

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





Housekeeping

- Attendees will be in listenonly 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: <u>https://www.chemcatbio.org/webin</u> <u>ars.html</u>

NOTICE: This webinar, including all audio and images of participants and presentation materials, may be recorded, saved, edited, distributed, used internally, posted on DOE's website, or otherwise made publicly available. If you continue to access this webinar and provide such audio or image content, you consent to such use by or on behalf of DOE and the Government for Government purposes and acknowledge that you will not inspect or approve, or be compensated for, such use.

Today's Speaker



Kurt Van Allsburg

Catalyst Scientist, National Renewable Energy Laboratory



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





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



ChemCatBio is accelerating the catalyst and process development cycle.

Released in 2018: CatCost



CatCost

CatCost™ is a cost estimation tool for pre-commercial catalysts.

CatCost combines industry-standard cost estimation methods and resources into an intuitive suite of tools to bring actionable cost insight to every step of catalyst research and development. It allows the rapid development of comprehensive catalyst cost estimates. It incorporates detailed insight into manufacturing methods, especially for precommercial catalysts, without requiring any process design experience.

CatCost is available in Excel spreadsheet and web app versions with user-selectable estimation methods, pricing libraries, and ease-of-use features. These features allow tailoring of an estimate to a user's experience and need for customization, empowering researchers of all skill levels to make better R&D decisions throughout the catalyst development cycle. Our costing methods are fully transparent and give power-users the ability to adjust any estimate to the economic realities of their operation.

New version! CatCost v1.1.0 released May 20, 2021

Many new features added, including process templates, sensitivity analysis, calculation and

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

ChemCatBio

Today's Webinar: the Catalyst Property Database



· Advanced search and filter logic available to help users find the data of interest, guickly

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

Catalyst Property Database Image: Catalyst Property Database Home Search Database Search the Catalyst Property Database Add Criteria to Search	Bioenergy	
Home Search Database Search the Catalyst Property Database Add Criteria In Search		
Search the Catalyst Property Database		
Add Citerion Remove Adsorbate Cite Enter search data		
· manage columns		
Buik Formula Adsorbate Adsorption Site Most Stable Adsorption Energy (eV) Reference Species Software 3	XC D	01
Ru CH2 bridge up true -4.55 CH2 VASP F	PW91 10	0.1021/jp013210m
> Ru CH2 top down true -4.35 CH2 VASP F	PW91 10	0.1021/jp013210m
> Fe CH2 undefined true -4.28 CH2 DACAPO F	PW91 10	0.1016/j.jcat.2010.0
Co CH2 undefined true -3.86 CH2 DACAPO F	PW91 10	0.1016/j.jcat.2010.0
> Ni CH2 top false -2.78 CH2 CASTEP F	PBE 10	0.1016/j.susc.2006.
NI CH2 hcp false -3.83 CH2 CASTEP F	PBE 10	0.1016/j.susc.2006.

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



Interpretation of Experimental Trends



Catalyst Screening

Applying reactivity data to identify new target compositions



Discovery of Scaling Relations and Reactivity Descriptors

Harnessing the power of a large, central database





✓ 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



Kurt Van Allsburg ΡI Experimentalist



Carrie Farberow Co-investigator Computational researcher



Sean Tacey Computational researcher



Qiyuan Wu Experimentalist Experienced web developer



Nalinrat Guba Software engineer (previously at Oracle)



Rachel Hurst Technical project lead, system architect



Alicia Key

Tom King

Software engineer with computational chemistry experience



UI/UX designer

Advisors:

Josh Schaidle, CCB Director Nick Wunder, NREL

Collaborations:



U.S. DEPARTMENT OF ENERGY **BIOENERGY TECHNOLOGIES OFFICE**

CatCost

The CPD: Fast, simple searching

 Catalyst Property Database x + Catalyst Property Database Catalyst Property Database Catalyst Property Database More Search Database Add Criteria to Search Add Criteria Bernove Attendate 	* 0				
Catalyst Property Database Home Search Database Add Criteria to Search Add Criteria Add Criteria Remove Advantate	* 0				
Catalyst Property Database Wore Search Database Add Criteria to Search Add Criteria Bernove Adjointate Enter search data					
Home Search Database Search the Catalyst Property Database Add Criteria to Search Add Criterion Berroye Adsorbate Fotor search data					
Add Criteria to Search					
	Add Criteria to Search Add Criterion Remove Adsorbate				
Manage Columns					
Bulk Formula Adsorbate Adsorption Site Most Stable Adsorption Energy (eV) Reference Species Software XC DOI					
> Ru CH2 bridge up true -4.55 CH2 VASP PW91 10.1021/jp013	10m				
> Ru CH2 top down true -4.35 CH2 VASP PW91 10.1021/jp013	10m				
> Fe CH2 undefined true -4.28 CH2 DACAPO PW91 10.1016/j.jcat.	010.0				
Co CH2 undefined true -3.86 CH2 DACAPO PW91 10.1016/j.jcat.	010.0				
Ni CH2 top false -2.78 CH2 CASTEP PBE 10.1016/j.susc	2006.				
Ni CH2 hcp false -3.83 CH2 CASTEP PBE 10.1016/j.susc					

- 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



ChemCatBio

Structuring the data

Developed a data structure for computed adsorption energies

Field	Data Type	Required Meld?
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
1 st 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
<u>Metadata</u>		
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



ChemCatBio

Parameter Guide for non-experts...

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



CPD Parameter Guide

This page gives an explanation for the parameters included in CPD entries. The CPD currently contains density functional theory-computed surface adsorption energies. The parameters reflect this application.

Bulk Properties

The properties of the bulk material from which a surface slab or cluster model has been generated for the adsorption energy calculation.



...and a User Guide for all

.

≡

Catalyst Property Database × Searching the Database

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

\leftrightarrow \rightarrow C \triangleq cpd-docs.chemcatbio.org/#/sea	rrch-database 🎓 💿 🕐 🗎 🖗		
Catalyst Property Database User Guide			
Home	Searching the Database		
About the CPD	Searching the Database		
Contributors and Funding			
Disclaimer	Search		
Searching the Database	Start a database search from the CPD homepage by clicking "Search Database" or going to		
- Search	https://cpd.chemcatbio.org/search_database.		
- Add Search Criteria			
- Live Updating	Add Search Criteria		
- Add More Criteria	Add criteria to your search by clicking the "Add Criterion" button. Select the criterion you want using the "Select a criterion" dropdown		
- Remove Criteria	See the <u>CPD Parameter Guide</u> or the <u>Parameter Guide</u> page on this site for an explanation of each parameter that can be added as a search criterion.		
- Clear the Search			
- Review Results	Example of adding an Adsorbate="HNUH" search criterion (GI-):		
- Adsorbate and Reference Details	Home Search Database Parameter Guide		
- Detail View	Search the Catalyst Broparty Database		
- Manage Columns	Search the Catalyst Property Database		
Parameter Guide	Add Criteria to Search		
Uploading Data	Add Criterion HNOH		
Curation	Remove Adsorbate V PHNO		
	Manage Columns		
	For many criteria, such as Adsorbate, you can continue adding multiple entries from the dropdown list. The CPD will return results		

× +

matching ANY of the entries within a single criterion.

☆ 🐵 🜔 📄 🕼 🔤 🌟 😩 🗄

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)	[00]
00	dioxygen	00
02	oxygen	0=0
HH	hydrogen (g)	[HH]
H2	hydrogen	[H][H]
0	oxygen, atomic	[0]
N2	nitrogen	N#N
СНЗСООН	acetic acid	CC(=O)O
НСО	formyl	[CH]=O
CH2	methylene	[CH2]
СООН	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

Catalyst Prope	ty Database X 🕏 Searching the Data	abase × O NREL/cpd-batch-uplo	ad: Part c 🗙 🕂	~
← → C 🔒 github.c	om/NREL/cpd-batch-upload			🖈 🖾 U 📄 🕼 🖬 🎘 🛎 :
Search or jump 1	o / Pull requests	Issues Marketplace Explore		Ċ +• � }•
NREL / cpd-bate Code Issues	ch-upload Public	ጦ Projects 🗆 Wiki ① Se	curity 🖂 Insights 🕸 Settings	⊘ Watch + 1 ☆ Star 0 ♥ Fork 0
۶۶ develop	→ P 2 branches ⓒ 6 tags		Go to file Add file • Code •	About ®
(e) kvanall	s Fix #N/A error in CPD_data_example_fr	om_API.json	6fac3dd 2 days ago 🕚 111 commits	Part of the Catalyst Property Database. Python 3 library for batch unload from files
🖿 cpduple	cpdupload Updatemainpy 20 days ago			
docs	docs Add documentation for authentication. 20 days ago			
input_e	mples Fix #N/A error in CPD_data_example_from_API.json 2 days ago			
🖿 integra	ion_examples/VASP Add file	es via upload	21 days ago	Releases 6
🗅 .gitigno	D .gitignore Update .gitignore 20 days ago			
🗅 READM	C README.md Update README.md 20 days		20 days ago	7 days ago
🗋 config.	Create config.yaml Create config.yaml for production. 20 days ag		20 days ago	+ 5 releases
🗋 mypy.in	mypy.ini MyPy and Black support 3 months ago			
🗅 pyproje	pyproject.toml MyPy and Black support 3 months ago		Packages	
🗋 setup.p	setup.py Update setup.py 21 days ago		No packages published Publish your first package	
	E.md		Ø	Contributors 4
cpdu	ıpload			🚯 akey7 Alicia Key
Python 3	library for batch upload from files			ChemE-SeanT
Instal	ation			

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

This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.



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 consensusbased 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

Check for updates

Towards Experimental Handbooks in Catalysis

Annette Trunschke¹ • Giulia Bellini¹ · Maxime Boniface¹ · Spencer J. Carey¹ · Jinhu Dong¹ · Ezgi Erdem^{1,2} · Lucas Foppa³ • Wiebke Frandsen¹ · Michael Geske² · Luca M. Ghiringhelli³ • Frank Girgsdies¹ • . Rania Hanna¹ · Maike Hashagen¹ · Michael Hävecker^{1,4} · Gregory Huff¹ · Axel Knop-Gericke^{1,4} · Gregor Koch¹ · Peter Kraus¹ · Jutta Kröhnert¹ · Pierre Kube¹ · Stephen Lohr⁵ · Thomas Lunkenbein¹ • . Liudmyla Masliuk¹ · Raoul Naumann d'Alnoncourt² · Toyin Omojola¹ · Christoph Pratsch¹ · Sven Richter¹ · Christian Rohner¹ · Frank Rosowski⁵ · Frederik Rüther² · Matthias Scheffler³ · Robert Schlögl^{1,4} · Andrey Tarasov¹ · Detre Teschner^{1,4} • Olaf Timpe¹ · Philipp Trunschke⁶ · Yuanging Wang^{1,2} • Sabine Wrabetz¹

Accepted: 19 September 2020 / Published online: 6 October 2020 © The Author(s) 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	<i>Ei</i> * (eV)	Difference (eV)
Н	-2.80	2.26
¹ ∕ ₂ H₂	- 0 _54	2.20



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.

H atom

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

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



Interpretation of Experimental Trends



Catalyst Screening

Applying reactivity data to identify new target compositions



Discovery of Scaling Relations and Reactivity Descriptors

Harnessing the power of a large, central database





Acknowledgements

TEAM CHEM



Kurt Van Allsburg ΡI Experimentalist



Carrie Farberow Co-investigator Computational researcher



Sean Tacey Computational researcher



Qivuan Wu Experimentalist Experienced web developer



FEAN

DEV

Nalinrat Guba

Software engineer (previously at Oracle)



Alicia Key

Software engineer with UI/UX designer computational chemistry experience



Rachel Hurst

Technical project lead, system architect



Tom King

Advisors:

Josh Schaidle, CCB Director Nick Wunder, NREL

BETO:

Trevor Smith Jesse Glover Andrea Bailey Nicole Fitzgerald





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.

35

ChemCatBio

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!





Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

BIOENERGY TECHNOLOGIES OFFICE