Graphical Materials Histories – Making the Invisible Visible

David Elbert: elbert@jhu.edu

CDO PARADIM Materials Innovation Platform (MIP) (2DLIM/Quantum Materials)
ARL HTMDEC Extreme Data (ceramics and alloys)
NASA IMQCAM STRI (additive metals digital twin)
NSF DMREF (streaming data alloys and recycled polymers)
DOE-BES Catalysis (zeolite scaling)
MaRDA (Materials Research Data Alliance)
MaRCN FAIROS-RCN

Scope: Data Production and Use

Scope: Community – People are Infrastructure, Too

Co-Conspirators!
Maggie Eminizer (JHU)
Ali Rachidi (JHU)
Matt Turk (UIUC)
Kacper Kowalik (NCSA/UIUC)
Craig Willis (UIUC)
Challenges

• Materials Data is Vast and Diverse
  
  Materials Research is Widely Distributed
  
  • Place/Time/Kind
  
Workflows are Numerous/Diverse/Changing

• AI/ML is Powerful yet Nascent
• Stakeholder/Expertise Fragmented
• Last-Mile Problem

First-Mile Problem – Model Instantiation Tools and Automation

• OpenMSI Streaming and Modeling
• Graphical Process Editors
• PID samples/instruments/anything linked or shared – Simplify the Simple

• Intellectual Interoperability
  
  • Incentivized to have solved the problem
  • “Do it my way”

Scientists are the worst judges of the reuse of their own data
Fundamental Materials Problems – How do you Discover a Novel Material?

- Understand how a material works
- Reveal structure-property-synthesis relationships
- Provide path to make designed materials
  - If you can’t make it, it doesn’t exist
- Expand knowledge of parameter and design space

Can AI/ML Address Fundamental Problems?
Material Histories Are Fundamental

Data alone is not enough
- History is directed
- Graphs underlie research
- Graphs underlie production

Histories are Connected Data

MGI Strategic Plan https://mgi.gov
Autonomy Revolution: *A Secret Superpower*

- Automating Everything
- High Throughput
- Automation + Decisions = Autonomy
- Data Infrastructure
  - Enables automation
  - Deploys decisions

---

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

Milad Abolhasani & Eugenia Kumacheva

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

14k Accesses | 9 Citations | 23 Altmetric | [Metrics]

---

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]

---

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]
Workflows Connect to Workflows

Data aggregation across complex, changing scientific workflows
Workflows Unlock Everything

• Three Types:
  • Lab Data Production
  • Data Analysis and Visualization
  • Data Curation/Sharing

• Explicitly Describe Workflows
  • Prescribed or Captured
  • Directed Acyclic Graphs
  • Transparent/Repeatable/Connectable
  • Automated/Runnable
Material History Data Model (GEMD https://citrineinformatics.github.io/gemd-docs)

Work the way materials scientists work!

Gannon Murray, rising junior
Earlham College
https://github.com/paradimdata/YVO4_project
Coding the History

Details that never make it to the literature

Crucible Growth of YVO₄ from V₂O₅ and Y₂O₃ in Air

YVO₄ was grown in an alumina crucible from Y and V oxides using a furnace.

The purpose of the experiment was to test the effects of vanadium deficiency on phase formation.

Performed 2023-06-09 @ ML by GBM

```python
# Imports
from tools atte_utils import *
from tools block block import Block
from tools utilities import *
from utils block builders import *
from utils base builders import *
from utils provenance import Provenance
from goods import FileLink
from goods json import JSON
import os

# Provenance
prv = Provenance()
prv = 'Gannon Murray',
prv = 'ggunn43@jhu.edu',
prv = 'GBM',
prv = 'YVO4',
prv = 'Crucible Growth of YVO4',
prv = '2023-06-09'
```

https://github.com/paradimdata/YVO4_project
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': 900.0, 'Type': 'Ramp'}, 'type': 'nominal_composition'}: "False"
Step 3, {'quantities': {'Duration': 1.0, 'Temp': 900.0, 'Type': 'Hold'}, 'type': 'nominal_composition'}: "False"
Step 4, {'quantities': {'Duration': 9.0, 'Temp': 0.0, 'Type': 'Ramp'}, 'type': 'nominal_composition'}: "False"
Step 5, {'quantities': {'Duration': 0.0, 'Temp': 0.0, 'Type': 'End'}, 'type': 'nominal_composition'}: "False"

label: "GBM1013_Flux_Medium Heating Process, 5e1"
Scientists Live in a Forest They’ve Never Seen
What if you could see the real structure of your research?

Concepts ➔ HTMDEC Data
HTMDEC Centers

Technical Thrusts:

• Data Driven Materials Design
• High-Throughput Synthesis and Processing
• High-Throughput Characterization
• ML-Augmented Physics Based Models

Scientists are the worst judges of the reuse of their own data
HTMDEC Different Philosophies, Same Goal: High Throughput

- Manual
- Structured

- Robotic Automation
- Variable
- Event Driven
What if you could see the structure of your research?

Birdshot:
Visualize Your Materials Histories

Make the Invisible Visible
Intuitive Data Provenance
Challenges

- Materials Data is Vast and Diverse
- Materials Research is Widely Distributed
  - Place/Time/Kind
- Workflows are Numerous/Diverse/Changing
- AI/ML is Powerful yet Nascent
- Stakeholder/Expertise Fragmented
- Last-Mile Problem
- **First-Mile Problem – Model Instantiation Tools and Automation**
  - OpenMSI Streaming and Modeling
  - Graphical Process Editors
  - PID samples/instruments/anything linked or shared – Simplify the Simple
- Intellectual Interoperability
  - Incentivized to have solved the problem
  - “Do it my way”
- **Boromir Syndrome – Don’t vilify the heroes**