Management of Reaction Data with the Open Reaction Database

https://open-reaction-database.org/
help@open-reaction-database.org

Benjamin J. Deadman (ben@bjdeadman.co.uk)

ChemSpider Webinar Data Storage and Management
19th May 2026

The State of Reaction Data in the 21st Century

Lincoln University Library, Teaching and Learning's Photostream, **2010**, CC BY-SA 2.0, flickr.com/photos/40028943@N07/5061683954

The need for more & better reaction data

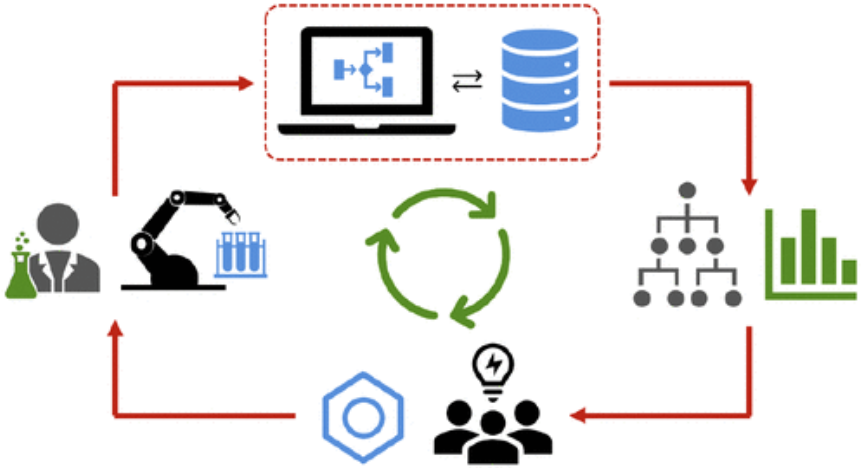
A growing chorus across the community calls for open, structured reaction data.

- **2019 —** *The digitization of organic synthesis* · Davies, *Nature*
- **2020 —** *Molecular Machine Learning: The Future of Synthetic Chemistry?* · Pflüger & Glorius, *Angew. Chem.*
- **2021 —** *Call for a Public Open Database of All Chemical Reactions* · Baldi, *J. Chem. Inf. Model.*
- **2023 —** *Data Sharing in Chemistry: A Case for Mandating Structured Reaction Data* · Mercado, Kearnes & Coley, *J. Chem. Inf. Model.*
- **2024 —** *AI for Retrosynthetic Planning Needs Both Data and Expert Knowledge* · Strieth-Kalthoff *et al., JACS*

The Open Reaction Database

- Building predictive models for chemistry relies on the availability of structured reaction data
- The ORD is an initiative to "support machine learning and related efforts in reaction prediction, chemical synthesis planning, and experiment design”

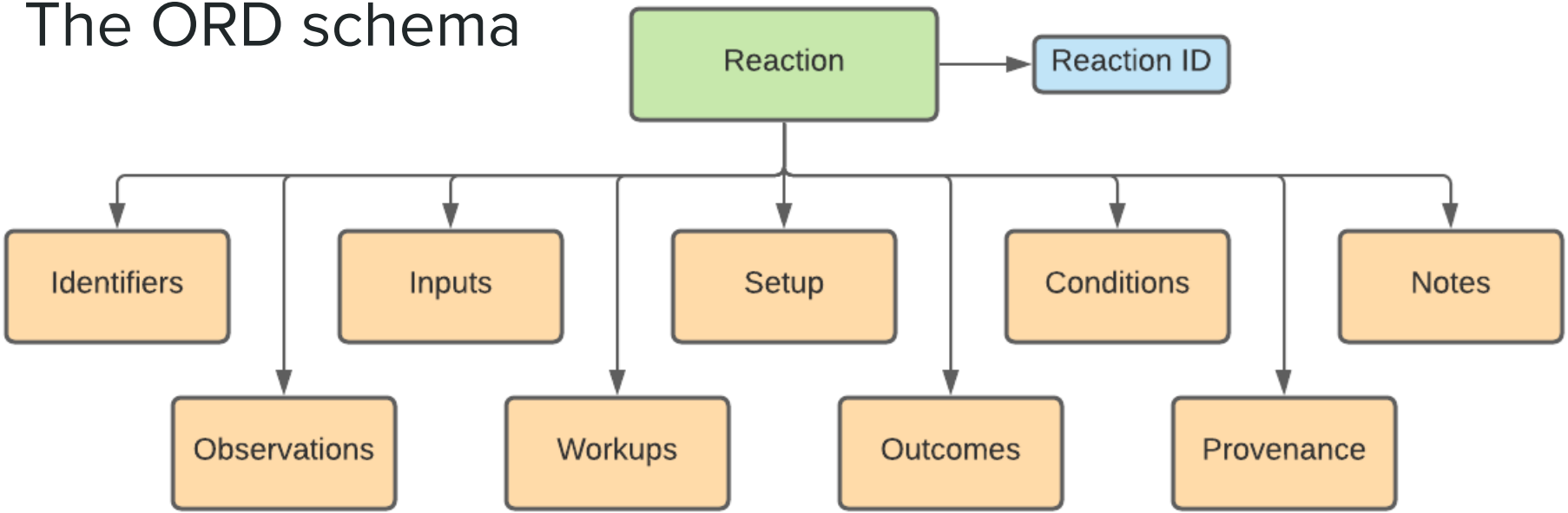
Experimenters



AI and Data Scientists

Kearnes et al. *JACS* **2021**, doi.org/10.1021/jacs.1c09820

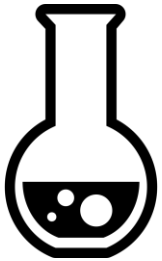
The ORD schema



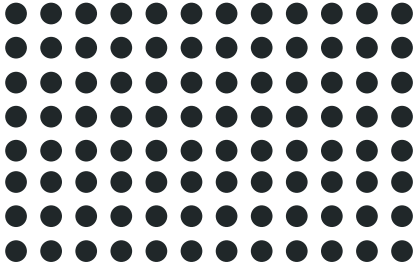
- Capture the most important aspects of reactions in a *structured* format.
- Allow additional details in a flexible, *unstructured* format.

Kearnes et al. *JACS* **2021**, doi.org/10.1021/jacs.1c09820

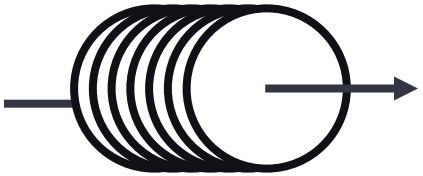
Traditional and Modern Synthesis



Batch Synthesis



Singleton or High-Throughput

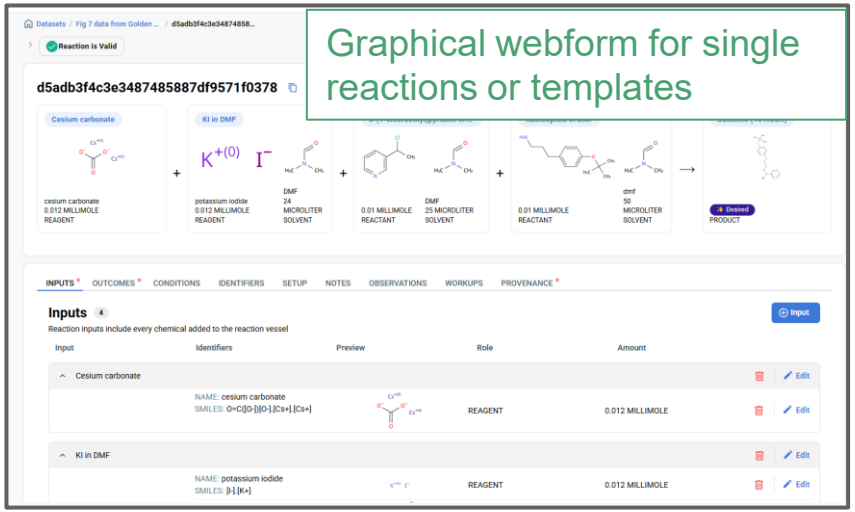


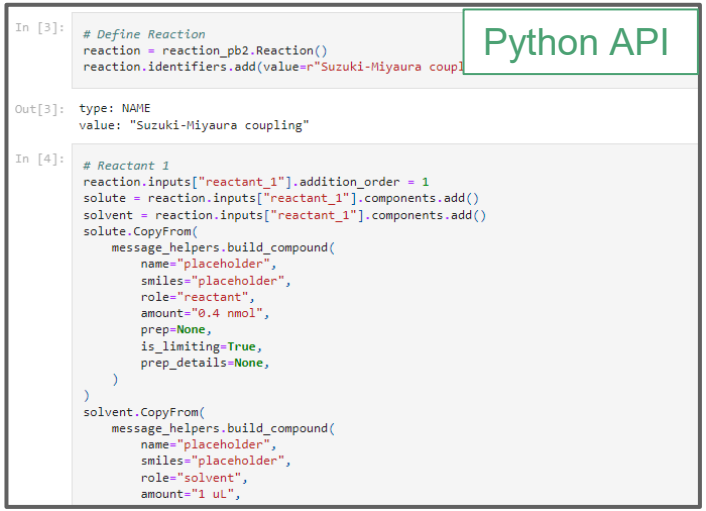
Flow Chemistry

Photochemistry

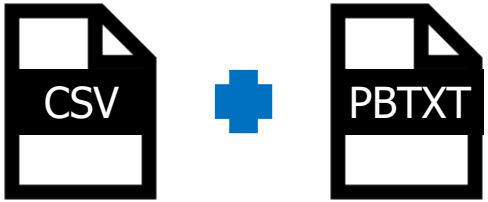
Electrochemistry

Data enumeration workflow





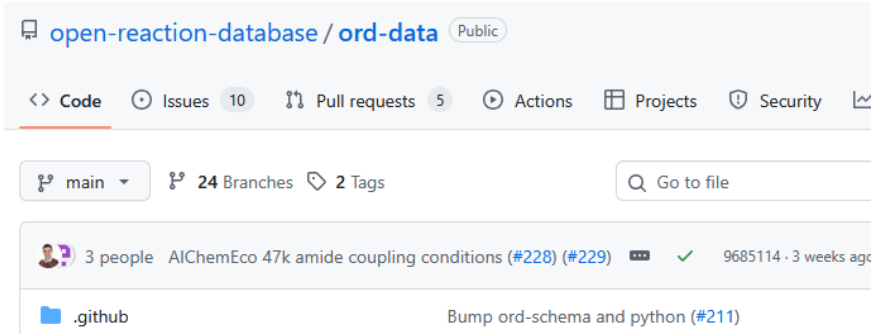
Dataset defined by iterating a template over a spreadsheet of Reactions and their changing parameters



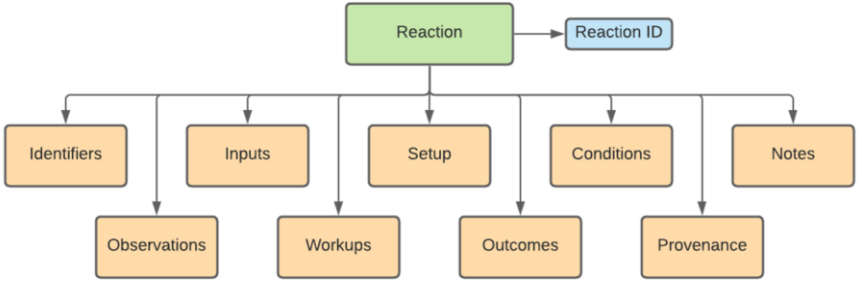
- HTE full factorial datasets
- Optimisation tables
- Substrate scope tables

ORD tools

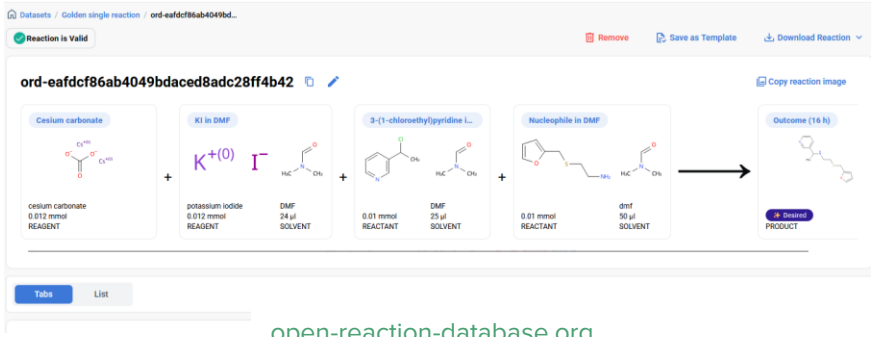
ord-data repository



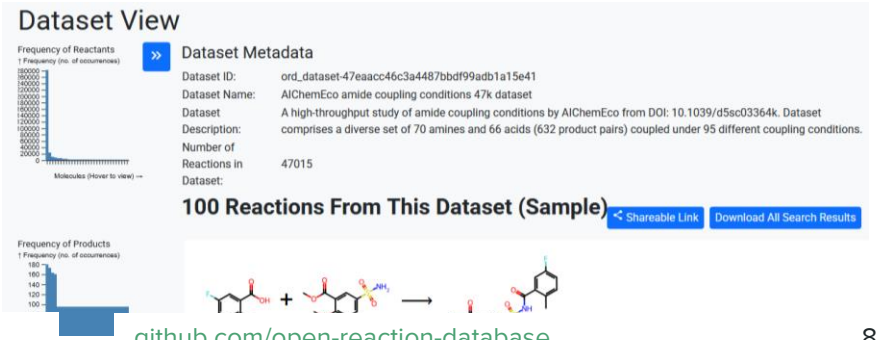
ord-schema package



ord-app reaction editor



ord-interface



open-reaction-database.org

github.com/open-reaction-database

Data stored as protocol buffers

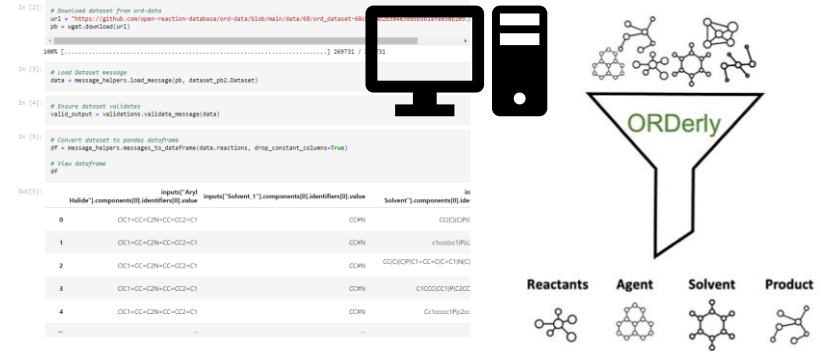
Exchangeable

ORD dataset

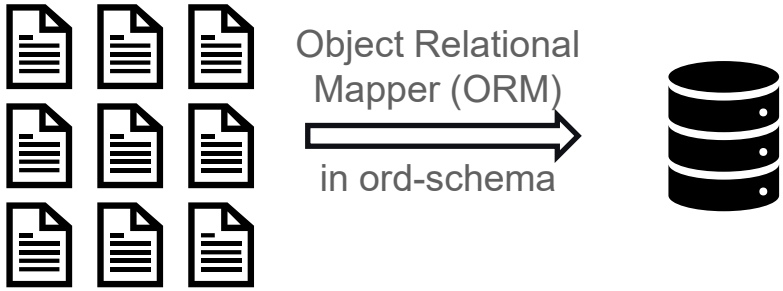
Extendable & Evolving

```
conditions {
  temperature {
    control {
      type: DRY_ALUMINUM_PLATE
      details: "additional free-text descriptions"    }
    setpoint {
      value: 80.0
      precision: 5.0
      units: CELSIUS
    } }}
```

Machine Actionable



Unpack datasets into a searchable database



collection of ORD datasets

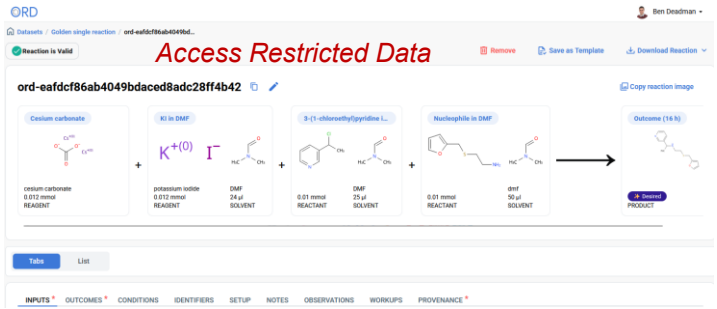
PostgreSQL database

Wigh et al. *J. Chem. Inf. Model.* **2024**, doi.org/10.1021/acs.jcim.4c00292

Open access policies

Software Vendors are also welcome to re-use ORD tools in their products under an Apache 2.0 license.

ord-app reaction editor
app.open-reaction-database.org



User initiated submission via GitHub

Public ORD Repository & Database

CC-BY-SA 4.0

optional release to public ORD

OR you can send your data to a private repository

Institutional or private repositories and/or reaction editors

Pool ORD formatted data from all repositories you have access to.

Some Quick Notices

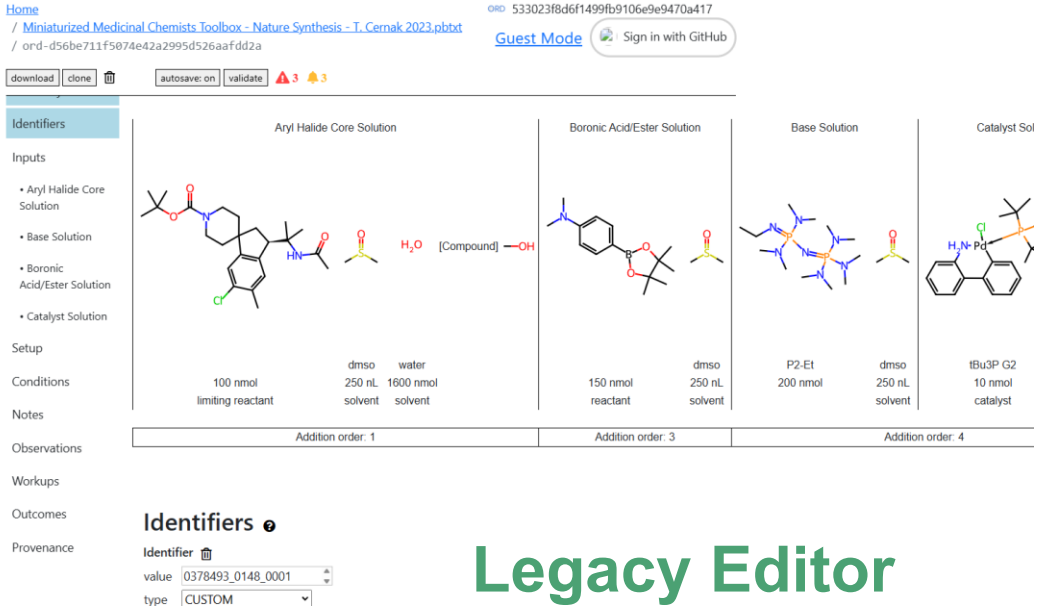
- **Dataset rewards —** $300 vouchers for eligible 2026 datasets
- **Legacy editor —** the old reaction editor is being retired

Dataset Rewards

- Receive a $300 Tremendous voucher for eligible datasets accepted in 2026
- Global participation
- 3 ways to get involved:
  - Pick from the Dataset Wishlist
  - Invited papers
  - Suggest a priority paper
- Terms and conditions apply

Closing the Legacy Reaction Editor

- New ORD Reaction Editor replaced the legacy editor in **July 2025**.
- Old editor still online at https://open-reaction-database.org/editor/dataset but will **shutdown after 31 July 2026.**
- Download your datasets if you want to keep them.



Acknowledgements

open-reaction-database.org
linkedin.com/company/open-reaction-database
help@open-reaction-database.org

Governing Committee
Connor Coley (MIT)
Abby Doyle (UCLA, C-CAS)
Spencer Dreher (Merck)
Klavs Jensen (MIT)
Steven Kearnes (Genesis Molecular AI)

Former Committee Members
Joel Hawkins (Pfizer)

Key Contributors
Mike Wleklinski (Merck)
Anton Kast (Google)
Brian Lee (Google)
Ian Pendleton (Alkermes)
Michael Maser (Caltech)
Michael Kennel (U. Notre Dame)
Olaf Wiest (U. Notre Dame)

Advisory Board
Alán Aspuru-Guzik (Toronto, MADNESS)
Tim Cernak (Michigan, Entos)
Lucy Colwell (Cambridge, SynTech, Google)
Werngard Czechtizky (AstraZeneca)
JW Feng (Consultant)
Matthew Gaunt (Cambridge, SynTech)
Alex Godfrey (Consultant)
Mimi Hii (A*STAR, Imperial, ROAR)
Greg Landrum (ETH Zurich)
Fabio Lima (Novartis)
Christos Nicolaou (Novo Nordisk)
Sarah Reisman (Caltech, C-CAS)
Francesco Rianjongdee
Matthew Sigman (Utah, C-CAS)
Jay Stevens (BMS)
Sarah Trice
Huimin Zhao (UIUC, MMLI)

ORDerly
J. Chem. Inf. Model., 2024, 64, 9, 3790.
https://doi.org/10.1021/acs.jcim.4c00292
(from Cambridge)
Daniel Wigh
Joe Arrowsmith
Alexander Pomberger
Kobi Felton
Alexei Lapkin

A special thank you to all our data contributors!