



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



**Current publication practices aren't
ready for AI**

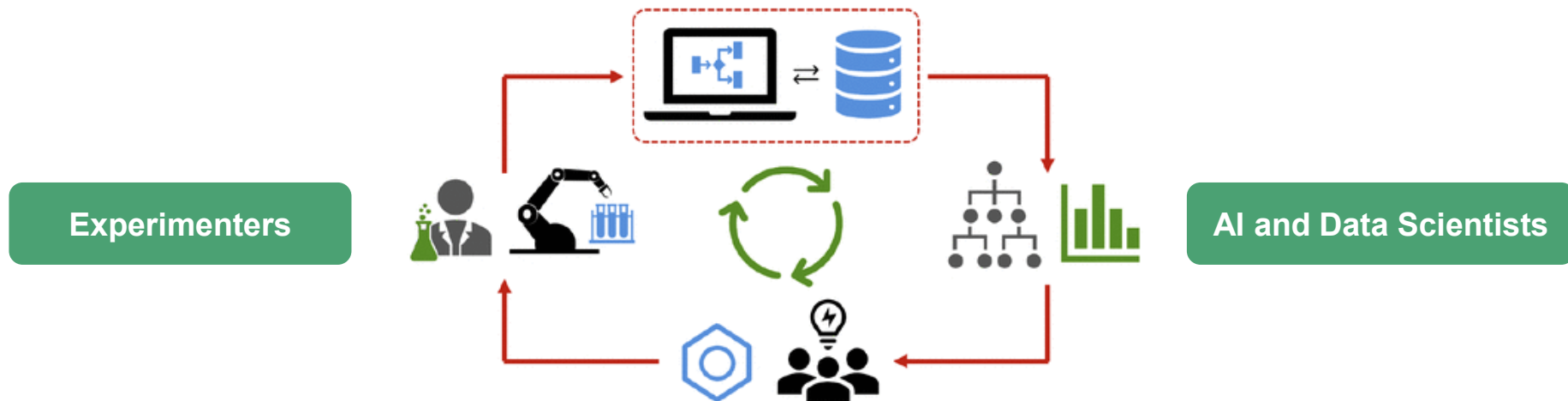
The need for more & better reaction data

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

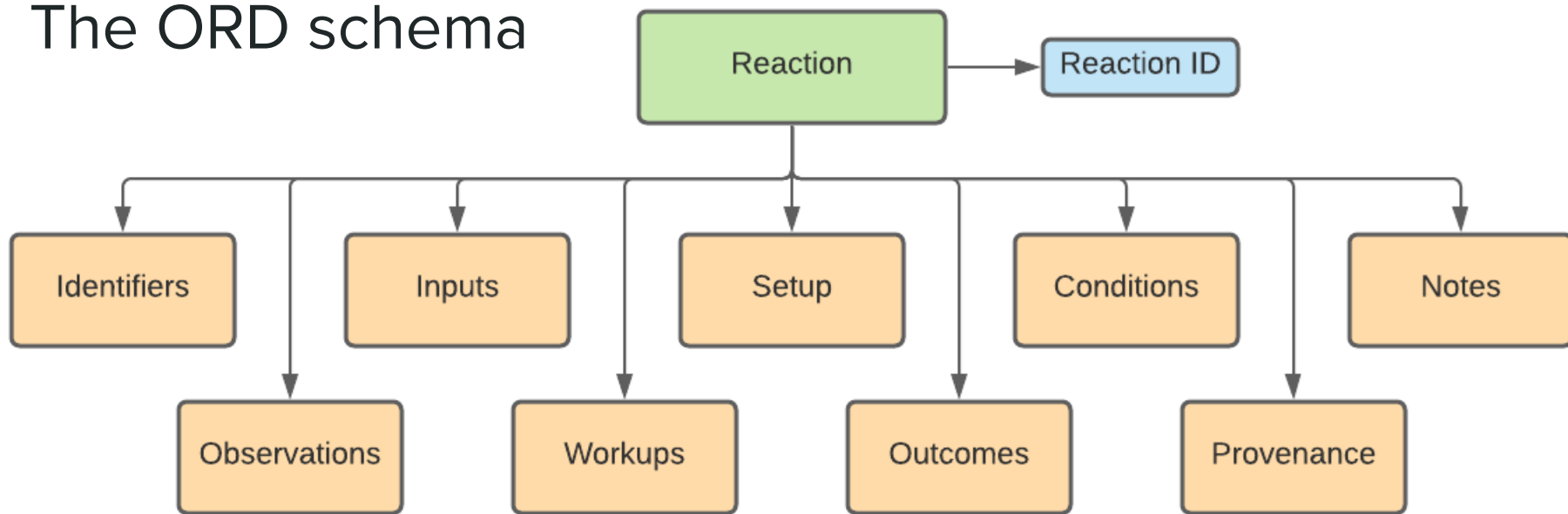
- **2019** — *The digitization of organic synthesis* · Davies, *Nature*
- **2020** — *Molecular Machine Learning: The Future of Synthetic Chemistry?* · Pflüger & Glorius, *Angew. Chem.*
- **2021** — *Call for a Public Open Database of All Chemical Reactions* · Baldi, *J. Chem. Inf. Model.*
- **2023** — *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"

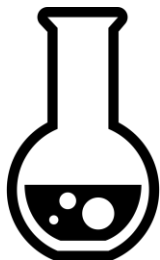


The ORD schema



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

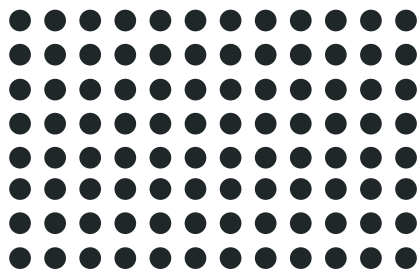
Traditional and Modern Synthesis



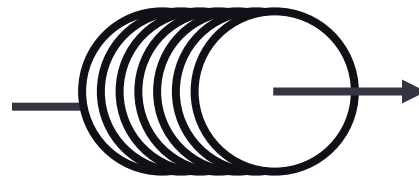
Batch Synthesis



Photochemistry



Singleton or High-Throughput



Flow Chemistry



Electrochemistry

Data enumeration workflow

Graphical webform for single reactions or templates

Input	Identifiers	Preview	Role	Amount
Cesium carbonate	NAME: cesium carbonate SMILES: O=C([O-])[K+].[Cs+]		REAGENT	0.012 MILLIMOLE
KI in DMF	NAME: potassium iodide SMILES: [I-].[K+]		REAGENT	0.012 MILLIMOLE

Python API

```
In [3]: # Define Reaction
reaction = reaction_pb2.Reaction()
reaction.identifiers.add(value="Suzuki-Miyaura coup]

Out[3]: type: NAME
value: "Suzuki-Miyaura coupling"

In [4]: # Reactant 1
reaction.inputs["reactant_1"].addition_order = 1
solute = reaction.inputs["reactant_1"].components.add()
solvent = reaction.inputs["reactant_1"].components.add()
solute.CopyFrom(
    message_helpers.build_compound(
        name="placeholder",
        smiles="placeholder",
        role="reactant",
        amount="0.4 nmol",
        prep=None,
        is_limiting=True,
        prep_details=None,
    )
)
solvent.CopyFrom(
    message_helpers.build_compound(
        name="placeholder",
        smiles="placeholder",
        role="solvent",
        amount="1 ul",
```



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

open-reaction-database / ord-data Public

<> Code Issues 10 Pull requests 5 Actions Projects Security

main 24 Branches 2 Tags Go to file

3 people AIChemEco 47k amide coupling conditions (#228) (#229) 9685114 · 3 weeks ago

.github Bump ord-schema and python (#211)

ord-app reaction editor

Reaction is Valid Remove Save as Template Download Reaction

ord-eafdcf86ab4049bdaced8adc28ff4b42

Cesium carbonate
cesium carbonate
0.812 mmol
REAGENT

KI in DMF
potassium iodide
0.812 mmol
REAGENT
DMF
24 µl
SOLVENT

3-(1-chloroethyl)pyridine L.
0.01 mmol
REACTANT
DMF
25 µl
SOLVENT

Nucleophile in DMF
0.01 mmol
REACTANT
DMF
90 µl
SOLVENT

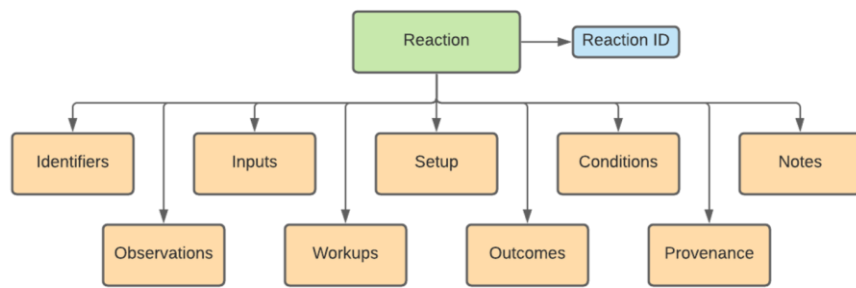
Outcome (16 h)
5.00000
PRODUCT

Copy reaction image

0.01 mmol REAGENT 0.01 mmol REACTANT 0.01 mmol REACTANT 0.01 mmol REACTANT

open-reaction-database.org

ord-schema package



ord-interface

Dataset View

Dataset Metadata

Dataset ID: ord_dataset-47eaacc46c3a4487bbdf99adb1a15e41
Dataset Name: AIChemEco amide coupling conditions 47k dataset
Dataset Description: A high-throughput study of amide coupling conditions by AIChemEco from DOI: 10.1039/d5sc03364k. Dataset comprises a diverse set of 70 amines and 66 acids (632 product pairs) coupled under 95 different coupling conditions.
Number of Reactions in Dataset: 47015

100 Reactions From This Dataset (Sample) [Shareable Link](#) [Download All Search Results](#)

Frequency of Reactants
Frequency of Products

github.com/open-reaction-database

Data stored as protocol buffers

Exchangeable



Machine Actionable

```
20 [21] # Download dataset from ord-data
url1 = "https://github.com/open-reaction-database/ord-data/raw/main/data/68/ord-dataset-68"
url2 = "https://github.com/open-reaction-database/ord-data/raw/main/data/68/ord-dataset-68"
pb = wget.download(url1)

22 [22] # Load Dataset message
data = message_helpers.load_message(pb, dataset_pb2.Dataset)

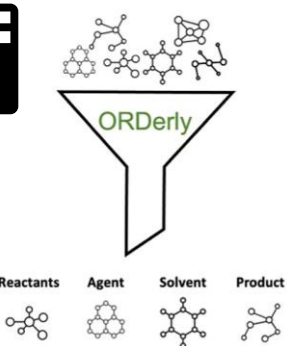
23 [23] # Ensure dataset is valid
valid_output = validations.validate_message(data)

24 [24] # Convert dataset to pandas dataframe
df = message_helpers.messages_to_dataframe(data.reactions, drop_constant_columns=True)

25 [25] # View dataframe
df
```

	Inputs["Aryl Halide"].components[0].identifiers[0].value	Inputs["Solvent"].components[0].identifiers[0].value	Solvent].components[0].label
0	OC1=CC=CRN=CC=CC=C1	CCN	CC(C)OPF
1	OC1=CC=CRN=CC=CC=C1	CCN	cHexoxTPig
2	OC1=CC=CRN=CC=CC=C1	CCN	CC(C)OPC1=CC=CC=C1N(C)
3	OC1=CC=CRN=CC=CC=C1	CCN	ClCC(C)OPC2C
4	OC1=CC=CRN=CC=CC=C1	CCN	CrHexoxPh2ox

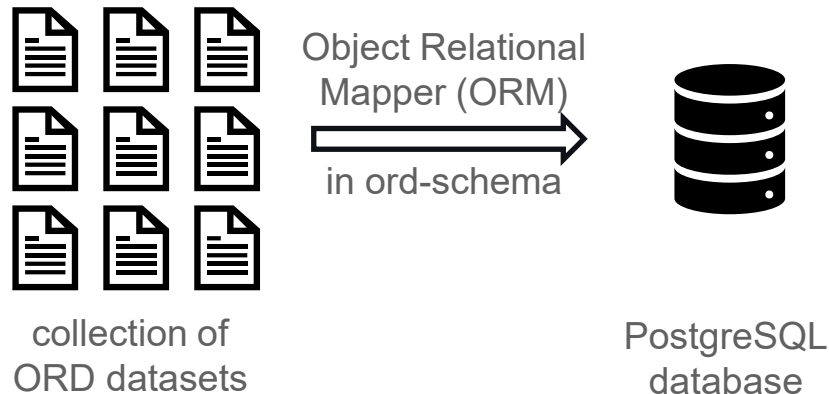
Chemically Informed Cleaning



Extendable & Evolving

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

Unpack datasets into a searchable database



Open access policies

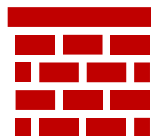
ord-app reaction editor
app.open-reaction-database.org

The screenshot shows the ORD app interface. At the top, it says "ORD" and "Database / Golden single reaction / ord-eafdcf86ab4049bdaced8adc28ff4b42". Below that, it says "Reaction is Valid" and "Access Restricted Data". The main area shows a chemical reaction with reagents: cesium carbonate, potassium iodide, DMF, 5-(1-chloroethyl)pyridine, and Nucleophile in DMF. The product is shown as a complex organic molecule. At the bottom, there are tabs for "Inputs", "Outcomes", "Conditions", "Identifiers", "Setup", "Notes", "Observations", "Workups", and "Provenance".

OR you can
send your data
to a private
repository



User initiated
submission via
GitHub



**Public ORD
Repository &
Database**

CC-BY-SA 4.0

optional release
to public ORD

**Institutional or private
repositories and/or
reaction editors**

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

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

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

WANTED
— FOR OPEN SCIENCE —
REACTION DATA

REWARD
\$300
per qualifying dataset

← Scan to claim

The Open Reaction Database is offering bounties for important reaction datasets.

HTE	DOE or BO Optimizations	Reaction Kinetics
Automated Synthesis	ELN Exports	Structured Literature Reviews

This offer is valid worldwide. Terms and Conditions apply. While stocks last.

GLOBAL REWARD

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.

Home
/ Miniaturized Medicinal Chemists Toolbox - Nature Synthesis - T. Cernak 2023.pbtxt
/ ord-d56be711f5074e42a2995d526aafdd2a

ORD 533023f8d6f1499fb9106e9e9470a417
Guest Mode Sign in with GitHub

download clone autosave: on validate 3 3

Identifiers

Inputs

- Aryl Halide Core Solution
- Base Solution
- Boronic Acid/Ester Solution
- Catalyst Solution

Setup

Conditions

Component	Amount	Notes
100 nmol limiting reactant		
dmsolimiting reactant	250 nL	
water solvent	1600 nmol	
150 nmol reactant		
dmsolimiting reactant	250 nL	
P2-Et	200 nmol	
dmsolimiting reactant	250 nL	
tBu3P G2 catalyst	10 nmol	

Notes

Observations

Workups

Outcomes

Provenance

Identifiers

Identifier

value 0378493_0148_0001

type CUSTOM

Legacy Editor

Acknowledgements

open-reaction-database.org

[linkedin.com/company/open-reaction-database](https://www.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!