflows?

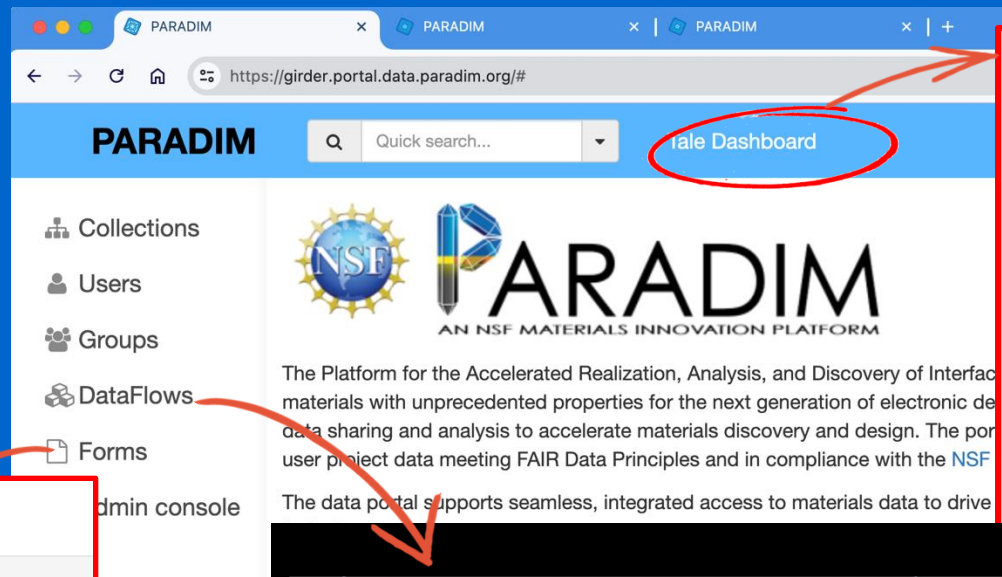


Automation in Data Portals

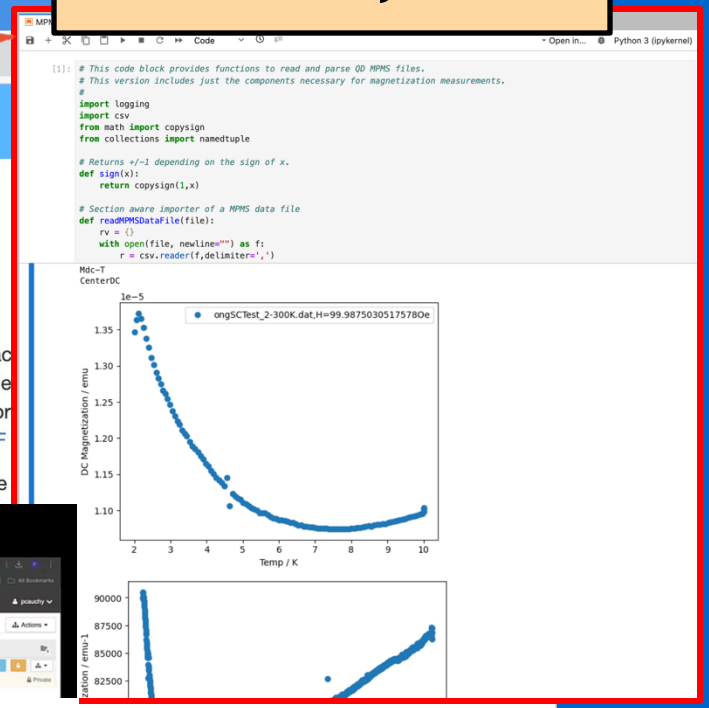
- JSON Forms
- One-Click Tales
- Streaming Ingest
- Chameleon Automation

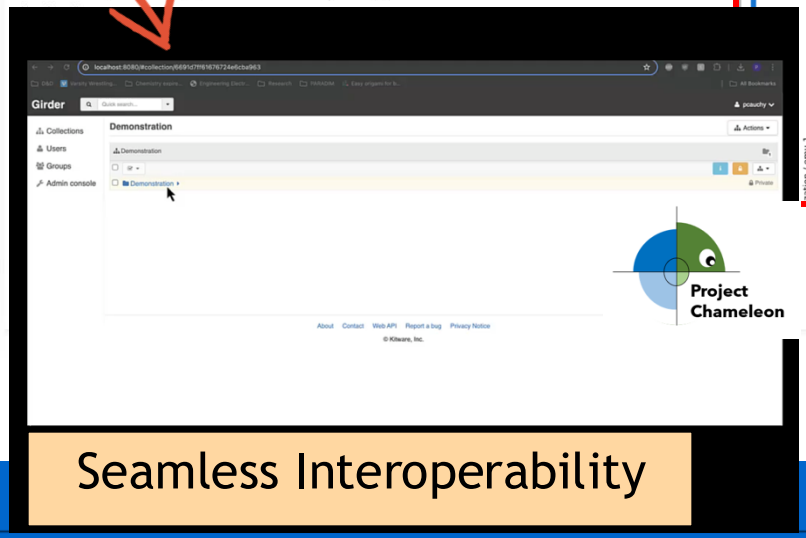


Schema Validated Metadata Entry



Click-to-Analyze Tales





Seamless Interoperability

MaRDA Materials Research Data Alliance

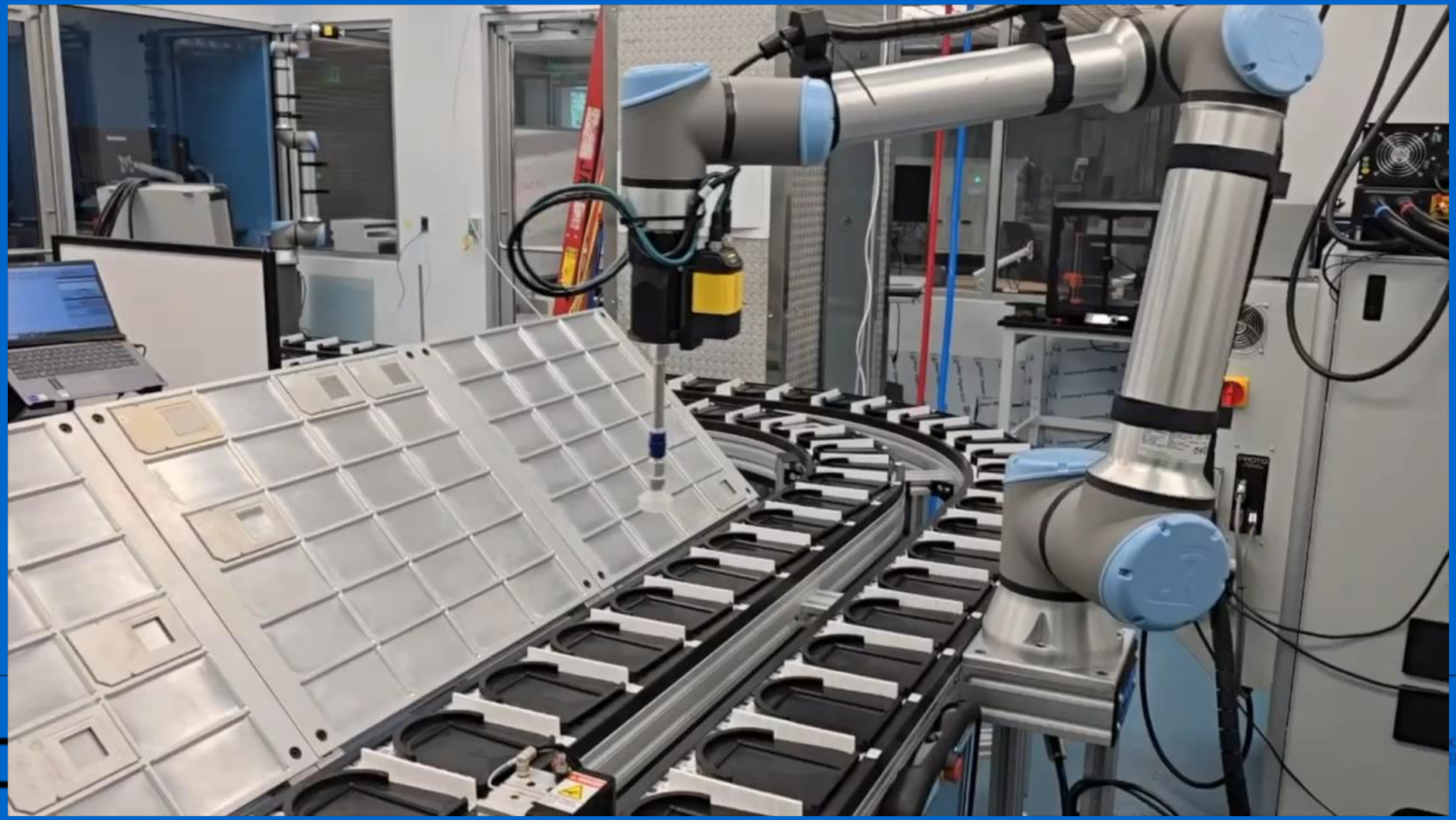
NIST

NSF

MaRCN

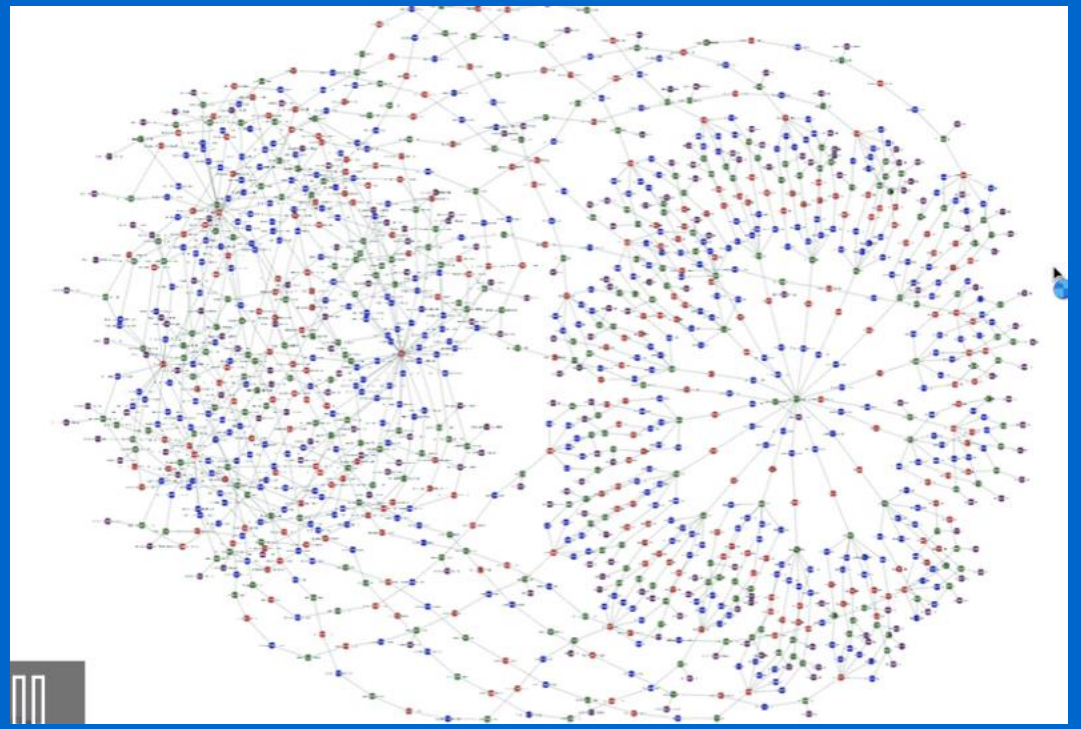
Automation in Labs

- High Throughput
- Autonomy



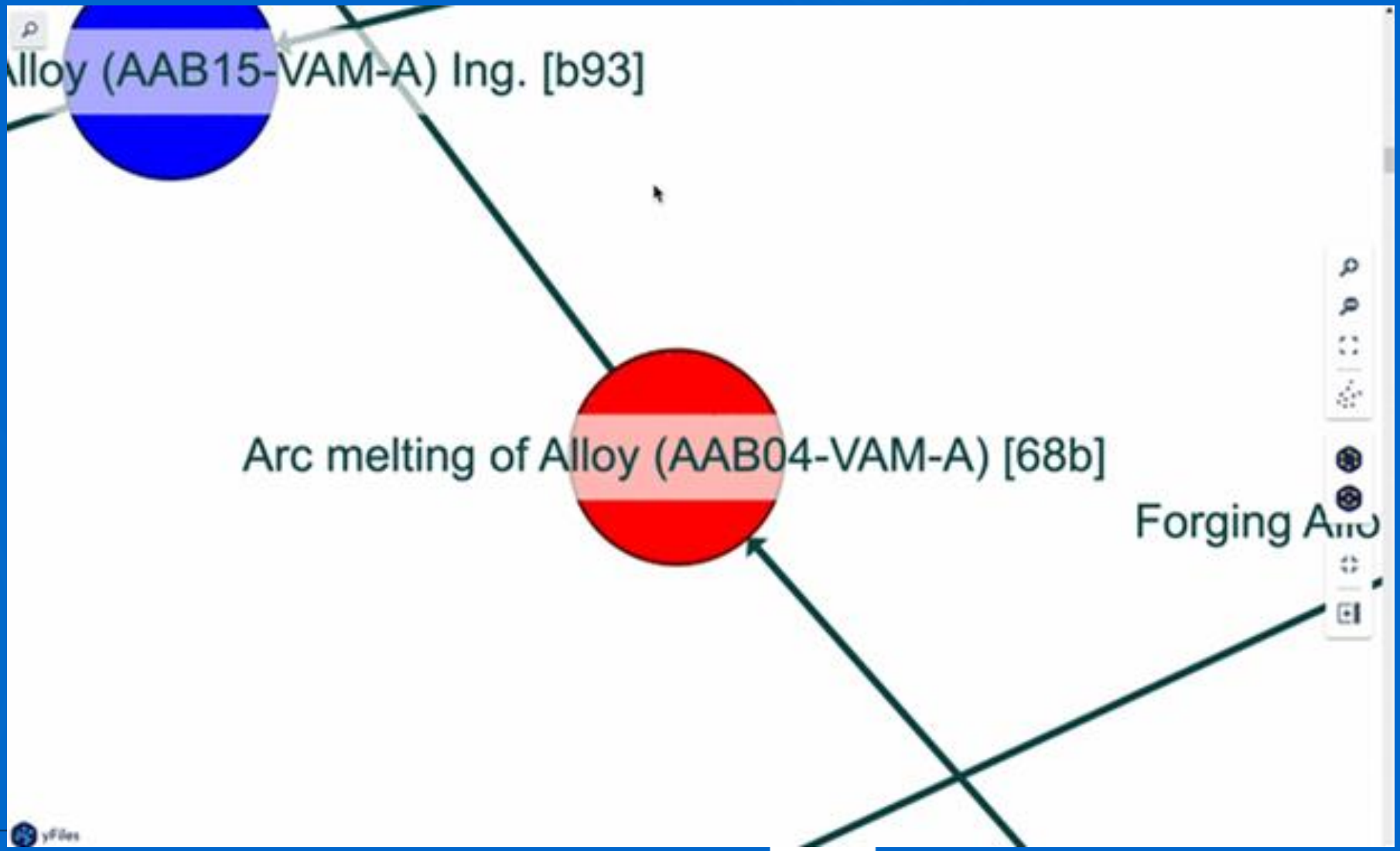
Automation in Knowledge Graphs

Concepts → HTMDEC Data



Automation in Knowledge Graphs

Birdshot:
Arroyave and Karaman



Legend	
Object Type	Node Color
Process	
Material	
Ingredient	
Measurement	

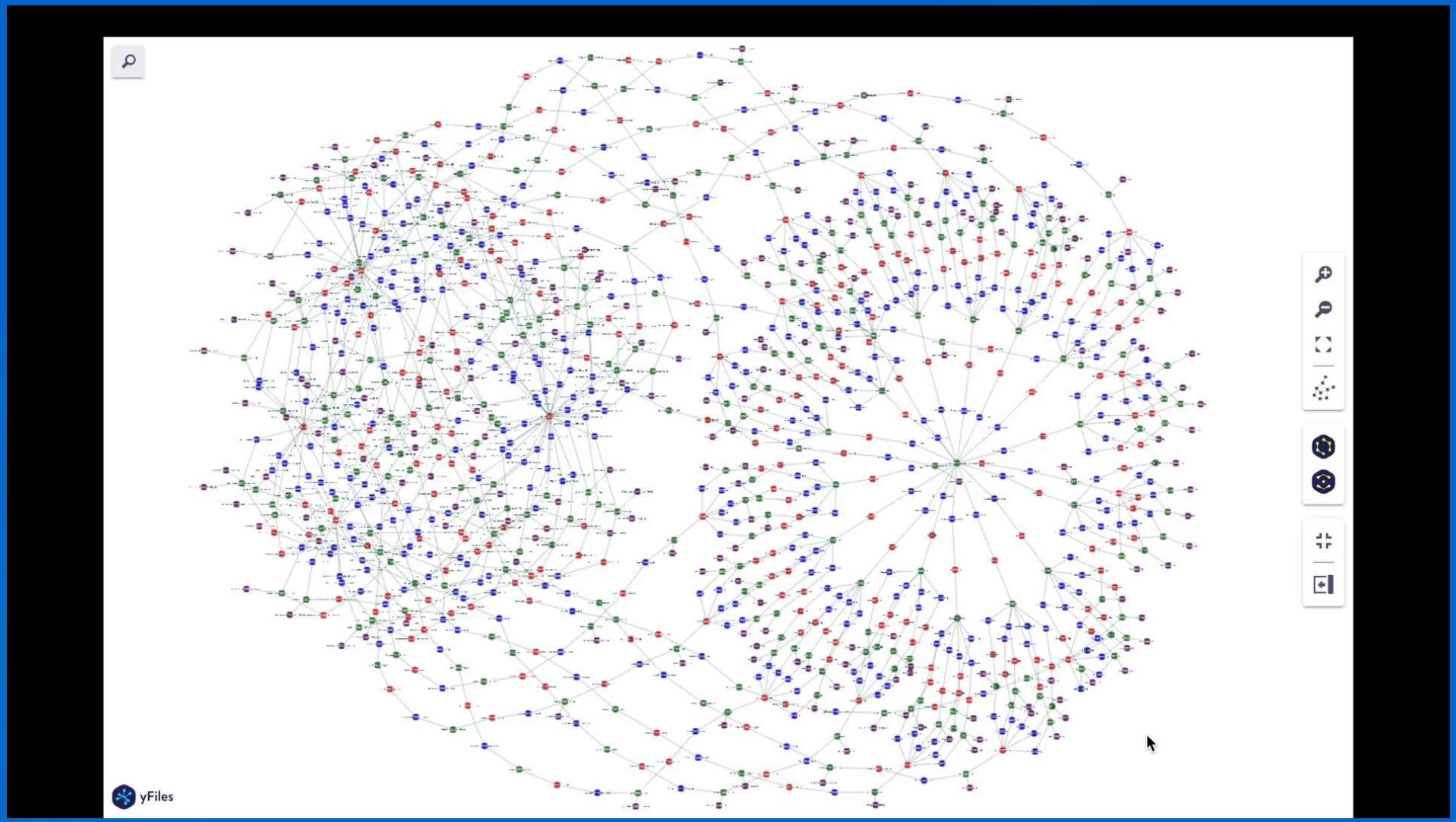
NSF PARADIM AN NSF MATERIALS INNOVATION PLATFORM

MaRDA Materials Research Data Alliance

NASA IMQCAM

HEMI HOPKINS EXTREME MATERIALS INSTITUTE

Automation in Knowledge Graphs



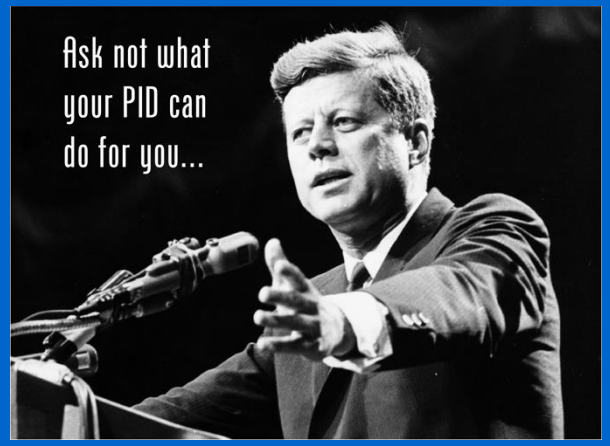
Legend	
Object Type	Node Color
Process	
Material	
Ingredient	
Measurement	

NSF PARADIM AN NSF MATERIALS INNOVATION PLATFORM MaRDA Materials Research Data Alliance NASA IMQCAM HEMI HOPKINS EXTREME MATERIALS INSTITUTE

PIDS Are Critical to any Roadmap to Foundational Knowledge Graphs

What We Have (OpenMSI/HTMDEC/VariMat/PARADM)

- GEMD Model
- Automated Processing
- Workflow Building Blocks



What We Need

- Simplify the Simple Parts (PIDs for samples and out-of-scope parts)
- Barrier Removal (portal and infrastructure at low/no cost)
- Scale out