



BETO 2021 Peer Review: ChemCatBio Data Hub 2.6.2.500

Kurt Van Allsburg, NREL
Catalytic Upgrading Review Panel
March 9, 2021
NREL/PR-5100-79303



The Data Hub Project within ChemCatBio

Integrated and collaborative portfolio of catalytic technologies and enabling capabilities

Catalytic Technologies

Catalytic Upgrading of Biochemical Intermediates

(NREL, PNNL, ORNL, LANL)

Upgrading of C1 Building Blocks
(NREL)

Upgrading of C2 Intermediates
(PNNL, ORNL)

Catalytic Fast Pyrolysis
(NREL, PNNL)

Electrocatalytic CO₂ Utilization
(NREL)

Enabling Capabilities

Advanced Catalyst Synthesis and Characterization

(NREL, ANL, ORNL)

Consortium for Computational Physics and Chemistry

(ORNL, NREL, PNNL, ANL, NETL)

Catalyst Deactivation Mitigation for Biomass Conversion
(PNNL)

Industry Partnerships

(Phase II Directed Funding)

Opus12 (NREL)

Visolis (PNNL)

Sironix (LANL)

Cross-Cutting Support

ChemCatBio Lead Team Support (NREL)

ChemCatBio Data Hub (NREL)

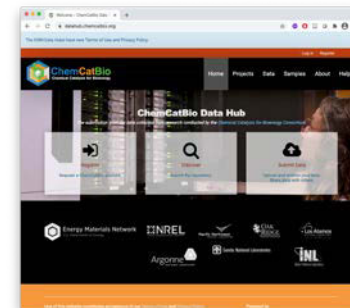
Project Overview

The Data Hub project:

Harnessing data to accelerate catalyst discovery

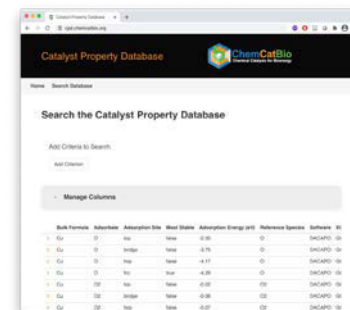
FY18–19: Develop the **Data Hub**, a general data repository

FY20–present: Develop the **Catalyst Property Database**, as part of a shift to focus on transformational tools enabling catalyst R&D



datahub.chemcatbio.org

Released in 2018



cpd.chemcatbio.org

Released in 2020

Project Overview: the Data Hub (FY18–19)

FY18–19: The Data Hub

- A requirement for ChemCatBio as an EMN consortium
- A dropbox for all kinds of scientific data, with advanced tools (e.g., visualization of common data types)
- Public and private datasets, curated by ChemCatBio

Data Hub, a framework for transformational tools enabling catalyst R&D

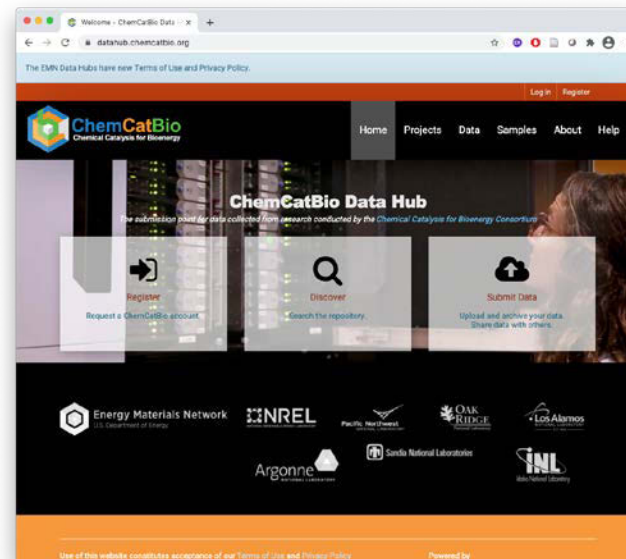
CatCost

(a free catalyst cost estimator, released in 2018)



The Catalyst Property Database

(current focus of this project)



datahub.chemcatbio.org

Released in 2018

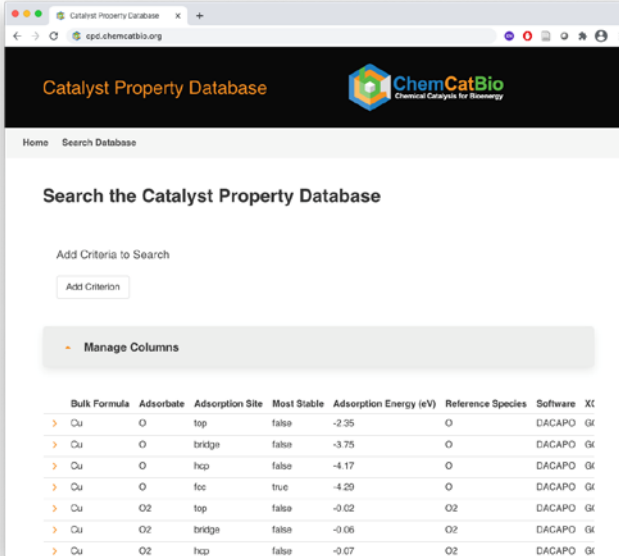
EMN = Energy Materials Network, a group of DOE-funded consortia

Project Overview: the CPD (FY20–present)

FY20–present: The **Catalyst Property Database**, a transformational tool enabling catalyst R&D

- Intended to accelerate materials discovery
- A resource for the entire catalyst community, not only ChemCatBio
- Open to public to view (FY20) and upload (FY21), subject to quality control

Informed by the vision for a **Catalyst Design Engine**



The screenshot shows the Catalyst Property Database (CPD) web interface. The page title is "Catalyst Property Database" and the URL is "cpd.chemcatbio.org". The interface includes a search bar, a "Manage Columns" button, and a table of catalyst data.

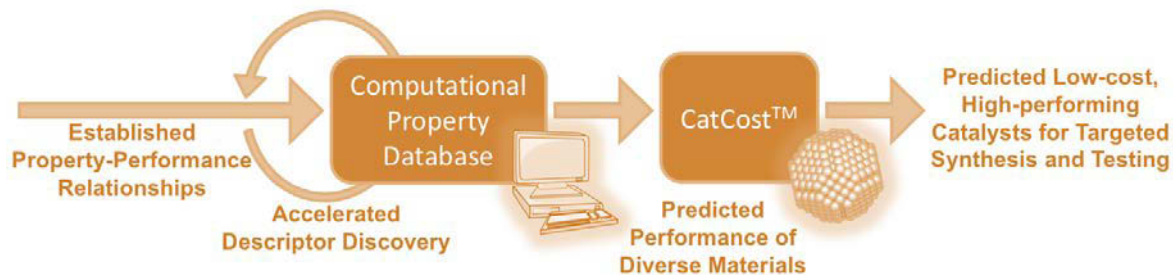
| Bulk Formula | Adsorbate | Adsorption Site | Most Stable | Adsorption Energy (eV) | Reference Species | Software | XT |
|--------------|-----------|-----------------|-------------|------------------------|-------------------|----------|----|
| > Cu | O | top | false | -2.35 | O | DACAPO | Gk |
| > Cu | O | bridge | false | -3.75 | O | DACAPO | Gk |
| > Cu | O | hcp | false | -4.17 | O | DACAPO | Gk |
| > Cu | O | fcc | true | -4.29 | O | DACAPO | Gk |
| > Cu | O2 | top | false | -0.02 | O2 | DACAPO | Gk |
| > Cu | O2 | bridge | false | -0.06 | O2 | DACAPO | Gk |
| > Cu | O2 | hcp | false | -0.07 | O2 | DACAPO | Gk |

cpd.chemcatbio.org

Released in 2020

The Catalyst Design Engine

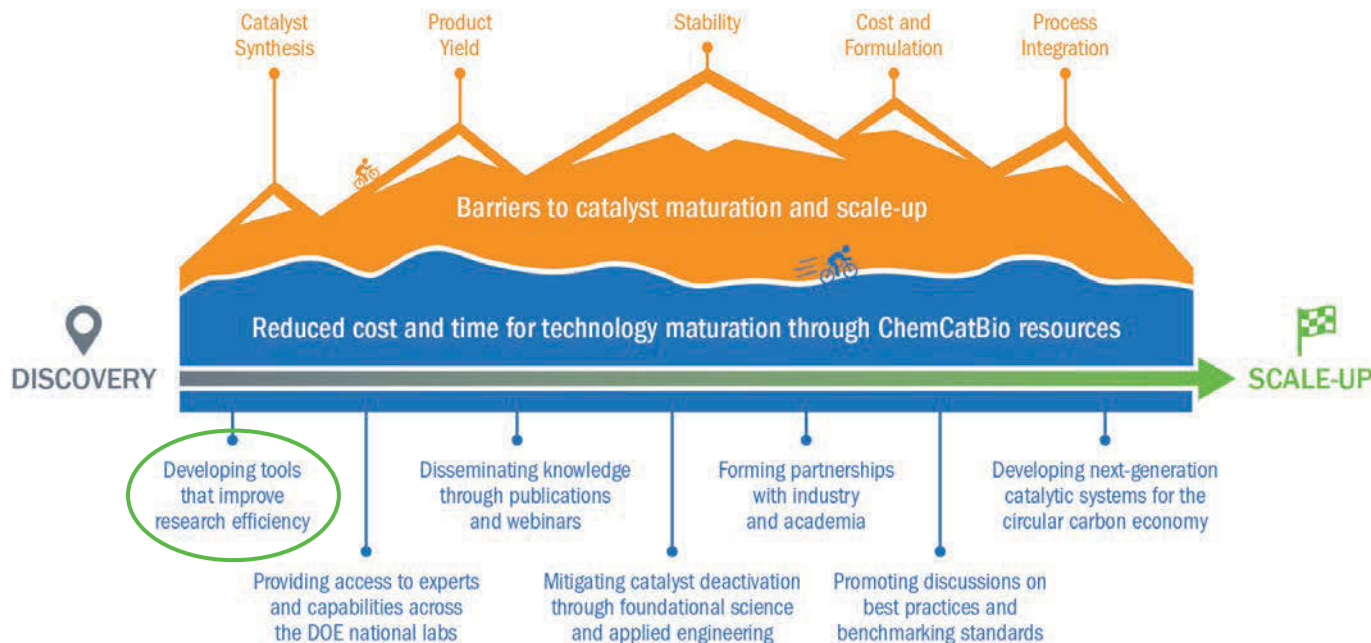
The **Catalyst Design Engine** is a vision for a free, publicly available tool that combines catalyst performance and cost insight



To work toward this vision, we first need to develop the building blocks: CatCost (released 2018) and the Catalyst Property Database (in development, initial release 2020)

Accelerating Catalyst Discovery

The path to catalyst deployment is slow and difficult.


















ChemCatBio is accelerating the catalyst and process development cycle.

Project Overview: Heilmeier Questions

- What are you trying to do?
Harnessing catalyst property data to accelerate catalyst discovery
- How is it done today and what are the limits?
When data is used in catalyst discovery, it is **often collected/computed from scratch** because of the challenges in finding **reliable, directly comparable** datasets
- Why is it important?
It's inefficient.
Current practice results in **redundant calculations** being repeatedly performed
Furthermore, only a **small subset** of published data is brought to bear
- What are the risks?
Users find the database irrelevant or too difficult to use

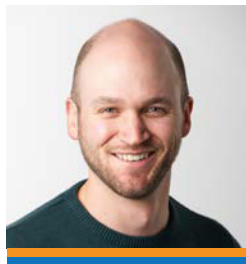
Market Trends Addressed by the Data Hub

| | |
|-----------------------|---|
| Product |  Gasoline/ethanol demand decreasing, diesel demand steady |
| |  Increasing demand for aviation and marine fuel |
| |  Demand for higher-performance products |
| |  Increasing demand for renewable/recyclable materials |
| Feedstock |  Sustained low oil prices |
| |  Decreasing cost of renewable electricity |
| |  Sustainable waste management |
| |  Expanding availability of green H ₂ |
| |  Closing the carbon cycle |
| |  Risk of greenfield investments |
| Capital |  Challenges and costs of biorefinery start-up |
| |  Availability of depreciated and underutilized capital equipment |
| |  Carbon intensity reduction |
| Social Responsibility |  Access to clean air and water |
| |  Environmental equity |

As a cross-cutting effort to enable accelerated catalyst discovery, the ChemCatBio Data Hub is most relevant to those market trends calling for new, advanced catalysts

1. Management: Team and Collaborations

Management Plan: A project team with diverse, targeted expertise



Kurt Van Allsburg, Ph.D.
PI
Experimentalist
Experienced developer of R&D tools such as CatCost



Nalinrat Guba, Ph.D.
Lead Developer
Software engineer
(previously at Oracle)



Qiyuan Wu, Ph.D.
Experimentalist
Experienced web developer



Carrie Farberow, Ph.D.
Technical expert,
former PI
Computational
researcher



Sean Tacey, Ph.D.
Technical expert
Computational
researcher

Advisors:

Josh Schaidle, CCB Director
Tom King, NREL, UI designer
[Nick Wunder](#), NREL, web dev expert
[Kathy Cisar](#), NREL Communications

Project Tasks:

- **Task 1: Data Hub and Computational Catalyst Property Database Development**
- **Task 2: Data Hub Maintenance, Security, and Oversight**

Project Collaborations:

- CCPC – Atomistic Modeling Task, (PI: Carrie Farberow)
- CatCost (PI: Fred Baddour)



CatCost

1. Management: Tasks

Project Tasks:

- **Task 1: Data Hub and Catalyst Property Database Development (\$325,000)**
 - New milestone-related features in Catalyst Property Database (CPD)
 - Bug fixes and usability improvements in CPD and Data Hub
 - Research demonstrations of CPD
 - Adding new datasets to CPD
 - External user interviews, user training & documentation, webinars/screencasts, quality control
- **Task 2: Data Hub Maintenance, Security, and Oversight (\$25,000)**
 - Security upgrades to Data Hub and Catalyst Property Database
 - Managing site hosting with Amazon Web Services

Tracked & prioritized
using Agile methods

1. Management: Risk Mitigation

Risk 1: Development Scope & Schedule Creep

Mitigation: Use Agile (Scrum) software development systems & practices

Risk 2: Database Does Not Match User Needs (irrelevant / wrong features)

Mitigation: Go/No-Go milestone (FY21Q2) focuses on seeking expert / potential user feedback on development direction & pitfalls

Risk 3: Database Is Too Difficult To Use or Does Not Justify Required Effort

Mitigation: Ongoing input from catalyst researchers as part of Go/No-Go, NREL technical software developers & UI experts, and ChemCatBio stakeholders

Risk 4: Data Quality/Quantity Issues

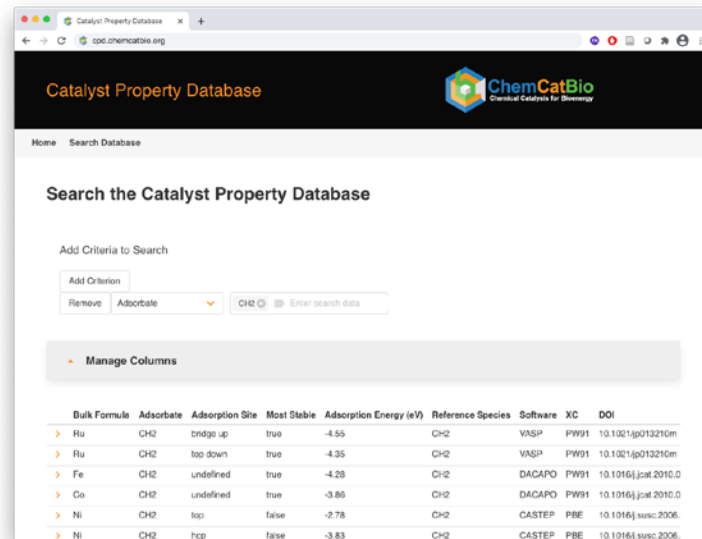
Mitigation:
Quality: Establish curation & training (FY21Q3)
Quantity: Use partnerships with CCPC, CCB, external interviewees to gain buy-in & users, leverage “prestige” factor for data uploads

2. Approach: the Catalyst Property Database

In FY20–22, the Data Hub project is focused on developing the **Catalyst Property Database** as a **collaboration and discovery tool**

The Catalyst Property Database (CPD)

- A centralized, searchable repository of catalyst properties
- Publicly accessible to view and upload
 - Uploads subject to quality control
- Initial release: DFT-computed adsorption energies for intermediates on catalyst surfaces



The screenshot shows the Catalyst Property Database (CPD) web interface. The page has a dark header with the ChemCatBio logo and the text 'Catalyst Property Database'. Below the header, there is a search bar and a table of catalyst properties. The table has columns for Bulk Formula, Adsorbate, Adsorption Site, Most Stable, Adsorption Energy (eV), Reference Species, Software, XC, and DOI. The table contains several rows of data, including Ru, Fe, Co, Ni, and Ni catalysts with various adsorption sites and energies.

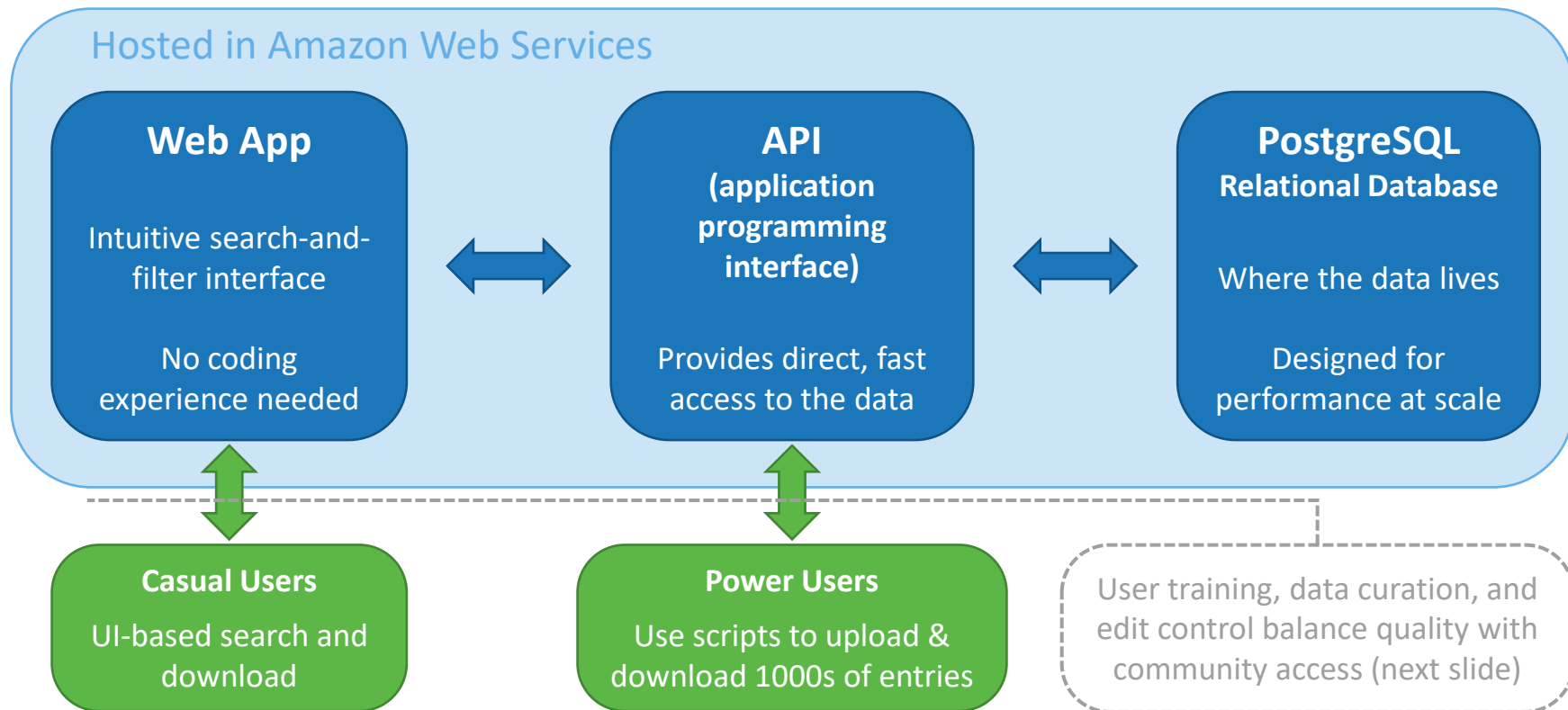
| Bulk Formula | Adsorbate | Adsorption Site | Most Stable | Adsorption Energy (eV) | Reference Species | Software | XC | DOI |
|--------------|-----------|-----------------|-------------|------------------------|-------------------|----------|------|-----------------------|
| > Ru | CH2 | bridge-up | true | -4.55 | CH2 | VASP | PW91 | 10.1002/ijp013210m |
| > Ru | CH2 | top-down | true | -4.35 | CH2 | VASP | PW91 | 10.1002/ijp013210m |
| > Fe | CH2 | undefined | true | -4.28 | CH2 | DACAPO | PW91 | 10.1016/j.jcat.2010.0 |
| > Co | CH2 | undefined | true | -3.86 | CH2 | DACAPO | PW91 | 10.1016/j.jcat.2010.0 |
| > 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.2006. |

DFT = density functional theory

Note: The CPD was previously called the *Computational* Catalyst Property Database

2. Approach: CPD Architecture

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



2. Approach: Data Growth & Curation

Problem:

Achieving sustainable growth for the CPD while maintaining high quality

CPD Solution:

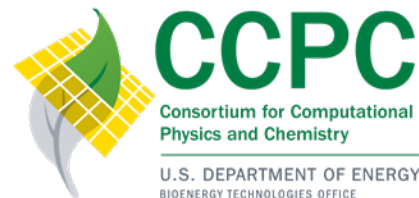
Open the database to public contributions, but use robust quality controls

Data Growth:

- Work with our partners to encourage large uploads

Quality Controls:

- Strict data requirements built into database
- Initial data curated by Data Hub team
- Long-term: use a Wikipedia model, but with qualified & trained curator-editors

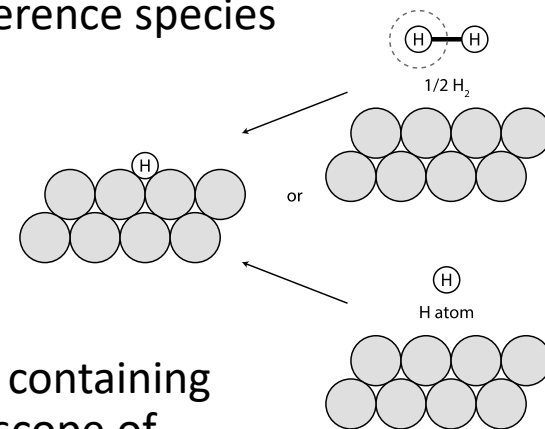


2. Approach: Reference Species Translation Feature

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

Adsorption energy (E_{r}^* , 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_{r}^* (eV) | Difference (eV) |
|-------------------------|-----------------------|------------------|
| H | -2.80 | 2.26 |
| $\frac{1}{2}\text{H}_2$ | -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.

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

2. Approach: Development Plan

Foundational

Transformational

FY20: Release CPD as a read-only resource

FY21: Focus on external users & enable uploads

FY21 Go/No-Go:
Interview external experts

FY22: Demo a research advancement enabled by CPD

FY21 Go/No-Go:

Interview 10+ experts to confirm that CPD development plans align with the needs of potential users and adjust as necessary.

Focus areas for interviews, to support CPD innovation:

- Preferred features
- Experience with competing solutions
- Strengths-Weaknesses-Opportunities-Threats
- Preferences for UI, scripting, etc.

3. Impact: Faster & Cheaper Catalyst Discovery

2019 Peer Reviewer Comment:

“We see considerable duplication of effort in catalysis research”

Every year, **more experimental and computational catalyst data is generated**, but the methods and **tools to apply this data have not kept up**.

The CPD is advancing the state of the art for application of computational data:

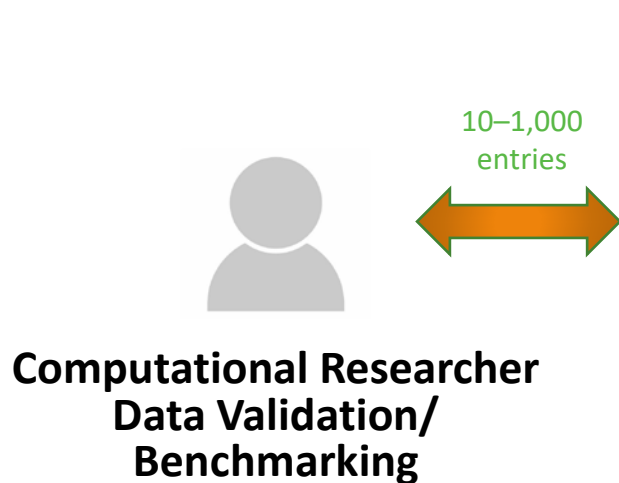
- Stop spending time and money **re-creating data** that already existed
- Enable **new approaches** to catalyst discovery that require large datasets

By harnessing the power of data, the Data Hub and the Catalyst Property Database are accelerating catalyst discovery



3. Impact: Diverse R&D Applications of the CPD

The Catalyst Property Database has diverse R&D applications, starting now and expanding as the database grows:



cpd.chemcatbio.org

| Bulk Formula | Adsorbate | Adsorption Site | Most Stable | Adsorption Energy (eV) | Reference Species | Software | XC | DOI |
|--------------|-----------------|-----------------|-------------|------------------------|-------------------|----------|------|-------------------|
| Ru | CH ₄ | top-down | true | -4.55 | CH ₄ | WASP | PW91 | 10.1021/gp101327n |
| Ta | CH ₄ | unbridged | true | -4.28 | CH ₄ | DACAPD | PW91 | 10.1019/gp101327n |
| Ce | CH ₄ | unbridged | true | 3.85 | CH ₄ | DACAPD | PW91 | 10.1019/gp101327n |
| Ni | CH ₄ | top | false | -0.78 | CH ₄ | CATSEP | PBE | 10.1019/gp101327n |
| Ni | CH ₄ | top | false | -3.83 | CH ₄ | CATSEP | PBE | 10.1019/gp101327n |



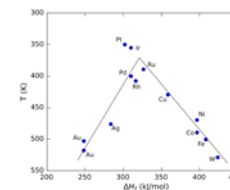
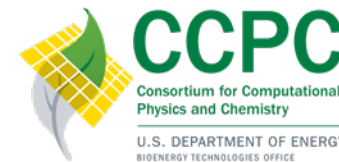
**Catalysis Machine
Learning**

1,000–100,000
entries



**High-Throughput
Calculations**

1,000–100,000
entries



Reactivity Descriptor Discovery

DFT = density functional theory

4. Progress: Create Database Data Structure

FY20: Developed the data structure for computed adsorption energies

| Field | Data Type | Required Field? |
|---|------------------|------------------------|
| <u>Material Properties</u> | | |
| 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 |
| <u>Surface/Particle Properties</u> | | |
| Nanoparticle Size | Numeric | No |
| 1st Layer Composition | String | No |
| 2nd Layer Composition | String | No |
| Facet | String | No |
| Termination | String | No |
| Cell Symmetry | String | No |
| <u>Methods</u> | | |
| 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? |
|---|------------------|-----------------------------|
| <u>Adsorbate and Reference Species</u> | | |
| 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 |
| <u>Adsorption Energy Data</u> | | |
| Adsorption Energy | Numeric | Yes |
| Most Stable? | Boolean | Yes, default True |

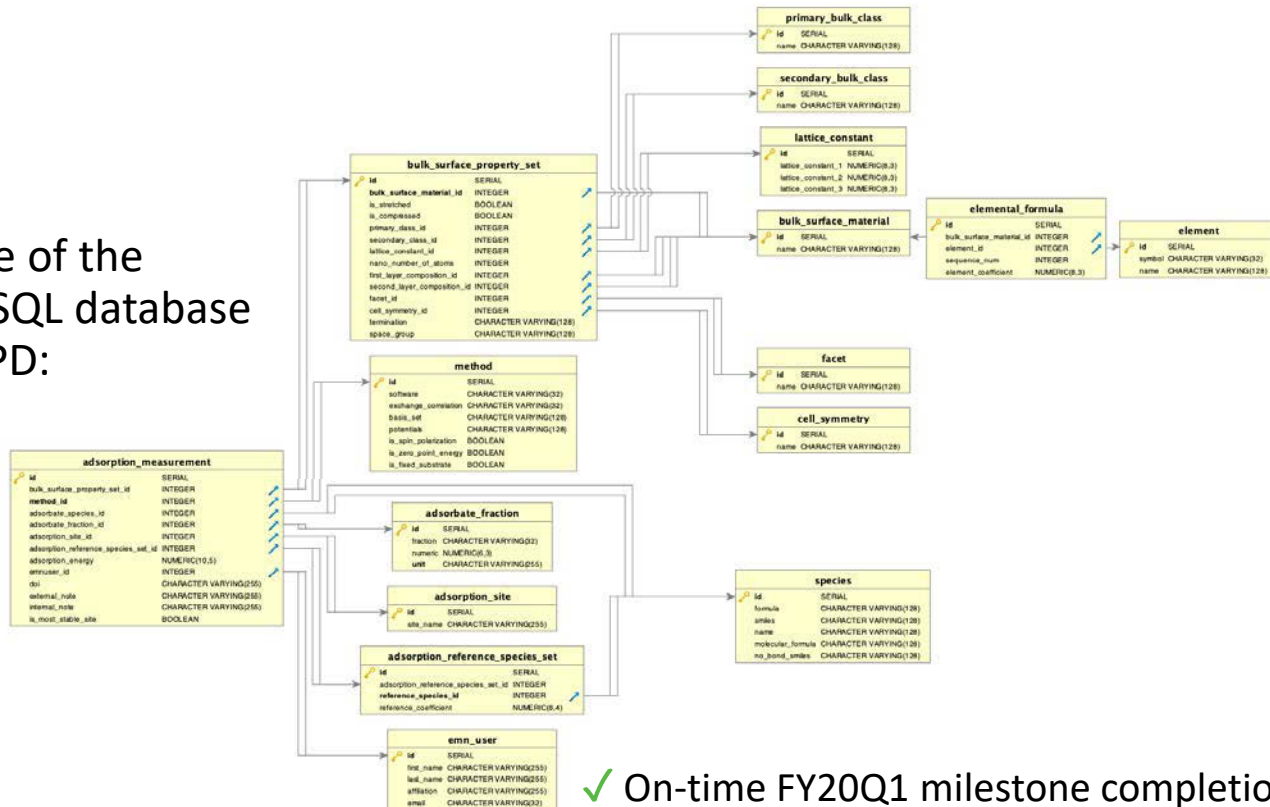
Outcome: This structure and the associated rules are essential to data quality control & curation

✓ On-time FY20Q1 milestone completion

4. Progress: Create Database Data Structure

FY20: Converted the data structure into a PostgreSQL database

Structure of the PostgreSQL database in the CPD:



Outcome: Database ready to efficiently scale up to millions of entries

✓ On-time FY20Q1 milestone completion

4. Progress: Standardize Data & Create Rules

FY20: 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
- Faster, easier search
- Gives users confidence in the data

✓ On-time FY20Q1 milestone completion

4. Progress: Create User Interface

FY20: Developed the UI for the Catalyst Property Database, including a live-updating search/filter approach:

Outcome:

The clean, modern interface allows catalyst researchers of all backgrounds to access valuable data

Search the Catalyst Property Database

Add Criteria to Search

Add Criterion

Remove Adsorbate

Enter search data

| | Bulk Formula | Adsorbate | Adsorption | Adsorption Energy (eV) | Reference Species | Software | XC | |
|---|--------------|-----------|------------|------------------------|-------------------|----------|----------|----------|
| > | Cu | O | top | | O | DACAPO | GGA-PW91 | |
| > | Cu | O | bridge | | O | DACAPO | GGA-PW91 | |
| > | Cu | O | hcp | false | -4.17 | O | DACAPO | GGA-PW91 |
| > | Cu | O | fcc | true | -4.29 | O | DACAPO | GGA-PW91 |
| > | Cu | O2 | top | false | -0.02 | O2 | DACAPO | GGA-PW91 |
| > | Cu | O2 | bridge | false | -0.06 | O2 | DACAPO | GGA-PW91 |
| > | Cu | O2 | hcp | false | -0.07 | O2 | DACAPO | GGA-PW91 |

✓ On-time FY20Q3 milestone completion

4. Progress: Initial Public Release

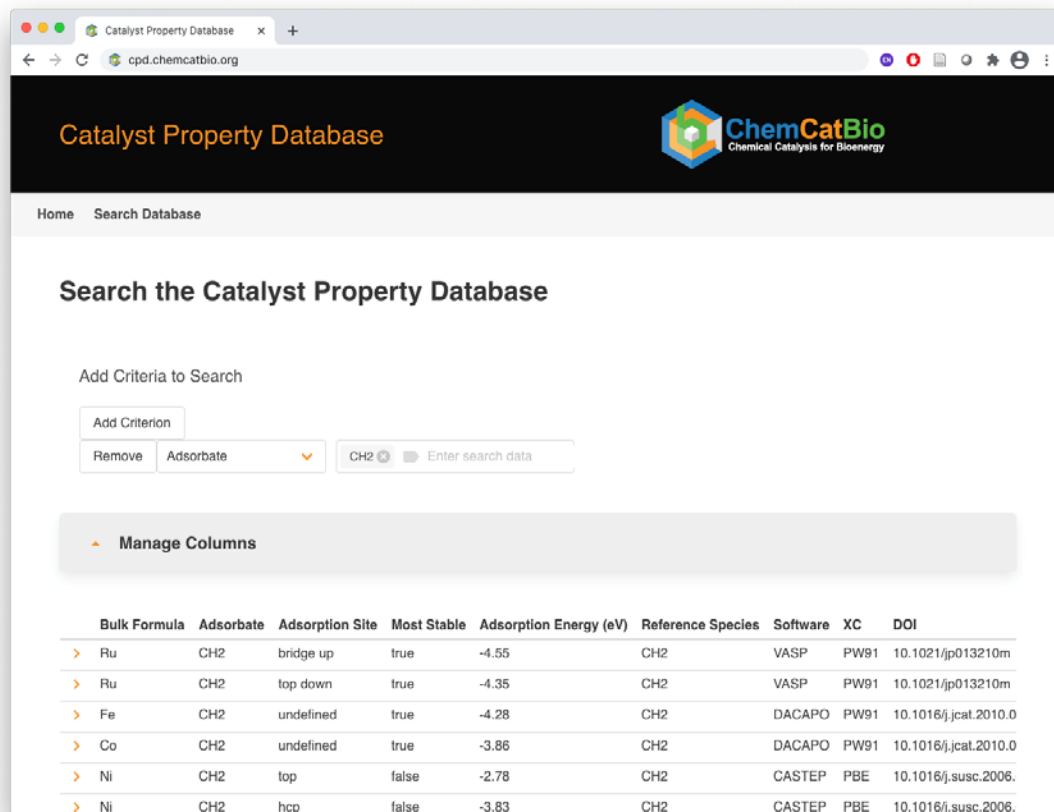
FY20: Released the Catalyst Property Database to the public in Sept 2020

cpd.chemcatbio.org

Outcome:

The CPD is now available for users to test and provide feedback.

✓ On-time FY20Q4 milestone completion



The screenshot shows the Catalyst Property Database website interface. The browser address bar displays "cpd.chemcatbio.org". The page header includes the site name and the ChemCatBio logo. Below the header, there is a search bar with the criteria "Adsorbate" set to "CH2". A "Manage Columns" button is visible above a table of search results.

| | 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/jp013210m |
| > | Ru | CH2 | top down | true | -4.35 | CH2 | VASP | PW91 | 10.1021/jp013210m |
| > | Fe | CH2 | undefined | true | -4.28 | CH2 | DACAPO | PW91 | 10.1016/j.jcat.2010.0 |
| > | Co | CH2 | undefined | true | -3.86 | CH2 | DACAPO | PW91 | 10.1016/j.jcat.2010.0 |
| > | 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.2006. |

4. Progress: Ongoing Development

FY21: Added a Manage Columns dialog to allow users to select columns for display...

Manage Columns

Bulk Formula Adsorbate Adsorption Site Most Stable Adsorption Energy (eV) Reference Species Software XC DOI

Primary Class Secondary Class First Layer Second Layer Space Group Facet Termination Cell Symmetry Potential

Basis Set Coverage Stretched Compressed Spin Pol ZPE Fixed Substrate Nanoparticle Size

... and a “detail view” to show all available metadata for a row

Cu O hcp false -4.17 O DACAPO GGA-PW91 10.1016/S0039-6028(0

Adsorption Measurement Detail

| Material Properties | Surface/Particle Properties | Methods | Adsorbate Species | Adsorption Reference Species | Adsorption Metadata | Adsorption E |
|---|--|---|---|--|--|-----------------------------|
| Formula: Cu Primary Class: transition metal Secondary Class: Stretched: false Compressed: false Space Group: Fm3m Lattice Constant 1: 3.660 Lattice Constant 2: Lattice Constant 3: | Nanoparticle Size: 1st Layer Composition: 2nd Layer Composition: Facet: {111} Termination: Cell Symmetry: 2X2 | Software: DACAPO Exchange Correlation: GGA-PW91 Potentials: Basis Set: SpinPol: ZPE: Fixed Substrate: false | Name: oxygen, atomic SMILES: [O] Formula: O NoBondSMILES: [O] Molecular Formula: O Adsorption Site: hcp Coverage: 1/4 | Overall Formula: O Adsorbate 1 Coefficient: 1.0000 Name: oxygen, atomic SMILES: [O] Formula: O NoBondSMILES: [O] Molecular Formula: O | DOI: 10.1016/S0039-6028(01)01464-9 User: Tuong Bui Email: Tuong.Bui@nrel.gov Affiliation: External Note: Internal Note: | Adsorption I MostStable: |

4. Progress: Reference Species Translation

FY21: Created a detailed development plan for this feature, including required data

| Reference species interconversion framework | Example: Atomic hydrogen (H) adsorption on Pt(111) Changing reference from H to molecular hydrogen (H ₂). | Example: Acetic acid adsorption on Pt(111) Changing reference from acetic acid to CO ₂ , O ₂ , and H ₂ . | Required data/changes for CCPD |
|--|---|--|---|
| <p>Step 1: Balance the stoichiometric coefficients (ν_i) for reaction of new gas-phase reference species (n.r.) to produce adsorbed species.</p> | <p>Step 1: Balance equation (* denotes adsorption site/species): $\nu_{H_2}H_2 + * \rightarrow H^*$ $\nu_{H_2} = 0.5$</p> | <p>Step 1: Balance equation $\nu_{CO_2}CO_2 + \nu_{O_2}O_2 + \nu_{H_2}H_2 + * \rightarrow CH_3COOH^*$ $\nu_{CO_2} = 2, \nu_{O_2} = -1, \nu_{H_2} = 2$</p> | Manual input field or automated script to balance stoichiometric coefficients |
| <p>Step 2: Calculate the base-reference energy ($E_{b,r}$) using stored stoichiometric coefficients (ν_j) and energies ($E_{g,b,k}$) for each b.r. species j.</p> $E_{b,r} = \sum \nu_j E_{g,b,k_j}$ | <p>Step 2: Base reference, H(g) $E_{g,b,H} = -1.12$ eV $\nu_H = 1$ $E_{b,r} = 1 * (-1.12$ eV) = -1.12 eV</p> | <p>Step 2: Base reference: Acetic acid(g); $E_{g,b,acetic\ acid} = -46.79$ eV $\nu_{acetic\ acid} = 1$ $E_{b,r} = 1 * (-46.79$ eV) = -46.79 eV</p> | Gas-phase energies for base reference |
| <p>Step 3: Calculate the new-reference energy ($E_{n,r}$) using calculated stoichiometric coefficients (Step 1, ν_i) and energies ($E_{g,b,k}$) for each n.r. species k.</p> $E_{n,r} = \sum \nu_i E_{g,b,k}$ | <p>Step 3: New reference, 0.5H₂ $E_{g,b,H_2} = -6.76$ eV $E_{n,r} = 0.5 * (-6.76$ eV) = 3.38 eV</p> | <p>Step 3: New reference: 2CO₂ + -1O₂ + 2H₂ $E_{g,b,CO_2} = -22.99$ eV; $E_{g,b,O_2} = -9.45$ eV; $E_{g,b,H_2} = -6.76$ eV $E_{n,r} = 2 * (-22.99$ eV) - 1 * (-9.45 eV) + 2 * (-6.76 eV) = 50.05 eV</p> | Gas-phase energies for new reference |
| <p>Step 4: Recalculate adsorption energy using $E_{b,r}$ and $E_{n,r}$ from Steps 2 and 3, respectively.</p> $E_p(n,r) = E_p(b,r) + E_{b,r} - E_{n,r}$ | <p>Step 4: $E_p(b,r) = -2.80$ eV $E_p(n,r) = -2.80$ eV + (-1.12 eV) - (-3.38 eV) $E_p(n,r) = -0.54$ eV</p> | <p>Step 4: $E_p(b,r) = -0.80$ eV $E_p(n,r) = -0.80$ eV + (-46.79 eV) - (-50.05 eV) $E_p(n,r) = +2.46$ eV</p> | Script/API to perform calculation |
| <p>Step 5: Output/display the data.</p> | | | Additional columns to CCPD output table |

Outcome: This key differentiator is complex to implement, so we broke it down into achievable development steps.

✓ On-time FY21Q1 milestone completion

Future Work

Foundational

Transformational



Upcoming work in the Data Hub project:

FY21Q2: Go/No-Go – Interviews with external experts; guide development direction

FY21Q3: Develop training, documentation, and curation procedures

FY21Q4: Add batch upload capability, including examples (e.g., Python script)

FY22: Demonstrate a research advancement enabled by the unique data-sharing capabilities of the CPD. Time savings of 2–10X compared to best alternative methods.

Quad Chart

Timeline

- Project start date: 10/1/2019
- Project end date: 09/30/2022

| | FY21 | FY20–FY22 |
|-------------|--------|-----------|
| DOE Funding | \$350k | \$1.05M |

Barriers addressed

- Ct-F – Increasing the yield from catalytic processes
- Ct-G – Decreasing the time and cost to develop novel industrially relevant catalysts

Funding Mechanism

AOP

Project Goal

Enable ChemCatBio and the bioenergy industry to accelerate the catalyst and process development cycle through development of publicly available advanced analytics tools. Demonstrate application of the Computational Catalyst Property Database (CPD) in a predictive design application that is dramatically accelerated (2–10X) by the use of the CPD's unique capabilities.

End of Project Milestone

Demonstrate a research advancement enabled by the unique data-sharing capabilities of the CPD. Pilot and document a research use case of the CPD that results in a time savings of 2–10X compared to the best alternative methods. Perform this demonstration with a suitable partner, such as a high-throughput computational research group.

Acknowledgements

Data Hub Team

Carrie Farberow

Nalinrat Guba

Sean Tacey

Matt Jankousky

Tuong Bui

Nick Wunder

Kris Munch

Courtney Pailing

Josh Schaidle

Advisors

Kathy Cisar

Fred Baddour

Tom King

BETO

Trevor Smith

Jesse Glover

Andrea Bailey

Nicole Fitzgerald

This research was supported by the DOE Bioenergy Technology Office under Contract no. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory

This work was performed in collaboration with the Chemical Catalysis for Bioenergy Consortium (ChemCatBio, CCB), a member of the Energy Materials Network (EMN)

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.



Energy Materials Network
U.S. Department of Energy

Thank you!



Data Hub: Informed by 2019 Peer Review

“We see considerable duplication of effort in catalysis research”

- The Catalyst Property Database aims to make data easier to find and reuse, cutting down on redundant computations/measurements

Should allow contributions from non-CCB researchers

- The CPD is designed for contributions by the broader catalysis community

“It would be nice to see...a more universal repository of the sort that NIST curates”

- The newly-renamed Catalyst Property Database is designed to work toward this vision

Should use Agile project management principles

- We are using Agile/Scrum in our development

User feedback and confirmation of usefulness is essential

- We have made external user feedback the core of our Go/No-Go milestone