Building an End-to-end Data Ecosystem to Support Materials Science Research

Ben Blaiszik (blaiszik@uchicago.edu), Logan Ward, Jonathon Gaff, Ian Foster

Michael Ondrejcek,
Ben Galewsky, Kenton McHenry

Rachana Ananthakrishnan, Steven Tuecke, Rick Wagner, Nick Saint, Eric Blau, Kyle Chard, Yadu Nand Babuji, John Towns,
Mike Papka
INTEGRATIVE MATERIALS DESIGN (IMAD)

Connecting the Midwest Materials Data Ecosystem

Integrative Materials and Design

Connect academic and industrial researchers, data services, and tooling

Provide simplified and unified access to high value materials datasets

Inform the community and public about the materials informatics work being done in the Midwest through videos, news articles, webinars, tutorials, workshops, etc.
Materials Data Facility
Build data services to

- **Empower researchers to publish data**, regardless of size, type, and location
- **Automate data and metadata ingest**, to enable capture of many valuable materials datasets
- **Enable unified search and discovery** across disparate materials data sources

Deploy with APIs to simplify connection to other data efforts and to enable automation
MDF Connect - Connecting Community Services

- Make it easy to deposit into many services from one location
- Strictly opt-in for cross-posting datasets

Submit Data [UI or API]

- .zip | link to archive | Globus endpoint path

Enrich Data

- Extract / Transform
- MDF Connect

Send to Community

- Mint DOIs
- Associate metadata
- Persist datasets

- Query
- Browse
- Aggregate

Forge

NIST MRR

Google Drive

MaterialsCommons

CITRINE INFORMATICS

4CeeD
MDF Connect Prototype

MDF CONNECT
It has never been easier to share your data with the community. Deposit data once, send to partner services.

Tell your research story.

Become a Contributor

HOW TO GET STARTED
Collect, publish, categorize models from many disciplines

Serve models via API to foster sharing, consumption, and access to data, training sets, and models

Simplify and automate training of models (using HPC and cloud)

Enable new science through reuse and synthesis of existing models

Funding: 2018 Argonne Adv. Computing LDRD
Predicting Glass-forming Ability

Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments

Fang Ren toolbox, Logan Ward, Travis Williams, Kevin J. Law, Christopher Wolverton, Jason Hattrick-Simpers and Apurva Mehta

10.1126/sciadv.aaq1566

- Where are the model and trained weights?
- How do I run the model on my data?
- Should I run the model on my data?
- How can I retrain the model on new data?
- How can I build on this work?

- How do I share my model with the community?

Funding: 2018 Argonne Adv. Computing LDRD
Predicting Glass-forming Ability

Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments

Fang Ren, Logan Ward, Travis Williams, Kevin J. Laws, Christopher Wolverton, Jason Hattrick-Simpers, and Apurva Mehta

10.1126/sciadv.aaq1566

Funding: 2018 Argonne Adv. Computing LDRD

```python
servable_name = "metallic_glass"
servable_id = dl.get_id_by_name(servable_name)
elems = ["V","Co","Zr"]
res = dl.run(servable_id, {"data":elems})
```

Predicted glass-forming ability

DLHub

[“Zr”, “Co”, “V”]
• Argonne Leadership Computing Facility: The Argonne Leadership Computing Facility is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357.

• Argonne Data Service: The Argonne Leadership Computing Facility is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357.

• Materials Data Facility: This work was performed under financial assistance award 70NANB14H012 from U.S. Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Material Design (CHiMaD).

• DLHub: Add LDRD support line...

• IMaD: This work was also supported by the National Science Foundation as part of the Midwest Big Data Hub under NSF Award Number: 1636950 "BD Spokes: SPOKE: MIDWEST: Collaborative: Integrative Materials Design (IMaD): Leverage, Innovate, and Disseminate".

• Parsl: This work was supported in part by NSF award ACI-1550588 and DOE contract DE-AC02-06CH11357.

• Petrel: The Argonne Leadership Computing Facility is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357.

• Globus: This research was supported in part by NSF grant ACI-1148484 (SciDaaS) and US Department of Energy contract DE-AC02-06CH11357.