ChemCatBio
Chemical Catalysis for Bioenergy

Reduce, Reuse, Recycle: Data Benchmarking and Accessibility for Faster Research With the Catalyst Property Database

October 13, 2021

ChemCatBio Webinar Series

Kurt Van Allsburg, NREL

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY
BIOENERGY TECHNOLOGIES OFFICE

Housekeeping

- Attendees will be in listen-only mode
- Audio connection options:
  - Computer audio
  - Dial in through your phone (best connection)

- Use the Q&A panel to ask questions
- Technical difficulties? Contact Erik Ringle through the chat section, lower right of your screen
- Recording will be available at: https://www.chemcatbio.org/webinars.html

Today's Speaker

Kurt Van Allsburg

Catalyst Scientist, National Renewable Energy Laboratory

ChemCatBio
Chemical Catalysis for Bioenergy

Reduce, Reuse, Recycle: Data Benchmarking and Accessibility for Faster Research With the Catalyst Property Database

October 13, 2021

ChemCatBio Webinar Series

Kurt Van Allsburg, NREL

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY
BIOENERGY TECHNOLOGIES OFFICE

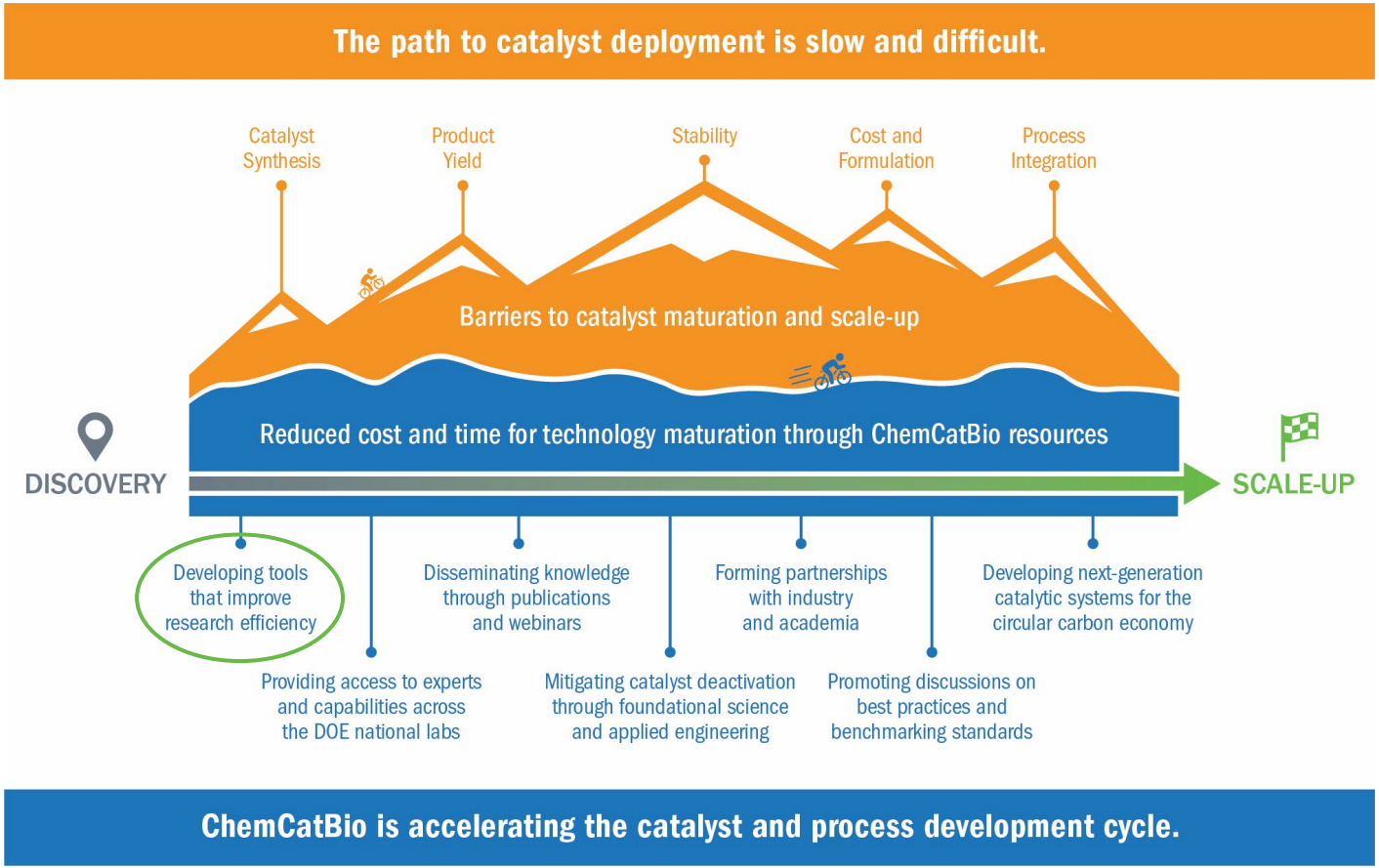
The Chemical Catalysis for Bioenergy consortium (ChemCatBio)

Accelerating the catalyst and process development cycle

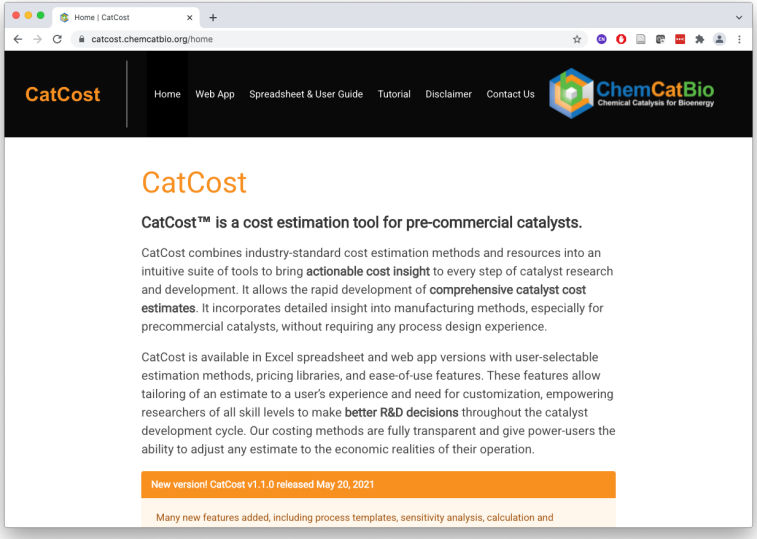
https://chemcatbio.org

Subscribe to our newsletter **The Accelerator**

ChemCatBio: Accelerating Catalyst Discovery



Released in 2018: CatCost



CatCost™

a free catalyst cost estimator

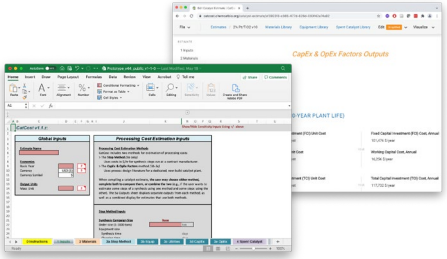
released in 2018

major update in May 2021

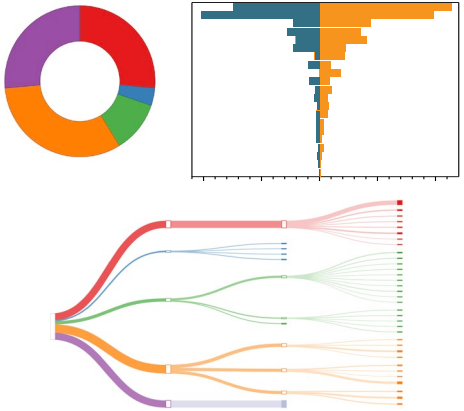
+ new mailing list

catcost.chemcatbio.org

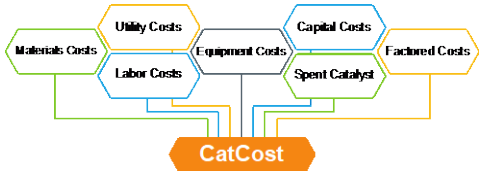
Excel and web app versions, with Python interconversion



Advanced visualizations built in



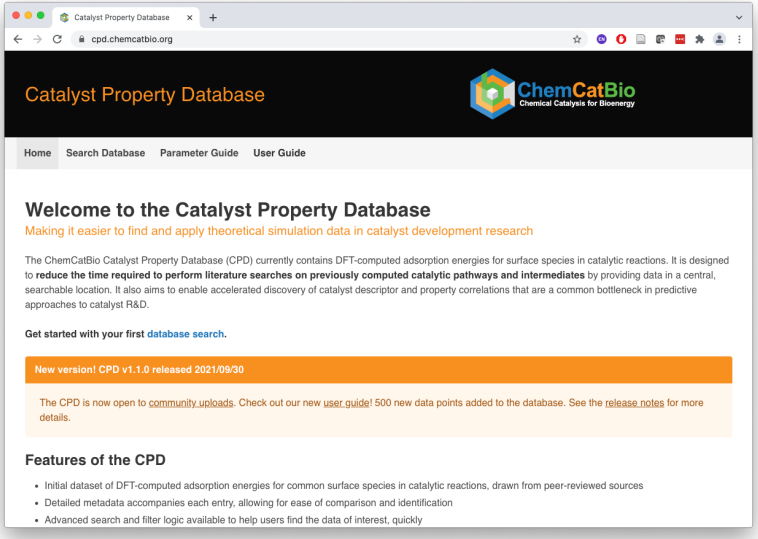
Detailed capital and operating costs



PI: Fred Baddour

Team Members: Kurt Van Allsburg, John Super, Eric Tan, Lesley Snowden-Swan, John Frye, James White, Joshua Schaidle, Michael Talmadge, Jesse Hensley, Susan Habas, Nicholas Wunder, Kenny Gruchalla, John Yarbrough, Kristi Potter, Matt Jankousky, Andy Young

Today's Webinar: the Catalyst Property Database



The Catalyst Property Database

a free database for catalyst research

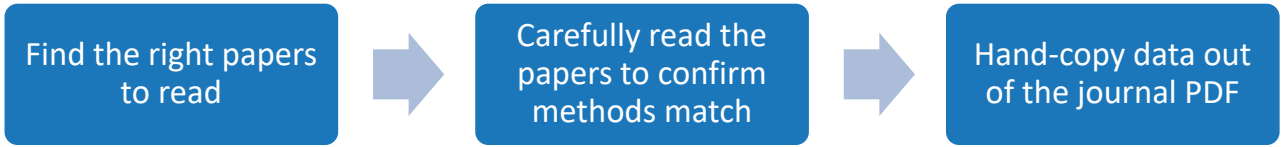
released in 2020

opened to external contributions Sept 2021

cpd.chemcatbio.org

Why build a catalyst database?

Problem: Finding the right catalysis data in the literature is often slow and cumbersome

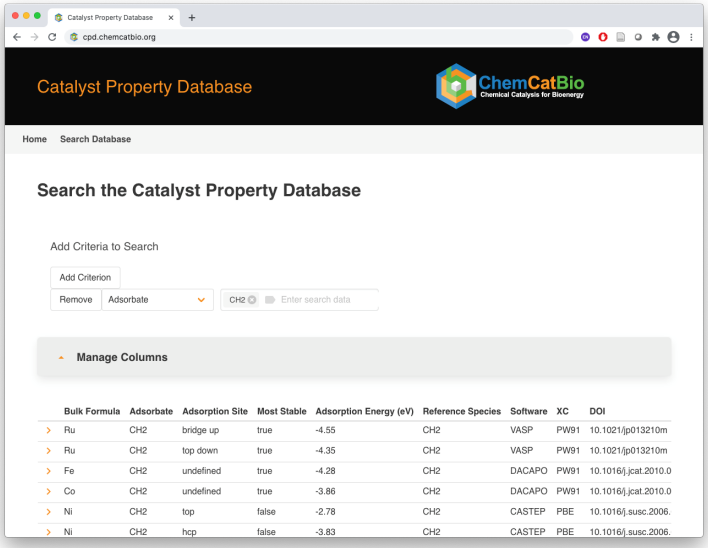


Result: Because of the challenges in finding **reliable, directly comparable** datasets, it's often easier to just **collect/compute it yourself (duplication of effort, $$$, time)**

Our goal: **Reduce, Reuse, Recycle**

Accelerate catalysis research by making it easier to find and reuse catalyst data

Our answer: make catalyst data more accessible



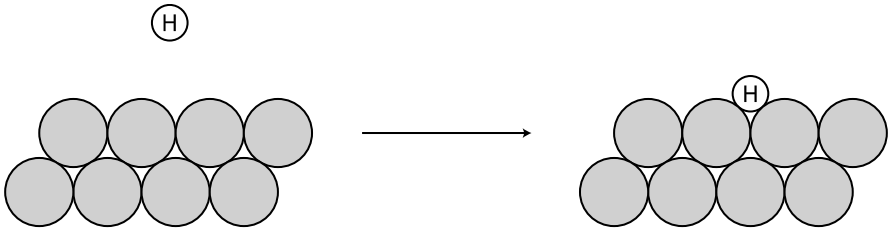
Data in the CPD is:

- Centralized
  - Enabling faster comparisons
- Searchable
  - Search dozens of metadata fields directly
  - Find the right data faster
- Publicly accessible to view and upload
  - Uploads subject to quality control

Data types:

- Computational catalysis (initial focus)
- Catalyst characterization
- Reaction performance

What sort of catalyst property data?



At initial release: Density functional theory-computed adsorption energies for intermediates on catalyst surfaces, from peer reviewed journals

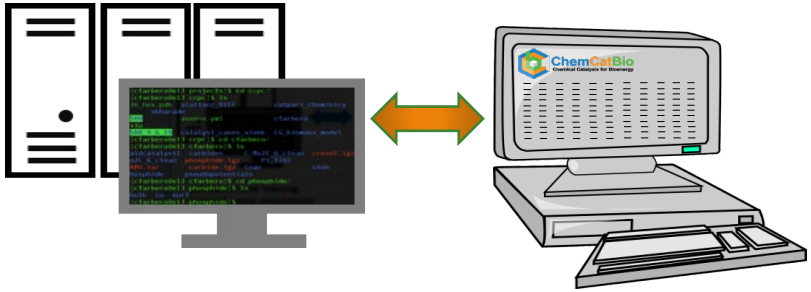
Long-term goals:

- Add new data types to enable scaling relation and reactivity descriptor discovery
- Experimental and computational results
- Catalyst property and preparation details
- Researcher-users of the CPD to guide data types and structure

Applications of the CPD

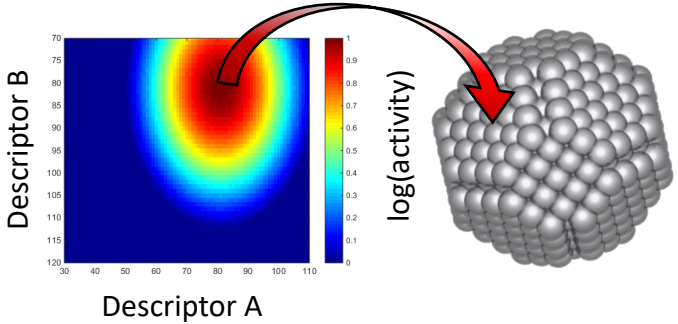
Benchmarking & Validation

Researchers wanting to check their results

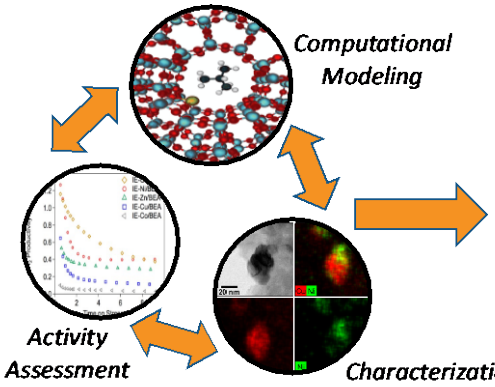


Catalyst Screening

Applying reactivity data to identify new target compositions



Interpretation of Experimental Trends

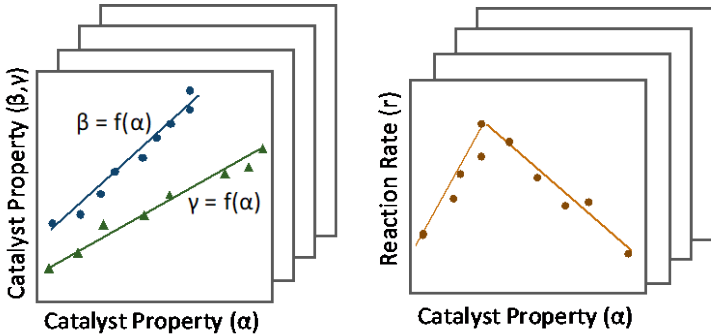


Investigate

- Active Site
- Mechanism
- Structure-Function Relationships

Discovery of Scaling Relations and Reactivity Descriptors

Harnessing the power of a large, central database



Vision to Reality: Not So Easy

✓ **Part I: Why we built it**

Part II: How we built it (and why you should care)

... and examples, upcoming additions, etc.

The CPD Team

CHEM TEAM

Kurt Van Allsburg

PI

Experimentalist

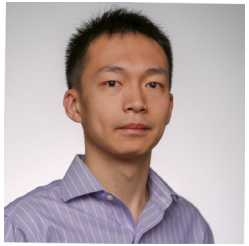
Sean Tacey

Computational researcher

Carrie Farberow

Co-investigator

Computational researcher



Qiyuan Wu

Experimentalist

Experienced web developer

DEV TEAM

Nalinrat Guba

Software engineer (previously at Oracle)

Rachel Hurst

Technical project lead, system architect

Alicia Key

Software engineer with computational chemistry experience

Tom King

UI/UX designer

Advisors:

Josh Schaidle, CCB Director

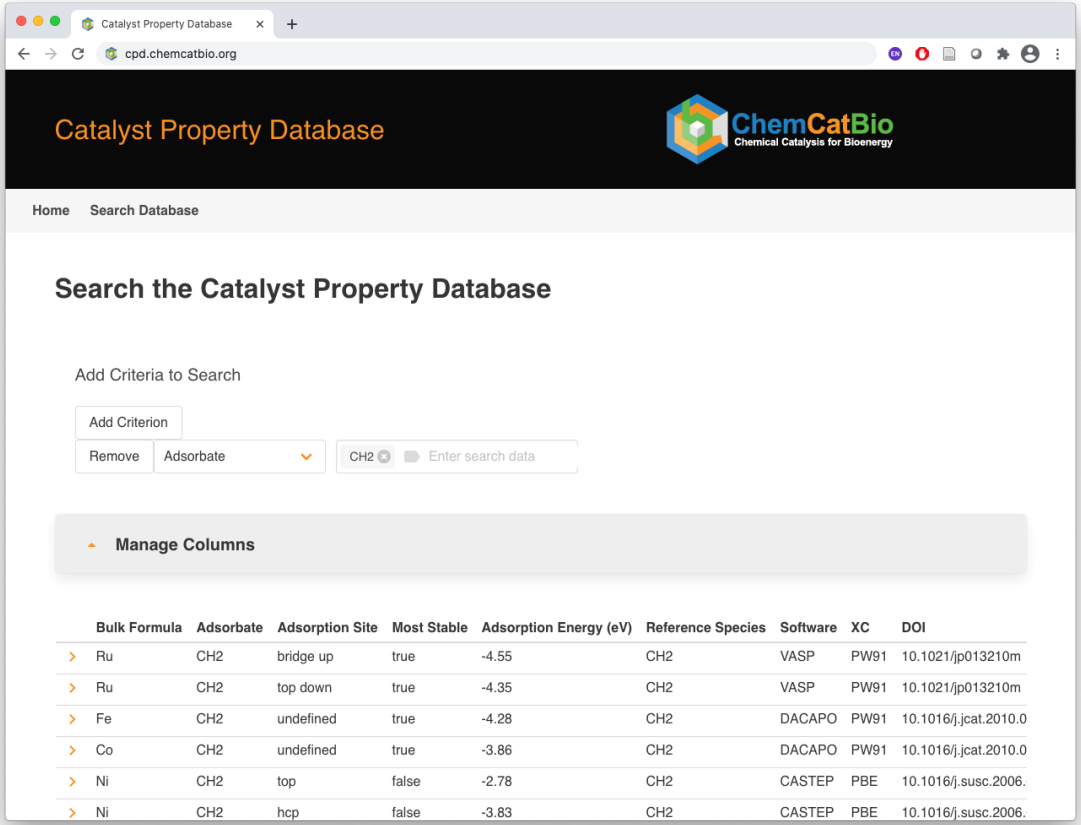
Nick Wunder, NREL

Collaborations:

CCPC
Consortium for Computational Physics and Chemistry
U.S. DEPARTMENT OF ENERGY
BIOENERGY TECHNOLOGIES OFFICE

CatCost

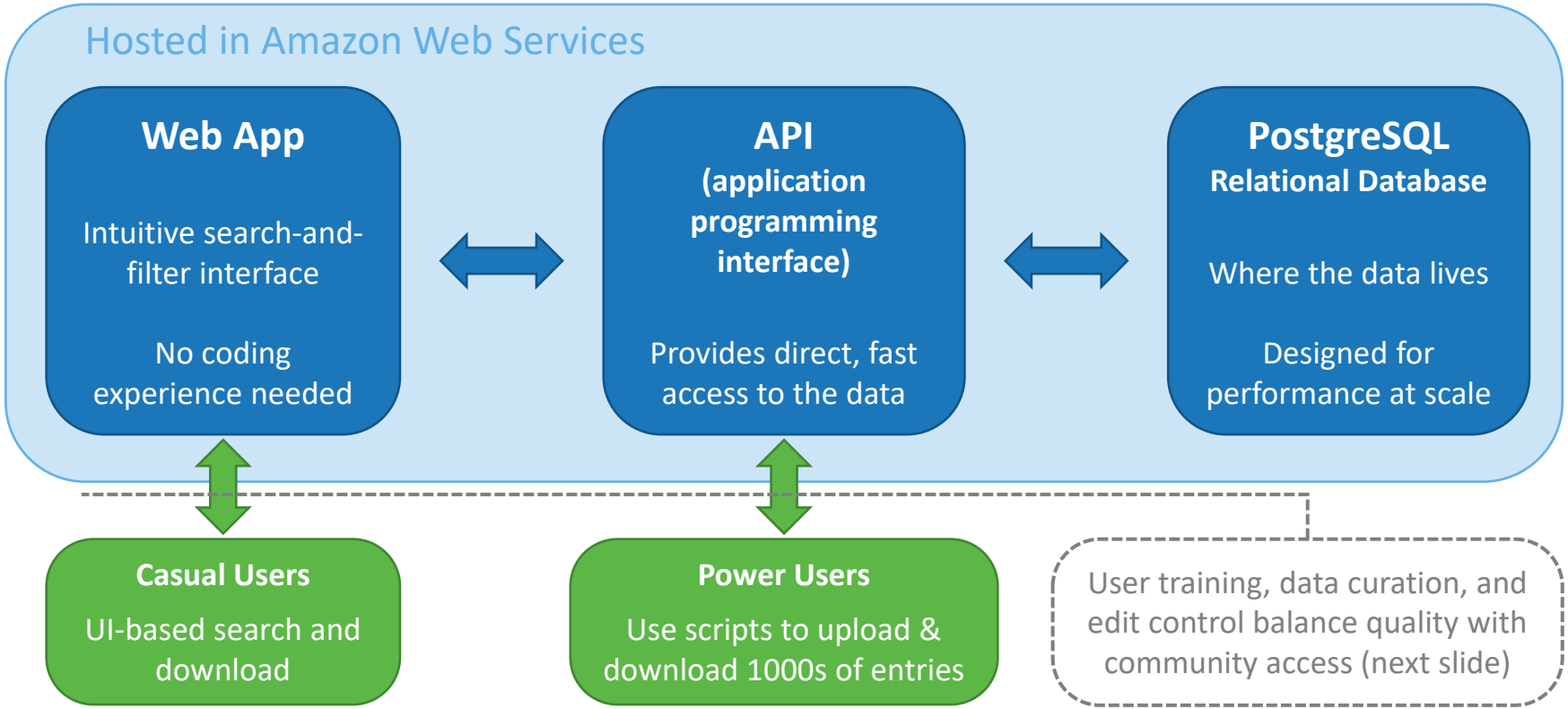
The CPD: Fast, simple searching



- Search updates live as you add criteria
- Designed for use by both computational and experimental researchers
- Simple interface

CPD Architecture: Designed for Scale

The CPD uses a modern design for fast performance as the database grows



Structuring the data

Developed a data structure for computed adsorption energies

Field	Data Type	Required Field?
Material Properties		
Formula	String	Yes
Primary Class	String	No
Secondary Class	String	No
Stretched?	Boolean	Yes, default False
Compressed?	Boolean	Yes, default False
Space Group	String	No
Lattice Constants (3)	Numeric	No
Surface/Particle Properties		
Nanoparticle Size	Numeric	No
1st Layer Composition	String	No
2nd Layer Composition	String	No
Facet	String	No
Termination	String	No
Cell Symmetry	String	No
Methods		
Software	String	Yes
Exchange correlation	String	Yes
Potentials	String	No
Basis Set	String	Yes
SpinPol?	Boolean	No
ZPE?	Boolean	No
Fixed Substrate?	Boolean	No

Field	Data Type	Required Field?
Adsorbate and Reference Species		
Adsorbate	Table	Yes
Adsorption Site	String	No
Coverage	Numeric	No
Reference Species (multiple)	Table	At least one entry required
Reference Species coefficient	Numeric	Yes
Metadata		
DOI	String	Yes
Notes	String	No
User	String	Yes
Adsorption Energy Data		
Adsorption Energy	Numeric	Yes
Most Stable?	Boolean	Yes, default True

This is key to our goal of helping users find **the right data, faster**

We'll work with users to define structures for new data types

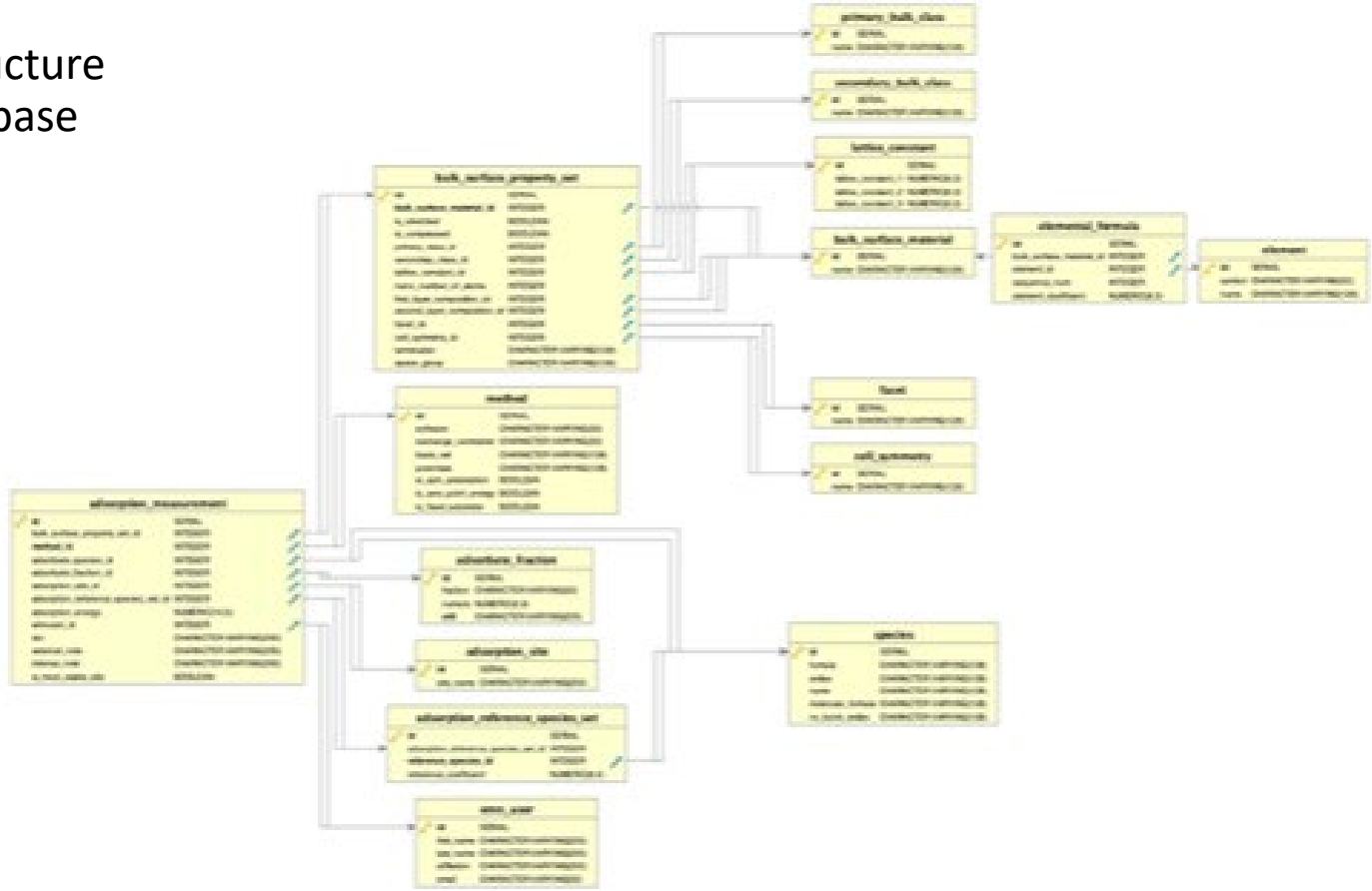
Structuring the data

Converted the data structure into a PostGreSQL database

Database ready to efficiently scale up to millions of entries

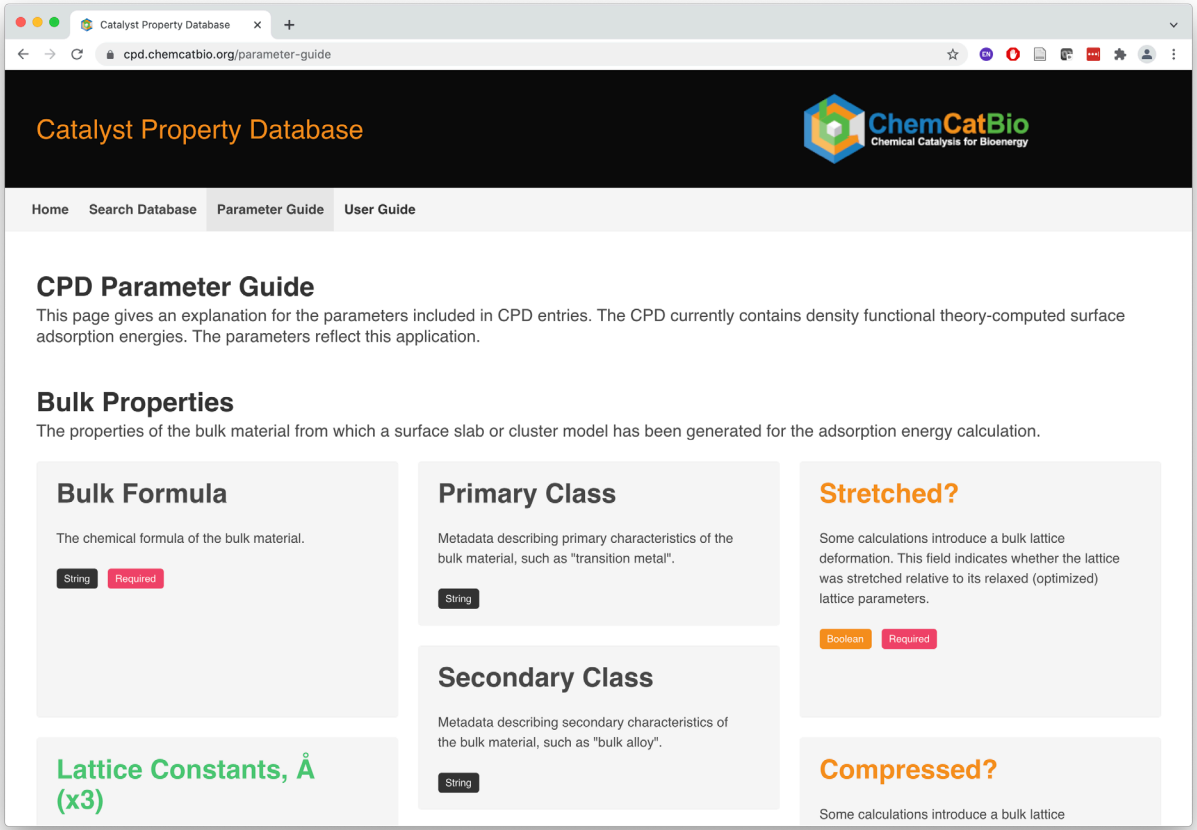
We can do this faster for new data types (characterization, synthesis, reactor performance)

1....2....*n*....



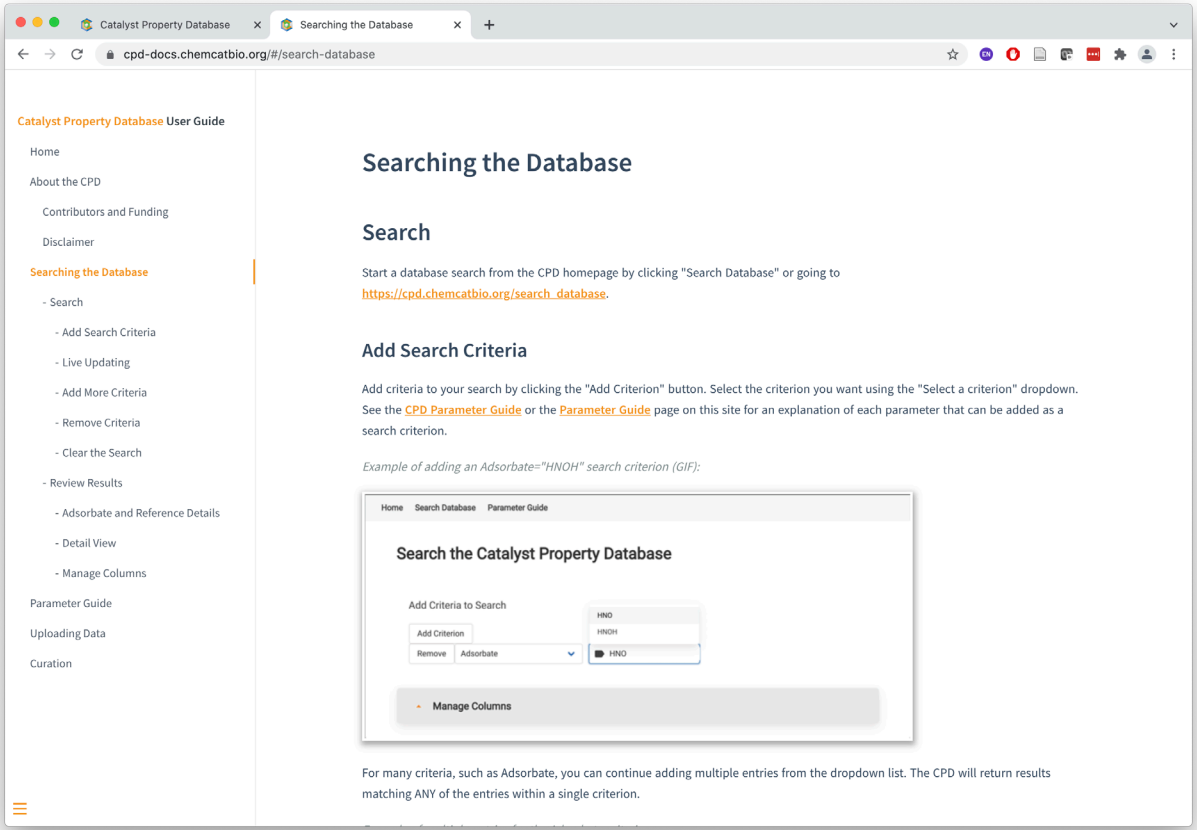
Parameter Guide for non-experts...

- Making computational chemistry results accessible to more people
- Explaining ourselves if we've been unclear (it happens)



...and a User Guide for all

- Explanation of features
- Documentation for power-users and uploaders



Standardizing data & creating rules

Created dictionaries of recurring entries, such as adsorbate/reference species, and rules for naming, data input/curation, etc.

	Molecular Formula	Name	SMILES
	~~O2 (g)~~	~~oxygen (g)~~	~~[OO]~~
	~~OO~~	~~dioxygen~~	~~OO~~
	O2	oxygen	O=O
	~~HH~~	~~hydrogen (g)~~	~~[HH]~~
	H2	hydrogen	[H][H]
	O	oxygen, atomic	[O]
	N2	nitrogen	N#N
	CH3COOH	acetic acid	CC(=O)O
	HCO	formyl	[CH]=O
	CH2	methylene	[CH2]
	COOH	carboxyl	[C](=O)O

Outcomes from dictionaries & data rules:

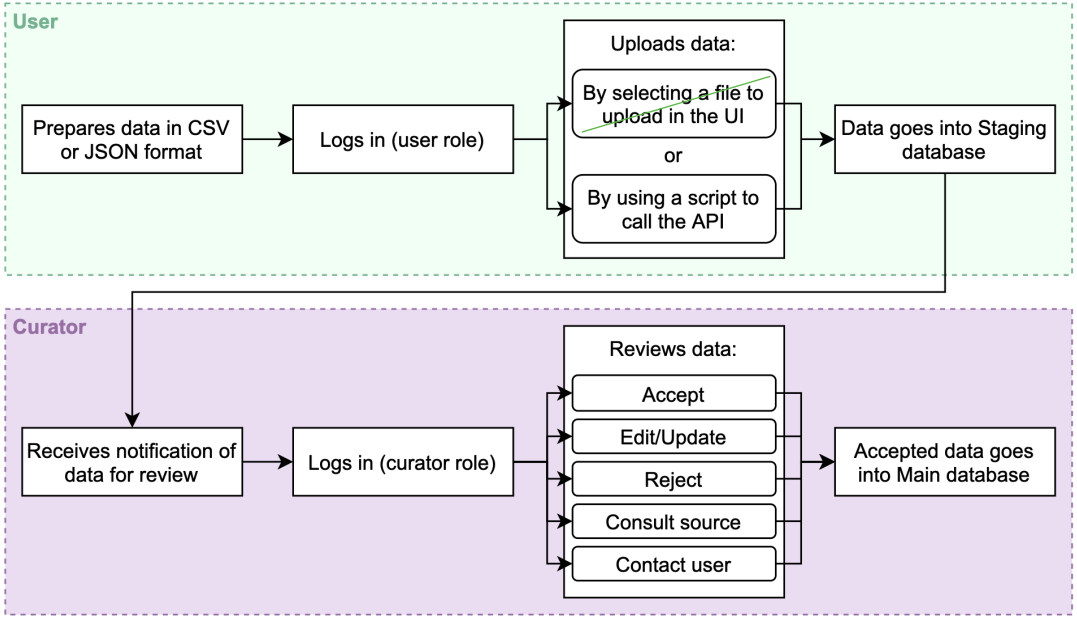
- Quality control
- Faster performance (fewer entries)
- Faster, easier search
- Gives users confidence in the data

Opening the CPD to community uploads

Sept 2021: CPD opened to uploads from the catalyst community

(previously, we added all data from the literature)

How data is uploaded and curated to ensure quality:



(still working out kinks here; it will get easier)

Hope to recruit a team of curators *a la* Wikipedia

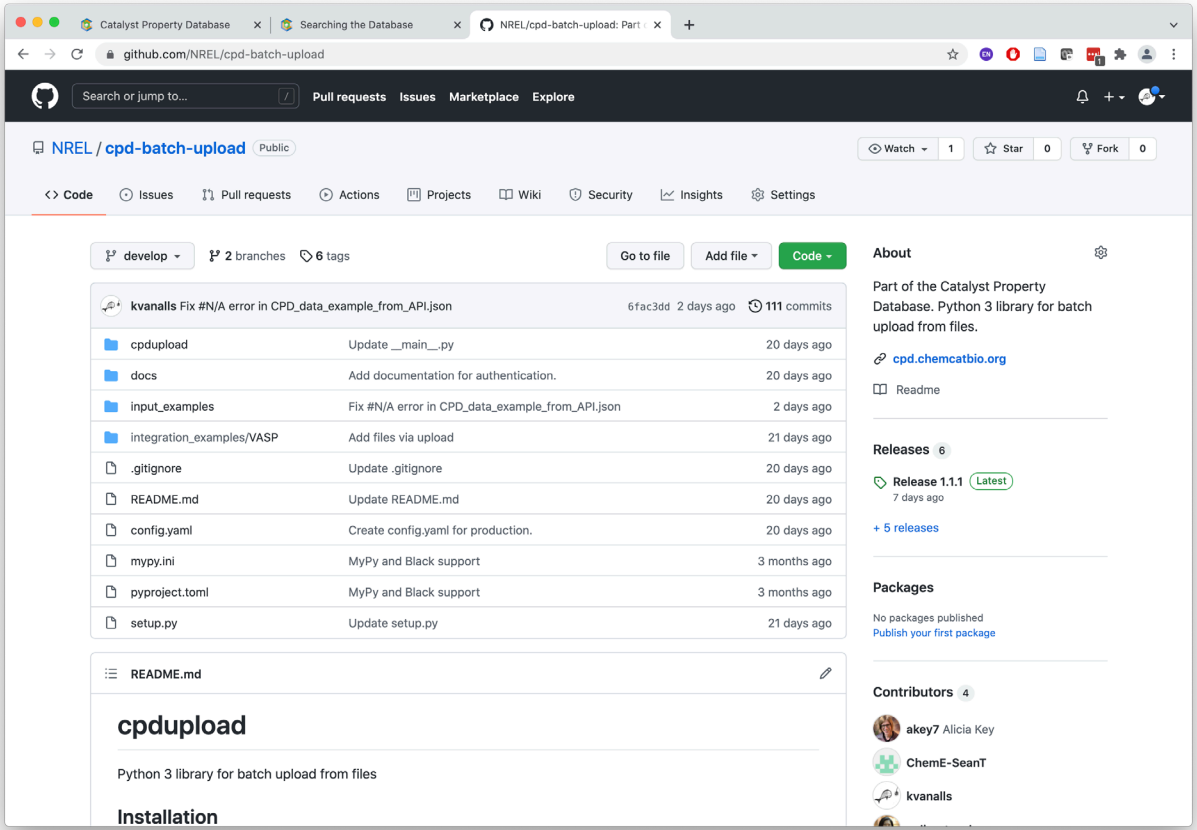
Want to upload data? (adsorption energies)

Get started by cloning this repository.

github.com/NREL/cpd-batch-upload

Reach out anytime for help!

cpd@nrel.gov

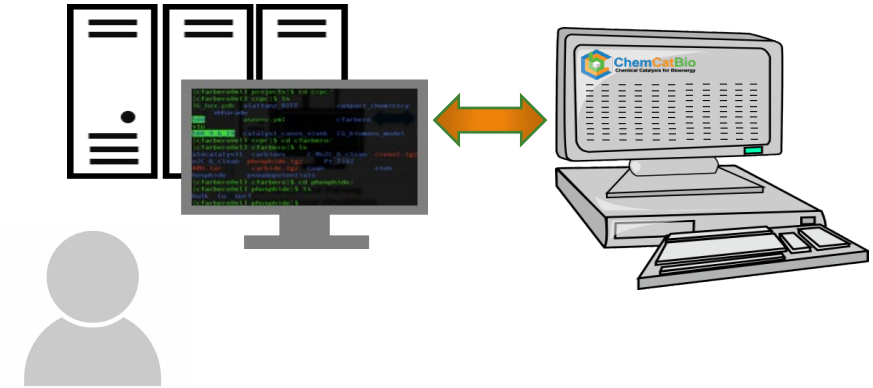


Use cases

Two use case examples

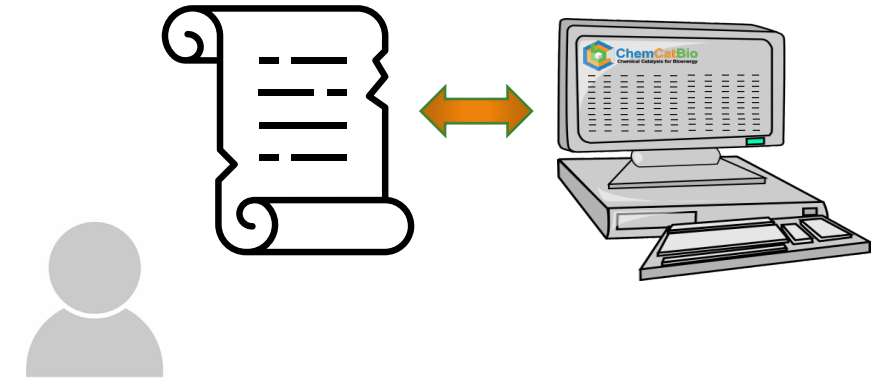
Use Case 1: Benchmarking

New computational chemistry graduate student: “leg up” on their project



Use Case 2: Benchmarking

Journal article reviewer: takes a lot of effort to do a “sanity check” on DFT papers



(we also hope editors and reviewers will encourage authors to upload their data)

Benchmarking is an important challenge facing the field

ACS Catalysis

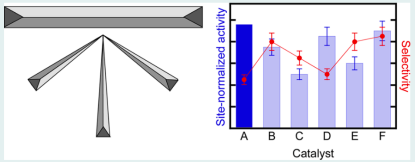
Perspective

pubs.acs.org/acscatalysis

Toward Benchmarking in Catalysis Science: Best Practices, Challenges, and Opportunities

Thomas Bligaard,[†] R. Morris Bullock,[‡] Charles T. Campbell,[§] Jingguang G. Chen,[*,∥,⊥] Bruce C. Gates,[#] Raymond J. Gorte,[g] Christopher W. Jones,[h] William D. Jones,[i] John R. Kitchin,[j] and Susannah L. Scott[*,k]

ABSTRACT: Benchmarking is a community-based and (preferably) community-driven activity involving consensus-based decisions on how to make reproducible, fair, and relevant assessments. In catalysis science, important catalyst performance metrics include activity, selectivity, and the deactivation profile, which enable comparisons between new and standard catalysts. Benchmarking also requires careful documentation, archiving, and sharing of methods and measurements, to ensure that the full value of research data can be realized. Beyond these goals, benchmarking presents unique opportunities to advance and accelerate understanding of complex reaction systems by combining and comparing experimental information from multiple, *in situ* and *operando* techniques with theoretical insights derived from calculations characterizing model systems. This Perspective describes the origins and uses of benchmarking and its applications in computational catalysis, heterogeneous catalysis, molecular catalysis, and electrocatalysis. It also discusses opportunities and challenges for future developments in these fields.



KEYWORDS: *benchmarking, catalytic performance, computational catalysis, heterogeneous catalysis, molecular catalysis, electrocatalysis*

Topics in Catalysis (2020) 63:1683–1699
https://doi.org/10.1007/s11244-020-01380-2

ORIGINAL PAPER

Towards Experimental Handbooks in Catalysis

Annette Trunschke[1] · Giulia Bellini[1] · Maxime Boniface[1] · Spencer J. Carey[1] · Jinhu Dong[1] · Ezgi Erdem[1,2] · Lucas Foppa[3] · Wiebke Frandsen[1] · Michael Geske[2] · Luca M. Ghiringhelli[3] · Frank Girgsdies[1] · Rania Hanna[1] · Maike Hashagen[1] · Michael Hävecker[1,4] · Gregory Huff[1] · Axel Knop-Gericke[1,4] · Gregor Koch[1] · Peter Kraus[1] · Jutta Kröhnert[1] · Pierre Kube[1] · Stephen Lohr[5] · Thomas Lunkenbein[1] · Liudmyla Masliuk[1] · Raoul Naumann d'Alnoncourt[2] · Toyin Omojola[1] · Christoph Pratsch[1] · Sven Richter[1] · Christian Rohner[1] · Frank Rosowski[5] · Frederik Rüther[2] · Matthias Scheffler[3] · Robert Schlögl[1,4] · Andrey Tarasov[1] · Detre Teschner[1,4] · Olaf Timpe[1] · Philipp Trunschke[6] · Yuanqing Wang[1,2] · Sabine Wrabetz[1]

Accepted: 19 September 2020 / Published online: 6 October 2020

Abstract
The "Seven Pillars" of oxidation catalysis proposed by Robert K. Grasselli represent an early example of phenomenological descriptors in the field of heterogeneous catalysis. Major advances in the theoretical description of catalytic reactions have been achieved in recent years and new catalysts are predicted today by using computational methods. To tackle the immense complexity of high-performance systems in reactions where selectivity is a major issue, analysis of scientific data by artificial intelligence and data science provides new opportunities for achieving improved understanding. Modern data analytics require data of highest quality and sufficient diversity. Existing data, however, frequently do not comply with these constraints. Therefore, new concepts of data generation and management are needed. Herein we present a basic approach in defining best practice procedures of measuring consistent data sets in heterogeneous catalysis using "handbooks". Selective oxidation of short-chain alkanes over mixed metal oxide catalysts was selected as an example.

Northwestern University

Currently in discussions that could lead to a larger collaboration:

Neil Schweitzer, Randy Snurr, Selim Alayoglu, Justin Notestein, Xijun Wang, Josh Kirkham

Reach out if you'd like to help us grow the database and this benchmarking effort!

Future updates: growing the database

- We plan to continue adding data from the literature

- We need your help – work with us and upload your results!

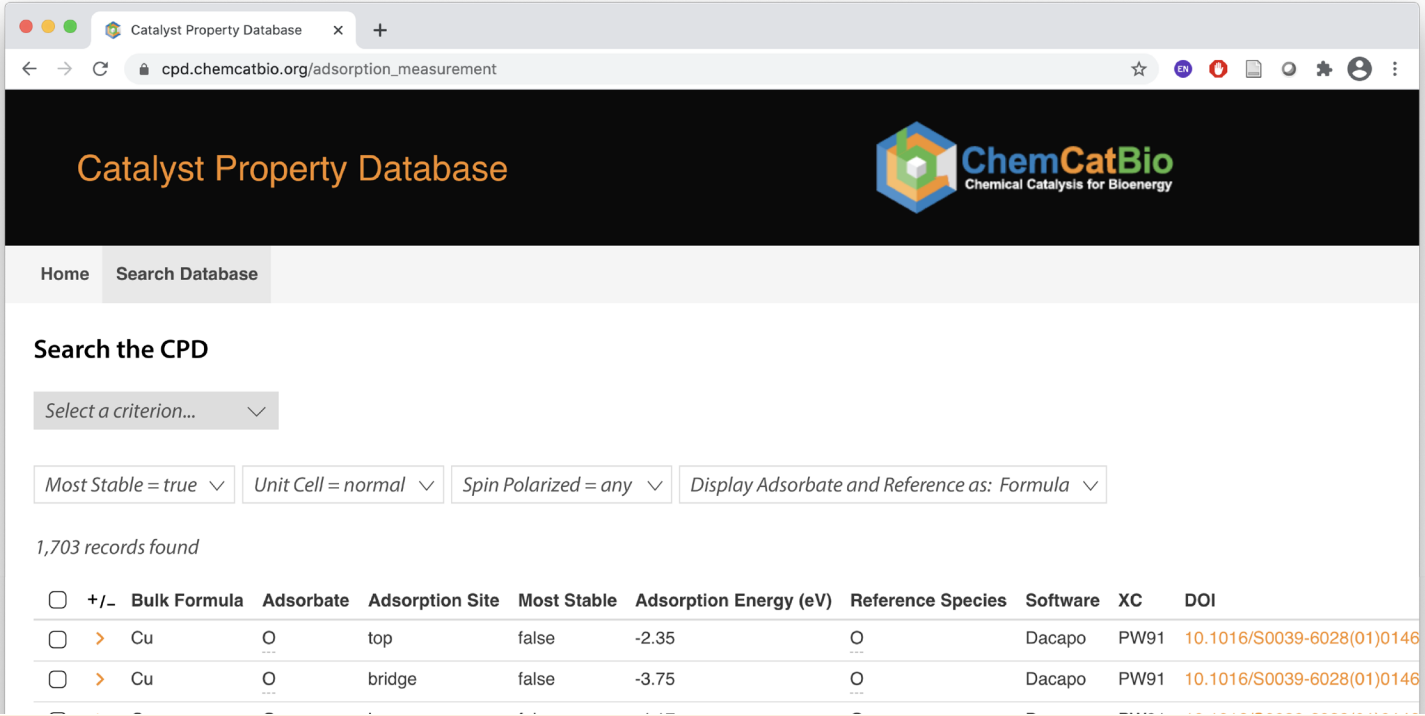
- May relax the peer-reviewed journal article requirement – what are your thoughts?

- Machine learning/natural language processing

Future updates: UI upgrades

Planned release this winter/spring

For example, adding pre-populated filters for the most common search preferences:

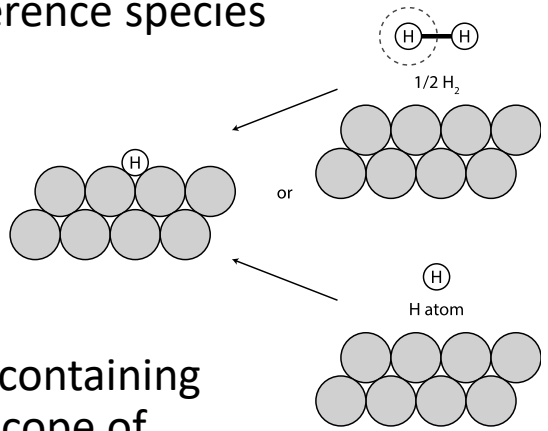


Future updates: Reference Species Translation

Background: Computed adsorption energies, the critical DFT output included in the Catalyst Property Database, may be reported with different reference species

Adsorption energy (E_{i^*}, in eV; 1 eV = 96.5 kJ/mol) for atomic H on a Pt(111) surface calculated with different gas-phase references.

Gas-phase reference	E_{i^*} (eV)	|Difference| (eV)
H	–2.80	2.26
½H_2	–0.54	



Problem: Data reported with different reference species, while containing similar information, **cannot be directly compared**, limiting the scope of data that can be applied in predictive catalyst applications

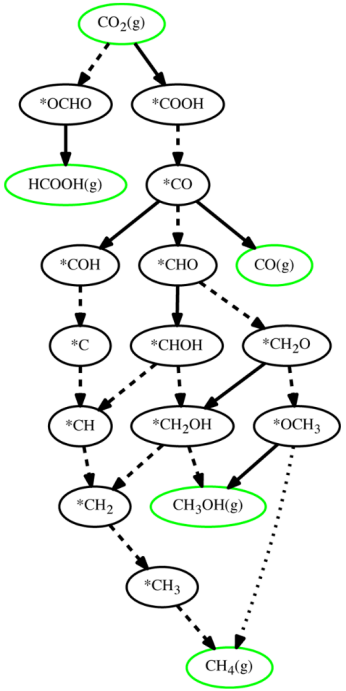
CPD Solution: Create a Reference Species Translation feature to enable interconversion between compatible reference species sets. Planned release next summer.

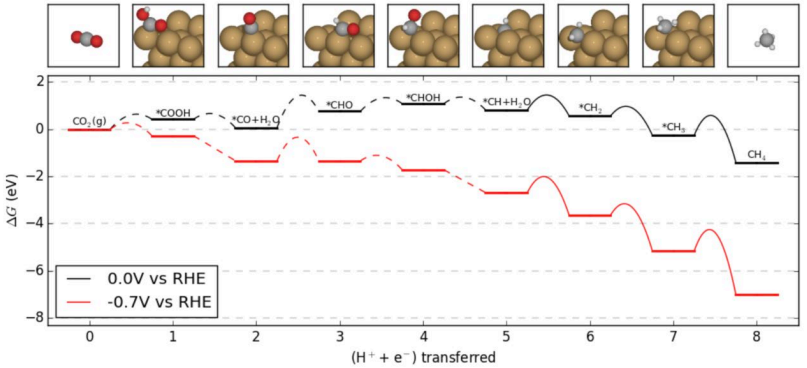
This is a key differentiator not found in any public database or resource.

Future updates: Reaction networks

Using the data generated for reference species translation, we could also generate reaction networks:

This would also allow us to cross-pollinate with databases containing reaction energies instead of adsorption energies, such as Catalysis Hub (SUNCAT)



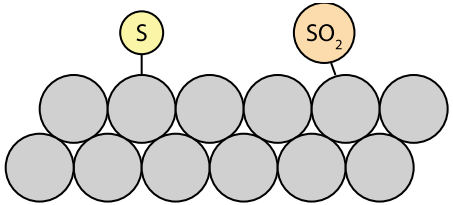


Figures from Norskov et al. *Org. Process Res. Dev.* **2016**, *20*, 1424–1430.

Future updates: Catalyst deactivation mitigation data

Create a catalyst deactivation mitigation resource in the CPD (release next fall)

- Adsorption energies for common catalyst poisons (alkali metals, halides, S, etc.) on common catalyst surfaces



Goal: enable researchers to make systematic predictions about catalyst deactivation rates

Future updates: more data types

Data types:

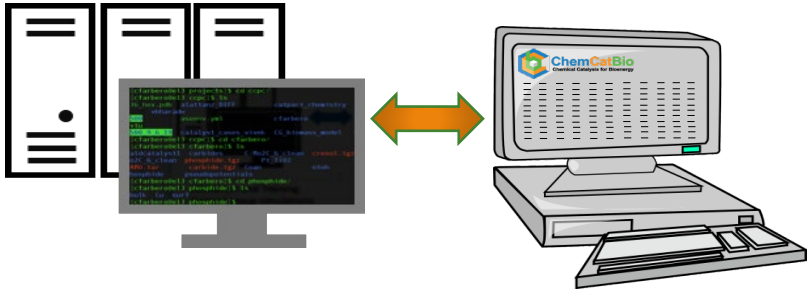
- Computational catalysis (initial focus)
- Catalyst characterization
- Reaction performance

... we need your help with this!

Applications of the CPD

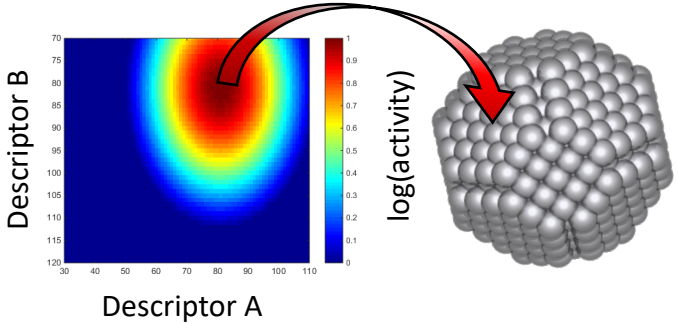
Benchmarking & Validation

Researchers wanting to check their results

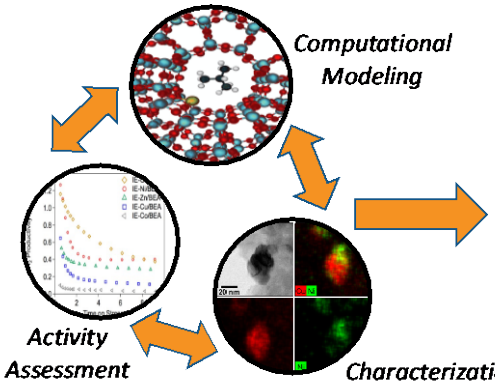


Catalyst Screening

Applying reactivity data to identify new target compositions



Interpretation of Experimental Trends

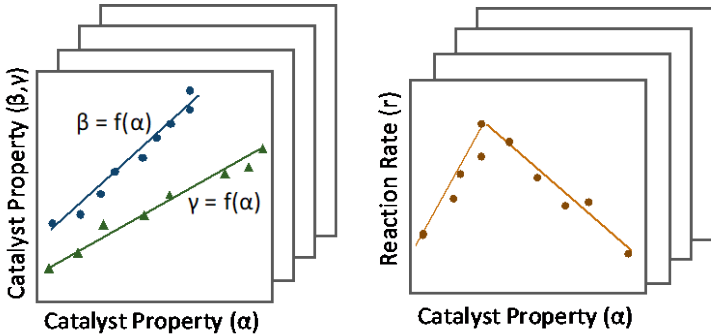


Investigate

- Active Site
- Mechanism
- Structure-Function Relationships

Discovery of Scaling Relations and Reactivity Descriptors

Harnessing the power of a large, central database



Acknowledgements

CHEM TEAM

Kurt Van Allsburg

PI

Experimentalist

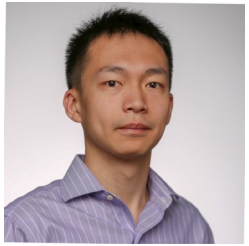
Sean Tacey

Computational researcher

Carrie Farberow

Co-investigator

Computational researcher



Qiyuan Wu

Experimentalist

Experienced web developer

DEV TEAM

Nalinrat Guba

Software engineer (previously at Oracle)

Rachel Hurst

Technical project lead, system architect

Alicia Key

Software engineer with computational chemistry experience

Tom King

UI/UX designer

Advisors:

Josh Schaidle, CCB Director

Nick Wunder, NREL

BETO:

Trevor Smith Jesse Glover

Andrea Bailey Nicole Fitzgerald

ChemCatBio Chemical Catalysis for Bioenergy

Energy Materials Network U.S. Department of Energy

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

BIOENERGY TECHNOLOGIES OFFICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308, in collaboration with the Chemical Catalysis for Bioenergy (ChemCatBio) Consortium, a member of the Energy Materials Network (EMN). Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Bioenergy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Thank you. Let’s Discuss.

Kurt Van Allsburg
kurt.vanallsburg@nrel.gov

Access the CPD: https://cpd.chemcatbio.org/

Subscribe to *The Accelerator*, a newsletter from ChemCatBio: https://public.govdelivery.com/accounts/USEERE/signup/21906

DISCLAIMER: This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Thank you!