
Aristana Scourtas (aristana@uchicago.edu), Ben Blaiszik (blaiszik@uchicago.edu), Ryan Chard, ZhuoZhao Li, Logan Ward, Tyler Skluzaček, Ben Galewska, KJ Schmidt, Nathan Pruyne, Jim James, Isaac Darling, Braeden Cullen, Aadit Ambadkar, Ethan Truelove, Marcus Schwarting, Eliu Huerta, Rebecca Willett, Rafa Gomez-Bombarelli, Dane Morgan, Ian Foster (foster@anl.gov)

https://www.dlhub.org
The Growing Importance of ML and Data in the Sciences

Data and ML are becoming key drivers of scientific progress

Methods and Data: https://github.com/blaiszik/ml_publication_charts
How do we use these models?

For a given study:

• Where is the code?
• Where are the trained models?
• Where are the training data?
• How can I reproduce these results?

Without all of these pieces, progress is drastically slowed

Need models and data to be FAIR:

Findable
Accessible
Interoperable
Reusable

Location of many ML models after a paper is finished
Just to clarify, I think your model issues are very valid and I was in no way blaming you for the issues with the publication -- I felt like the way Dane responded made it kind of seem like that? But I very much do not feel that way lol.

I 1000% agree that it's hard for publishers to figure out what's going on when a model hangs, and that's on us as developers.
we then register the container and the function `dlhub_run()` with funcX

dlhub_run(event)

```python
from home_run import create_servable

with open("dlhub.json") as fp:
    shim = create_servable(json.load(fp))
```

dlhub.json contains all servable-specific info
Garden for FAIR Models and Data

An ecosystem to connect researchers, provide resources, and validate models for the community

Molecular Solubility Prediction Model Garden

1. **Publish**
   - Molecular Solubility Predictor Model
   - Linked to training data
   - Tests specified
   - DOI minted
   - Model page created
   - Metric tracking enabled

2. **Test / Validate**
   - Create containers
   - Run testing
   - Molecular Solubility Predictor Model
   - All tests passing

3. **Catalog**
   - Catalog metadata
   - Enable discovery

4. **Run**
   - Inference tasks
   - Web & Notebook Interfaces
   - Hugging Face

**Publisher**
- Publish models and functions. Receive DOI for citation and landing page
- Track usage metrics and obtain credit.
- Share models

**Consumer**
- Discover tested and validated models
- Explore model reliability, UQ, and testing information
- Run models

Globus Auth

PyTorch, GitHub, Keras, TensorFlow, colab, jupyter

- Run models
Our goal is to make it easy for the user to assess applicability, compatibility, and reliability.

Then, the user can take action:
- Cite/share
- Discover
- Understand
- Decide
We’re hiring 🌱

Interested in building Garden with us, or know someone who is? We’re currently hiring for software developers to help us deliver FAIR infrastructure and models to the community.

Reach out if you’re interested in learning more!

Contact: Aristana Scourtas (aristana@uchicago.edu)
KJ Schmidt (kjschmidt@uchicago.edu)
Ben Blaiszik (blaiszik@uchicago.edu)
This work was supported in part by Laboratory Directed Research and Development funding from Argonne National Laboratory under U.S. Department of Energy under Contract DE-AC02-06CH11357. We also thank the Argonne Leadership Computing Facility for access to the PetrelKube Kubernetes cluster and Amazon Web Services for providing research credits to enable rapid service prototyping. This research used resources of the Argonne Leadership Computing Facility, a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

https://www.dlhub.org
Funding: 2018 Argonne Adv. Computing LDRD

https://www.foundry-ml.org
This work was supported by the National Science Foundation under NSF Award Number: 1931306 “Collaborative Research: Framework: Machine Learning Materials Innovation Infrastructure”.

https://www.materialsdatafacility.org
This work was performed under financial assistance award 70NANB14H012 and 70NANB19H005 from U.S. Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Material Design (CHiMaD).

Contact: Aristana Scourtas (aristana@uchicago.edu)
KJ Schmidt (kjschmidt@uchicago.edu)
Ben Blaiszik (blaiszik@uchicago.edu)