



# Engaging with MSD-LIVE

**Casey Burleyson, Carina  
Lansing, Zoë Guillen, Matt  
Jugovic, Matthew Macduff,  
Devin McAllester, and Jon  
Weers**



PNNL is operated by Battelle for the U.S. Department of Energy



# Open Science in the MSD CoP

*“This report summarizes current MSD research, synthesizes insights from these early investments, defines key concepts, identifies areas for further research, and **proposes a collaborative research strategy based on open science principles and growing a diverse workforce to accelerate progress.**”*

*“**Expand MSD’s commitment and capacity for open science to continue developing interoperable and more easily reusable data, models, and analysis methods;** leverage emerging computational innovations (e.g., AI, natural language processing, reinforcement learning (RL), and interactive visual analytics); and grow and diversify the MSD workforce to broaden the backgrounds, technical skills, and expertise/experiences available to advance our understanding of societal risks.”*



Reed, P. M., et al. (2022). MultiSector Dynamics: Scientific Challenges and a Research Vision for 2030, A Community of Practice Supported by the United States Department of Energy's Office of Science. Zenodo. <https://doi.org/10.5281/zenodo.6144309>



# The Vision for MSD-LIVE



About News & Events Policies Metrics Acknowledgement [Log In](#) [Sign Up](#)  
[Forgot Password?](#)

**Living  
Intuitive  
Value-adding  
Environment**

*A collaborative data and computational platform for the MultiSector Dynamics community*

<https://msdlive.org>



## Data & Code Repository

Discover and share curated MSD datasets, codes, and workflows.



## Computational Resources

Use Jupyter Notebooks to analyze or visualize data stored in MSD-LIVE.



## Project Services

Create and manage project teams and their collaborative tools and resources.



## Need Help?

Learn how to start using MSD-LIVE to manage your data.

- A cloud-based data management system and advanced computing platform
- Will enable researchers to document and archive their data, run their models and analysis tools, and share data, software, and multi-model workflows
- A cornerstone capability of the MSD Community of Practice



# Top-10 Most Pressing Use Cases

Core Capabilities	Use Cases
Data Repository	Find Data
	Archive Data
	Version Control
	Training
	Control Access
Advanced Computing	Share Data
	Analyze Data
	Multi-Model Workflows
	ML/AI
	Move Code to the Data

 = available now

 = current focus

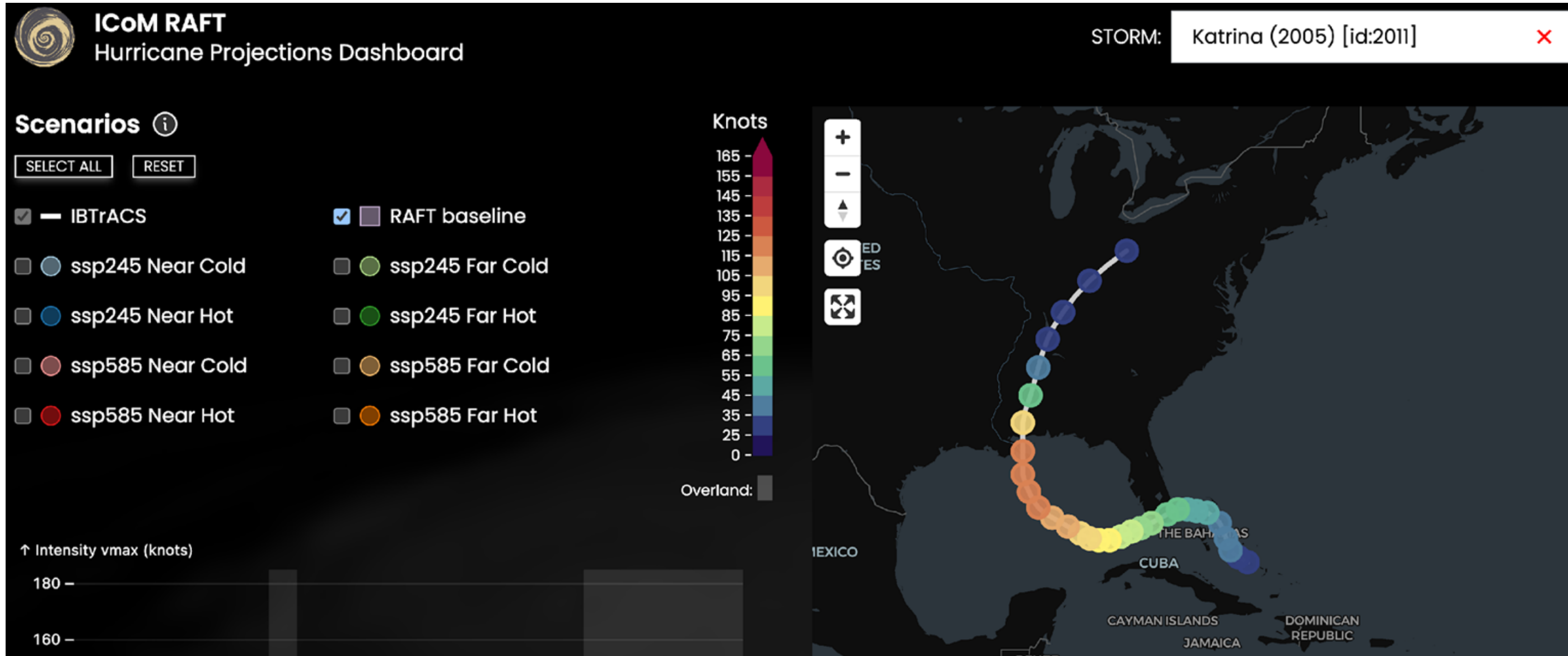


Our use cases are derived from extensive conversations with researchers from the MSD community.

# Current Capabilities

- Data repository:
  - <https://data.msdlive.org/>
- Data dashboards:
  - PCHES - <https://lafferty-sriver-2023-downscaling-uncertainty.msdlive.org/user-redirect/lab/tree/dashboard.ipynb>
  - ICoM - <https://raft-hurricane-projections.msdlive.org/>
- Interactive model training notebooks:
  - IM3 - Uncertainty Characterization eBook: <https://uc-ebook.org/>
  - GCIMS - Xanthos: <https://xanthos.msdlive.org>
  - GCIMS - Hector: <https://hector.msdlive.org>
  - GCIMS - GCAM Wrapper: <https://gcamwrapper.msdlive.org>
  - GCIMS - Stitches: <https://stitches.msdlive.org>
  - IM3 - StateModify: <https://statemodify.msdlive.org>

# Interactive Data Dashboards



<https://raft-hurricane-projections.msdlive.org/>



# Jupyter Notebooks

The MSD community is leveraging the cloud-computing capabilities in MSD-LIVE to train new people and teams to understand and extend MSD data and models.



**Sandbox**



**Classroom**

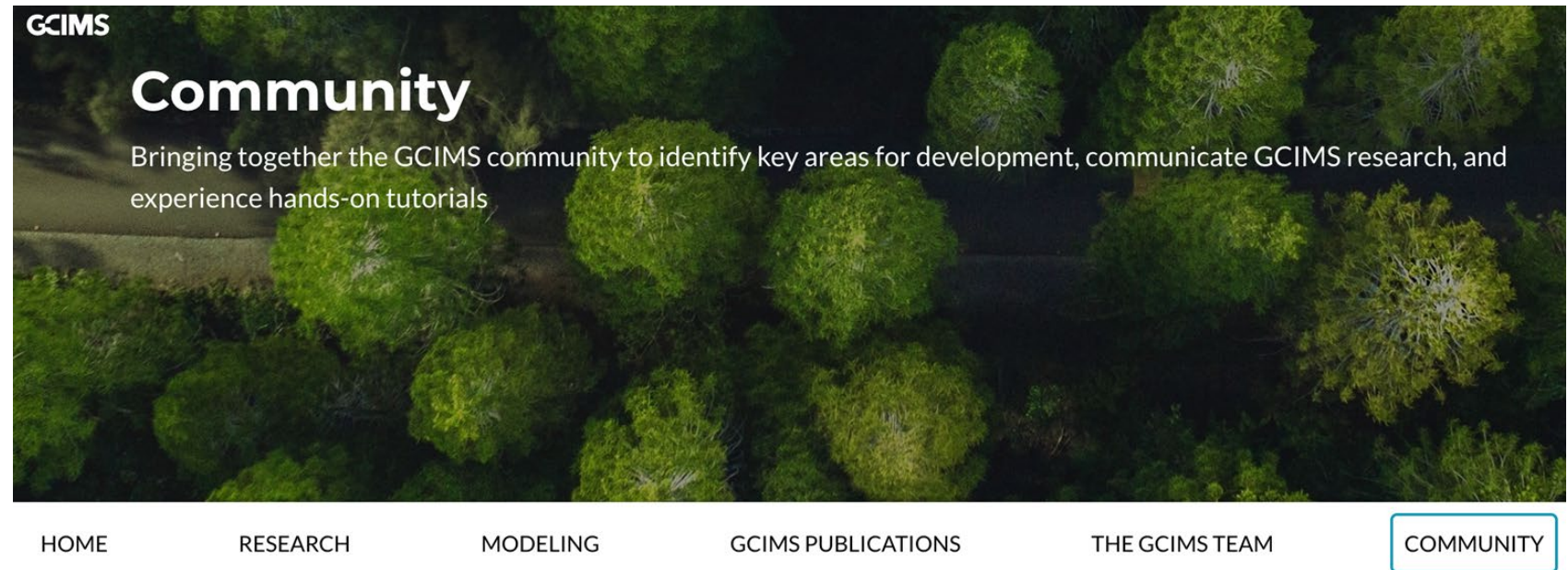


**Library**



# Training New Users

- At the 2024 GCAM Annual Meeting we used MSD-LIVE to train attendees on seven GCAM ecosystem models (demeter, gcamwrapper, hector, matilda, rgcam xanthos, and stitches).
- Attendees were able to run the models and analyze the output in real-time on the AWS cloud.
- 250+ people launched live computing instances over the course of the two-day training.



## Global Change Analysis Model (GCAM) Annual Meetings

The Joint Global Change Research Institute's next annual GCAM Community Modeling Meeting will be held June 5-6, 2024 virtually. [Registration is now open.](#)

The GCAM Community Modeling Meeting provides a forum for the GCAM community to hear about the latest developments and current uses of GCAM. For those new to GCAM, this workshop continued to serve as an excellent in-depth opportunity to learn and gain a better understanding of the model.

Check out what was presented here:

- [2023 GCAM Annual Meeting Presentations and Posters](#)
- [2022 GCAM Annual Meeting Presentations](#)
- [2022 GCAM Annual Meeting Posters](#)
- [2019 \(and earlier\) GCAM Annual Meeting Presentations.](#)

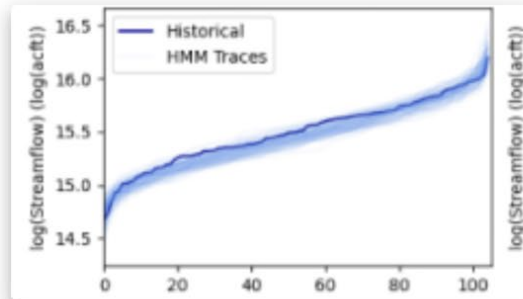




Computational Resources / Model Training Notebooks

## Model Training Notebooks

MSD projects are using the cloud-computing capabilities in MSD-LIVE to create interactive Jupyter notebooks to train new users to configure, run, and analyze MSD models.



### statemodify

statemodify is an open-source Python package that provides users with a way to easily modify StateMod's input and output files to enable exploratory modeling.

```
Plot up the supply and demand curves

import matplotlib.pyplot as plt

fig, ax = plt.subplots()
ax.plot(bio_sd.price, bio_sd.supply, label="supply")
ax.plot(bio_sd.price, bio_sd.demand, label="demand")
ax.set_ylabel("quantity")
ax.set_xlabel("price")
ax.legend()
```

### gcamwrapper

gcamwrapper contains all C++ and R/python source code to wrap the GCAM model such that simulations can be run interactively.

```
!param coord:      data frame for coordinates [lat, lon]
!param legend_label: string for legend label
!param cmap:       string for color palette
...

cols_str, cols_dict = name_columns(tas, start_yr=1971)
df = pd.DataFrame(data=arr, columns=cols_str, mean(axis=1), rename={"value"})
df = pd.concat([df, coord], axis=1)

fig = plt.figure(figsize=(15,6), constrained_layout=True)
spec = fig.add_subplot(1, 1)
ax = fig.add_subplot(spec[0], 0)
p = ax.scatter(df.lat, df.lat, s=1, c=df.value, cmap=cmap)
legend_binds = [{"label": legend_label,
                  "shrink": 1}]
```

### xanthos

xanthos is an open-source hydrologic model written in Python that simulates historical and future global water availability on a monthly time step.

- MSD-LIVE is becoming a living library for MSD models.
- Our innovative approach using Jupyter notebooks is making it easier to train new users to configure, run, and analyze MSD models.
- Can be used to entrain new users, including those from small and resource-limited colleges, into the MSD research community.

# Ways You Can Use MSD-LIVE Today

1. Archive your data.
2. Get a data DOI.
3. Find data produced by other MSD projects.
4. Share working data with collaborators.
5. Create a data dashboard so that others can visualize or interact with your data without having to download it.
6. Create a model training notebook to teach others to install, configure, run, and analyze your model.
7. Access training and documentation on open science best practices (e.g., writing readmes, using DOIs, licenses).
8. Find and watch archived MSD CoP webinars.

Core Capabilities	Use Cases
Data Repository	Find Data
	Archive Data
	Version Control
	Training
	Control Access
Advanced Computing	Share Data
	Analyze Data
	Multi-Model Workflows
	ML/AI
	Move Code to the Data



# Lessons Learned

- Facilitate, not complicate
- Demonstrate value to avoid box-checking
- Listen to your users: Often, actively, and with an open mind
- Don't underestimate how much training you will need to do
- Encourage small steps while you wait for “the big thing”



# Backup Slides



# How Do We Make a Transition to Open Science?

Journals largely skipped to this end of the pyramid...



Make it required

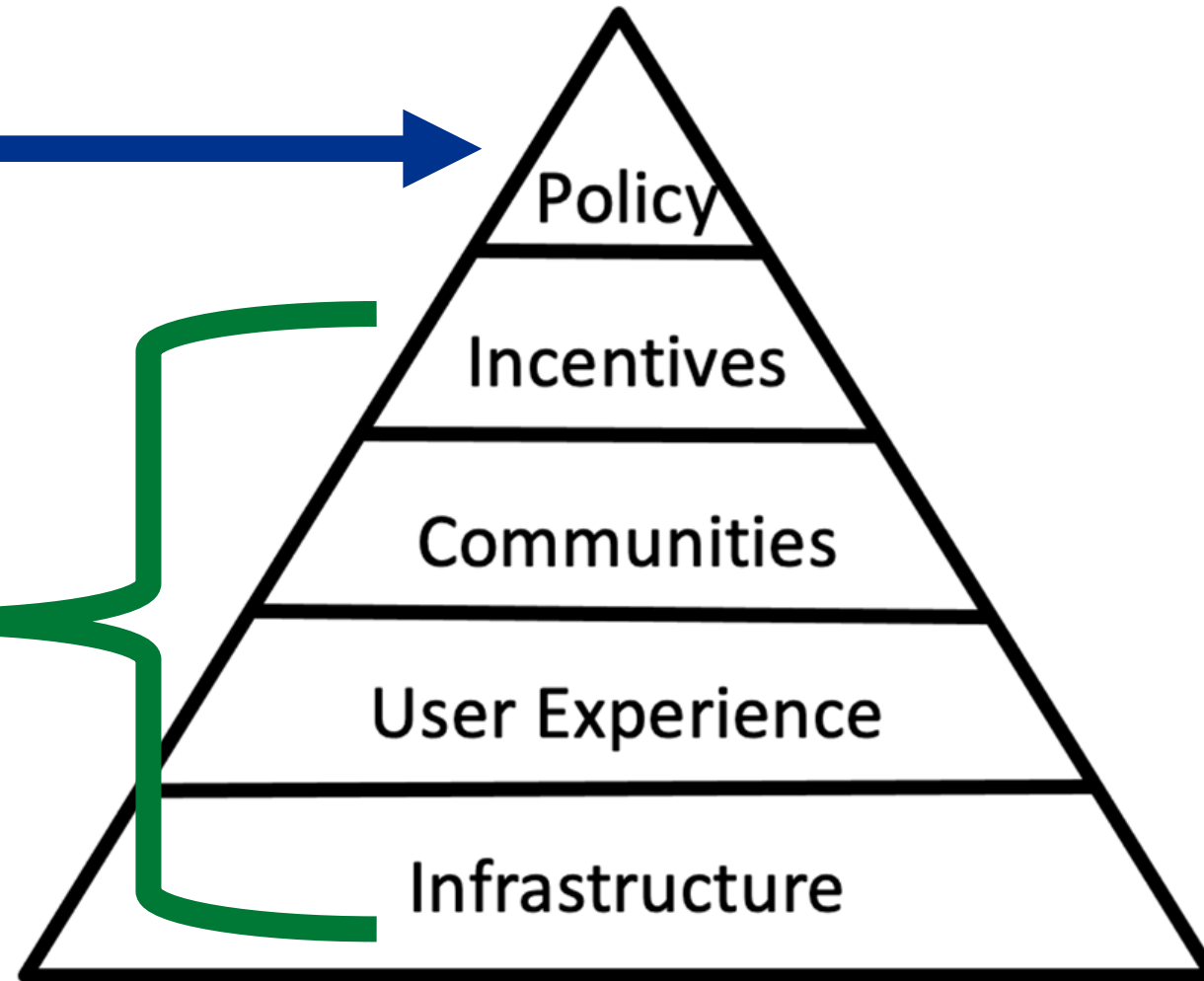
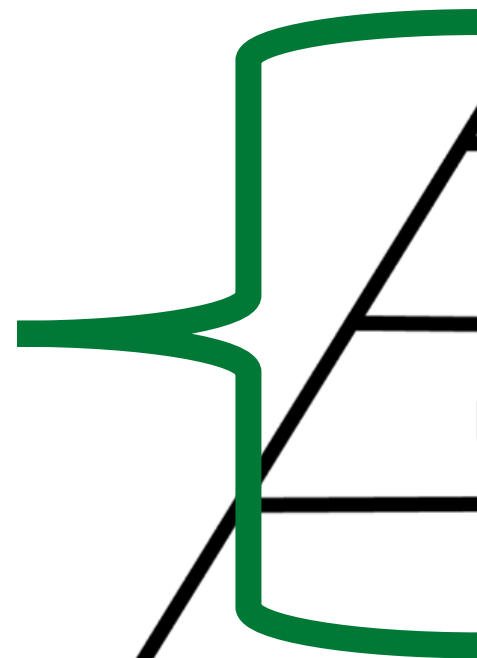
Make it rewarding

Make it normative

Make it easy

Make it possible

MSD-LIVE and the MSD CoP are tackling these foundational elements of the pyramid...



*Conceptual diagram from Brian Nosek of the University of Virginia and the Center for Open Science*

# Steppingstones to Open Science





## Data and Code Repository



Discover

Search the MSD-LIVE data and code repository



Submit Data

Upload, archive, and share your data and code

### Access Status

- Open 50
- Restricted 2

### File Status

- In MSD-LIVE 43
- Metadata-only 9

### Resource Types

- Dataset 38
- Presentation 7
- Software 5
- Other 2

### Keywords

- Water 15
- Climate Change 11
- Urban 11
- Energy 10
- Earth Systems 9
- Economics 7
- Land 7
- GCAM-USA 6
- Scenario 6
- Electricity Demand 5

### Status

- Published 52

### Project

- IM3 30
- PCHES 10
- BSEC UIFL 7

## Recent Contributions

February 18, 2024 (v1) Model Output Open

### Globally Gridded Groundwater Extraction Volumes and Costs under Six Depletion and Poned Depth Targets

Niazi, Hassan ; Ferencz, Stephen; Yoon, Jim

This repository contains simulated outputs from superwell – a hydro-economic tool for long-term assessment of groundwater cost and supply – providing globally gridded groundwater extractable volumes and associated unit costs (\$/km<sup>2</sup>) for accessible groundwater production, based on a variety of user-defined depletion and ponded depth scenarios. Ke...

Energy Water GCIMS

Uploaded on February 23, 2024

February 8, 2024 (v1) Model Output Open

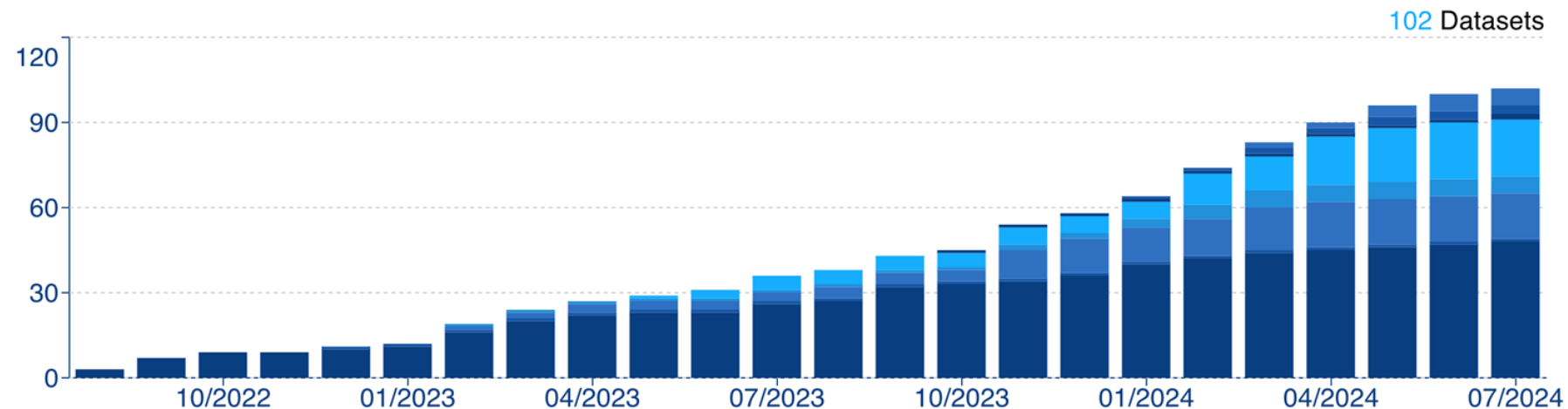
### Model run output for WM-ABM 2024 HESS Publication

Yoon, Jim

Model output for computational experiments introducing WM-ABM. See associated HESS publication here: <https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1604/>

# Usage Statistics


- v1 released 22-Aug 2022
- v2 released 28-Jun 2023
- 10 projects
- 210 registered users
- 102 published datasets
- 185+ TB of total data
- ~1150 usages of the Jupyter notebooks in the past three months




**210 Users**




**10 Projects**



**102 Datasets**



**75.9k Files**



**185.5 TB Stored**



**Downloads**



**1166 Launches**

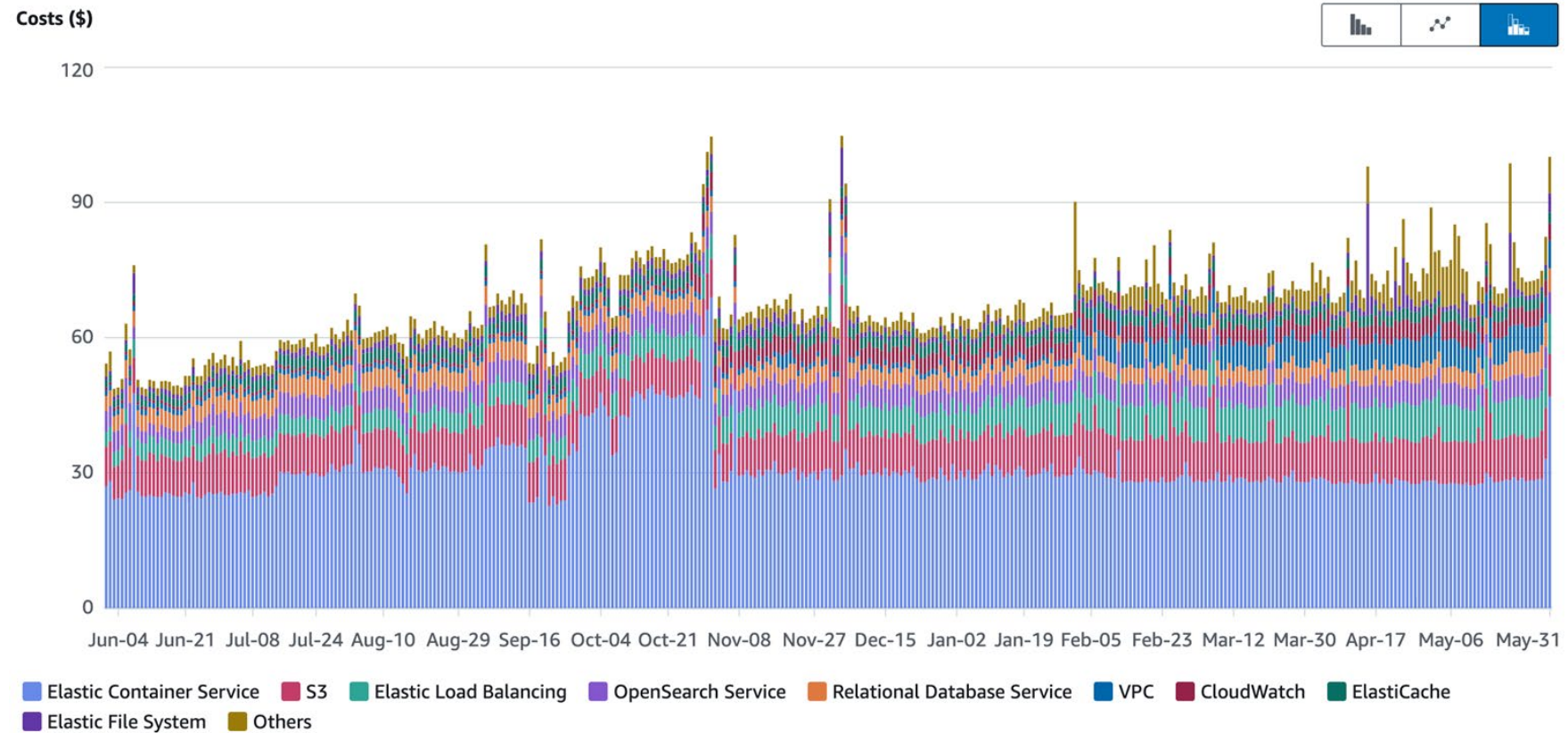


# Cloud Cost Statistics

Total cost  
**\$24,697.14**

Average daily cost  
**\$67.48**

Service count  
**14**



AWS Service	Daily Cost
ECS (runs basically everything on the site)	~\$30
S3 (data storage)	~\$8
Jupyter notebooks	~\$9
Everything else	~\$20