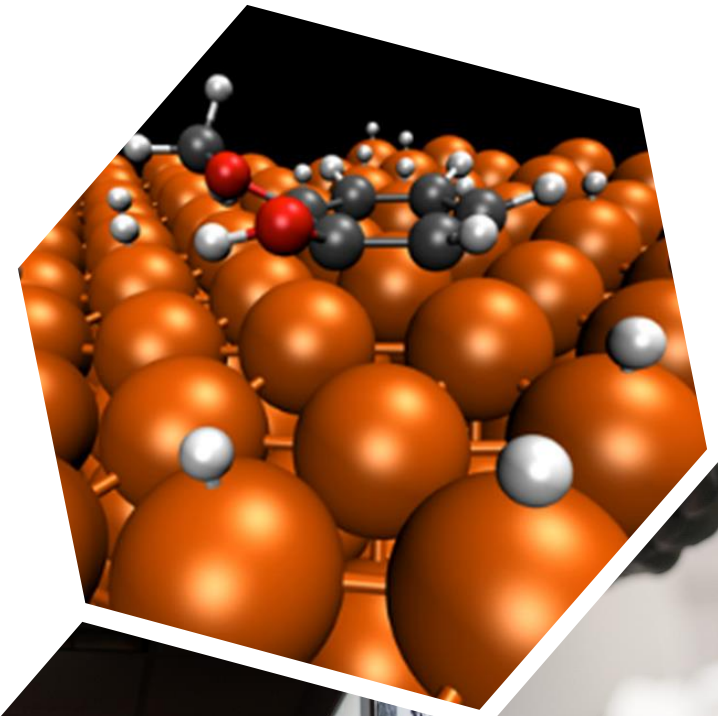
ChemCatBio
Chemical Catalysis for Bioenergy



ChemCatBio DataHub

WBS: 2.6.2.500

U.S. Department of Energy (DOE)
Bioenergy Technologies Office (BETO)
2019 Project Peer Review

Carrie Farberow

March 4th, 2019

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

ChemCatBio Foundation

Integrated and collaborative portfolio of catalytic technologies and enabling capabilities

Catalytic Technologies

- **Catalytic Upgrading of Biochemical Intermediates** (NREL, PNNL, ORNL, LANL, NREL*)
- **Catalytic Upgrading of Indirect Liquefaction Intermediates** (NREL, PNNL, ORNL)
- **Catalytic Fast Pyrolysis** (NREL, PNNL)
- **Electrocatalytic and Thermocatalytic CO_2 Utilization** (NREL, ORNL*)

*FY19 Seed Project

Enabling Capabilities

- **Advanced Catalyst Synthesis and Characterization** (NREL, ANL, ORNL, SNL)
- **Catalyst Cost Model Development** (NREL, PNNL)
- **Consortium for Computational Physics and Chemistry** (ORNL, NREL, PNNL, ANL, NETL)
- **Catalyst Deactivation Mitigation for Biomass Conversion** (PNNL)

Industry Partnerships (Directed Funding)

- **Gevo** (NREL)
- **ALD Nano/JM** (NREL)
- **Vertimass** (ORNL)
- **Opus12**(NREL)
- **Visolis** (PNNL)
- **Lanzatech** (PNNL) - Fuel
- **Gevo** (LANL)
- **Lanzatech** (PNNL) - TPA
- **Sironix** (LANL)

Cross-Cutting Support

ChemCatBio Lead Team Support (NREL)

ChemCatBio DataHUB (NREL)

Goal Statement and Outcomes

Goal

Enable **ChemCatBio** and the **bioenergy industry** to **accelerate** the **catalyst** and **process development** cycle by establishing the DataHub for:

1. **Centralized** and **secure** data **storage, sharing, and analysis**
2. Development and application of publicly available advanced analytics tools that provide **predictive capabilities** for catalyst R&D

Outcomes

- Data repository for **easy** and **fast storage and sharing** of data **amongst project teams** and, as desired, with the **public** for **knowledge dissemination**
- **Customized plug-ins** and **data analysis tools** to **increase research efficiency**
- **Integrated catalyst design engine** that accelerates the catalyst development process by **optimizing for predicted performance and material cost**

Relevance to Bioenergy Industry

- **Pathway-independent tools** developed **reduce** the **time and cost** to develop active, selective, low-cost and durable **real-world catalysts** to produce **desired end-products from biomass and waste resources**

Quad Chart Overview

Timeline

- Project start date: 10/1/2017
- Project end date: 9/30/2020
- Percent complete: 50%

	Total Costs Pre FY17**	FY 17 Costs	FY 18 Costs	Total Planned Funding (FY 19-Project End Date)
DOE Funded	$0	$0	$200k	$400k

Barriers addressed

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

Objective

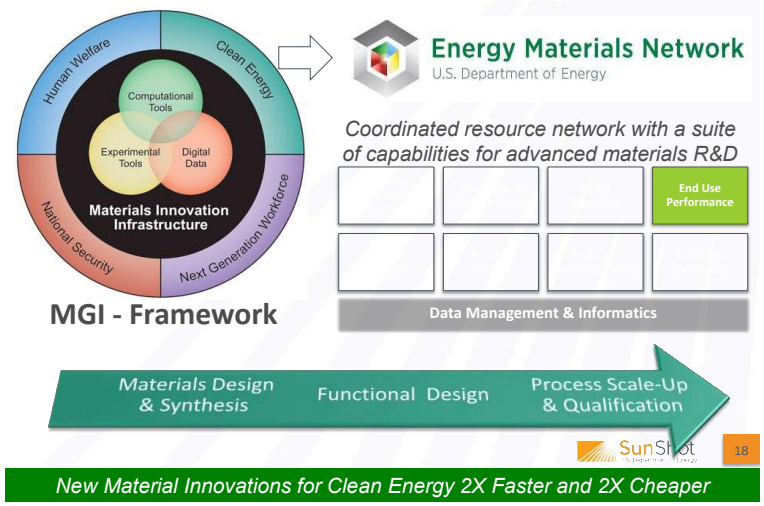
Develop and provide access to **accelerated data storage, sharing, and advanced analytics tools** to **reduce** the **time** and **cost** to **develop catalysts** for the production of desired end-products from biomass and waste resources.

End of Project Goal

Produce a **data repository** alongside **customized plug-ins and data tools** that are **accessible** to all ChemCatBio project teams and industry partners, including completion of the infrastructure of an **integrated catalyst design engine** that simultaneously **considers predicted performance and catalyst material cost** to accelerate the catalyst design cycle.

Overview: The ChemCatBio DataHub

In support of the Materials Genome Initiative, each DOE EMN consortium is tasked with providing an accessible, searchable data resource.

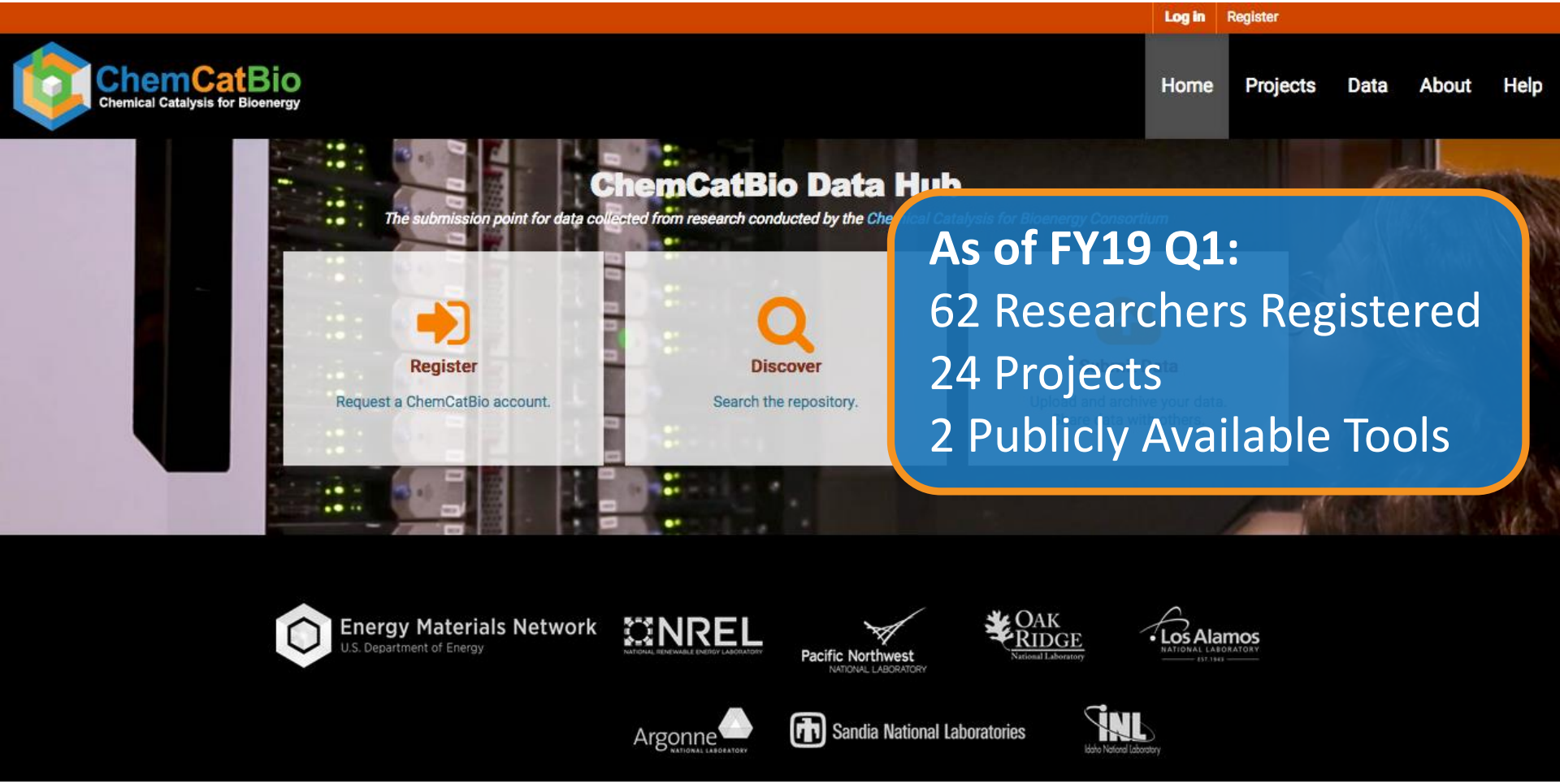


Key Project Outcomes

- Provides a **secure** project **space for storing and sharing project data**
  - ➢ *Outcome: Faster, less burdensome cross-institution collaboration*
- Enables **controlled public accessibility** of selected datasets
  - ➢ *Outcome: Distribution of knowledge and reduction in research redundancy*
- Facilitates development of and access to **advanced data analysis tools**
  - ➢ *Outcome: Application of state-of-the-art predictive approaches that accelerate catalyst R&D*

Overview: The ChemCatBio DataHub

The ChemCatBio DataHub is publically accessible at *datahub.chemcatbio.org*

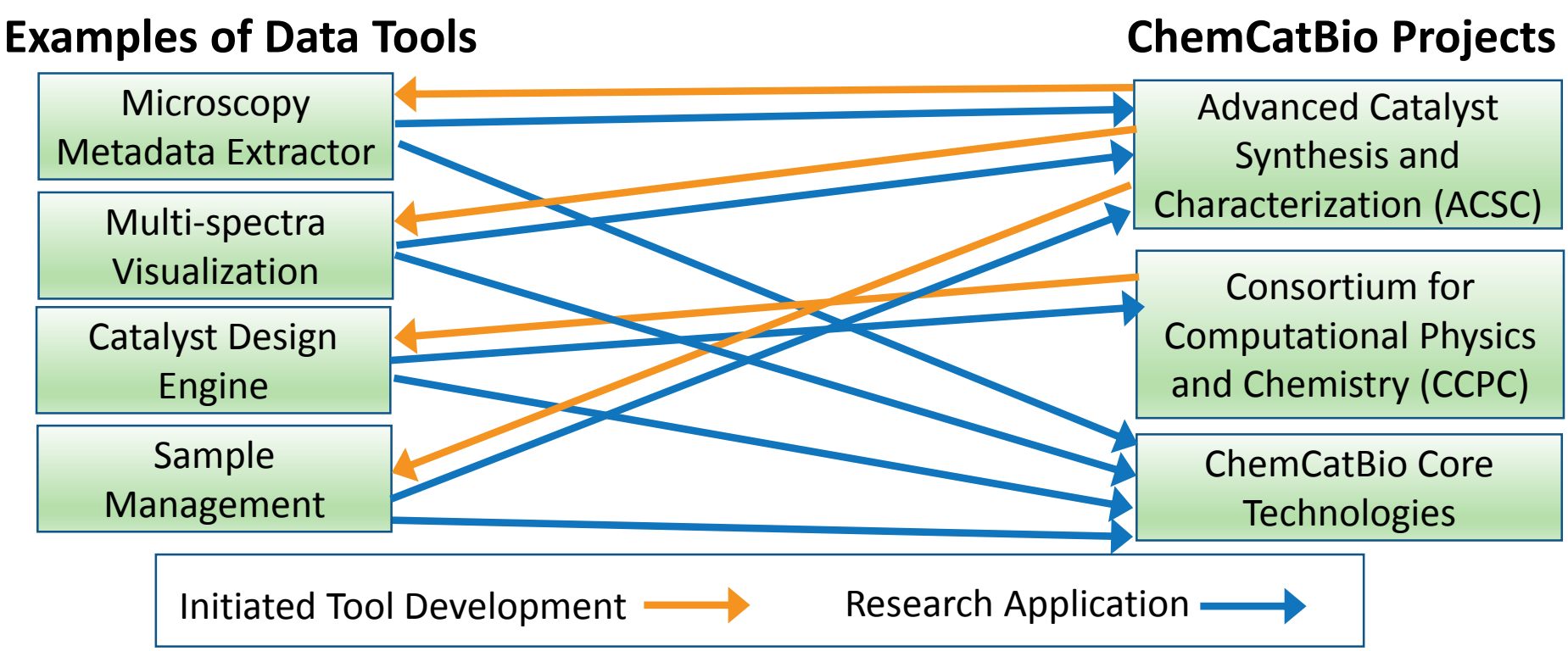


Management Approach

*Develop tools responsive to researcher needs through **researcher-centric approach***

Task 1: Build DataHub Infrastructure and Tools
- Build DataHub framework
- Develop plug-ins for storing, searching, and processing
- Develop advanced analytics tools

Task 2: Interface with Researchers
- Solicit regular researcher feedback to guide tool development
- Ensure deliverables are responsive to researcher needs

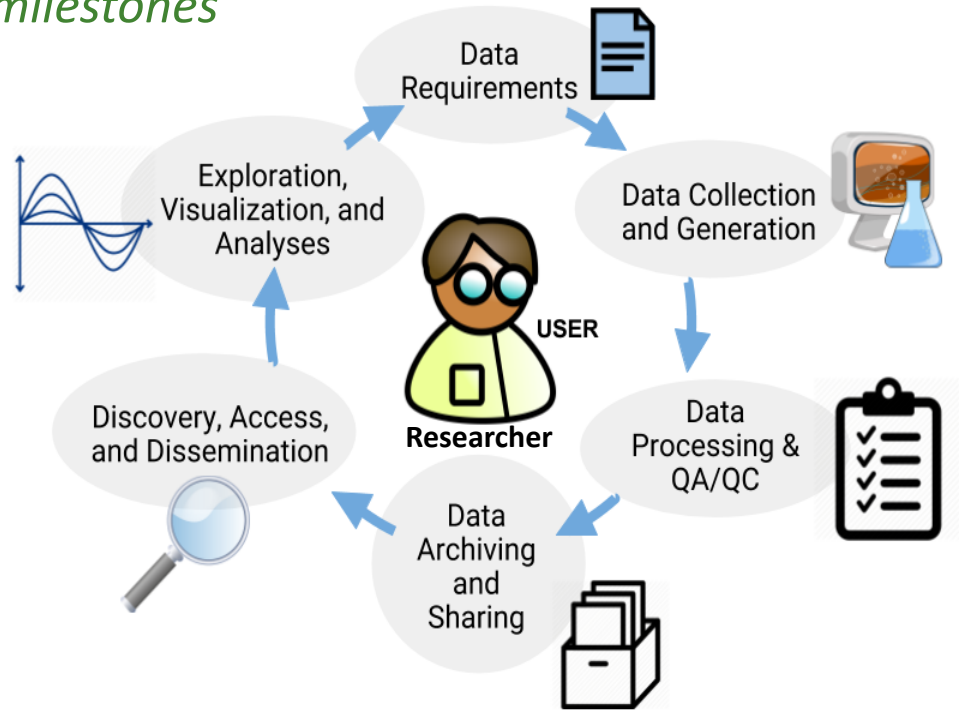


Management Approach

*Develop tools responsive to researcher needs through **researcher-centric approach***

FY18 Go/No-Go: The need for additional scope in FY19 will be assessed based on feedback from the ChemCatBio technical teams and steering committee.

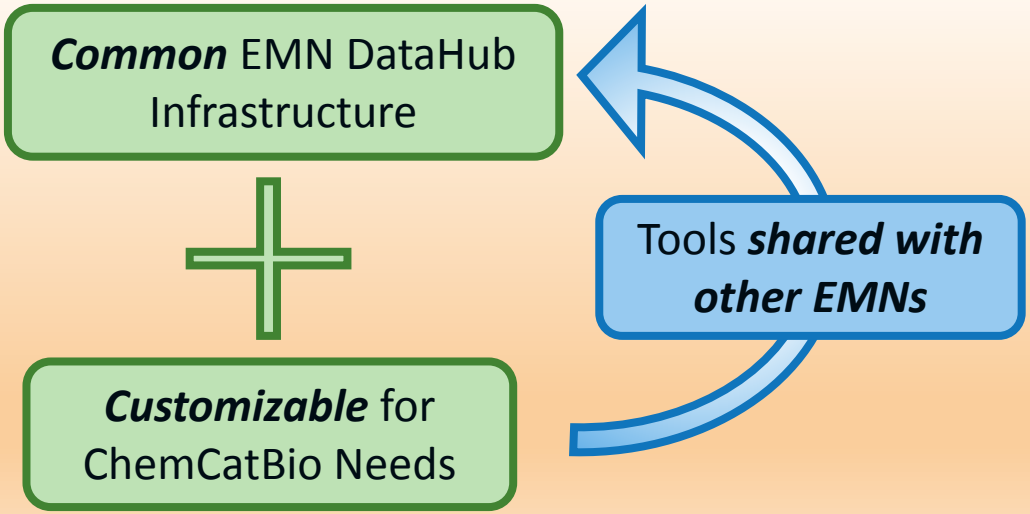
- *Led Data needs discussion at face-to-face meeting and teleconference and solicited specific researcher input through survey*
- *Feedback clearly indicated that **greatest value** is in **tool development***
- *Applied results to define FY19/20 milestones*



Technical Approach: Critical Challenges and Risk Mitigation

Critical Challenges

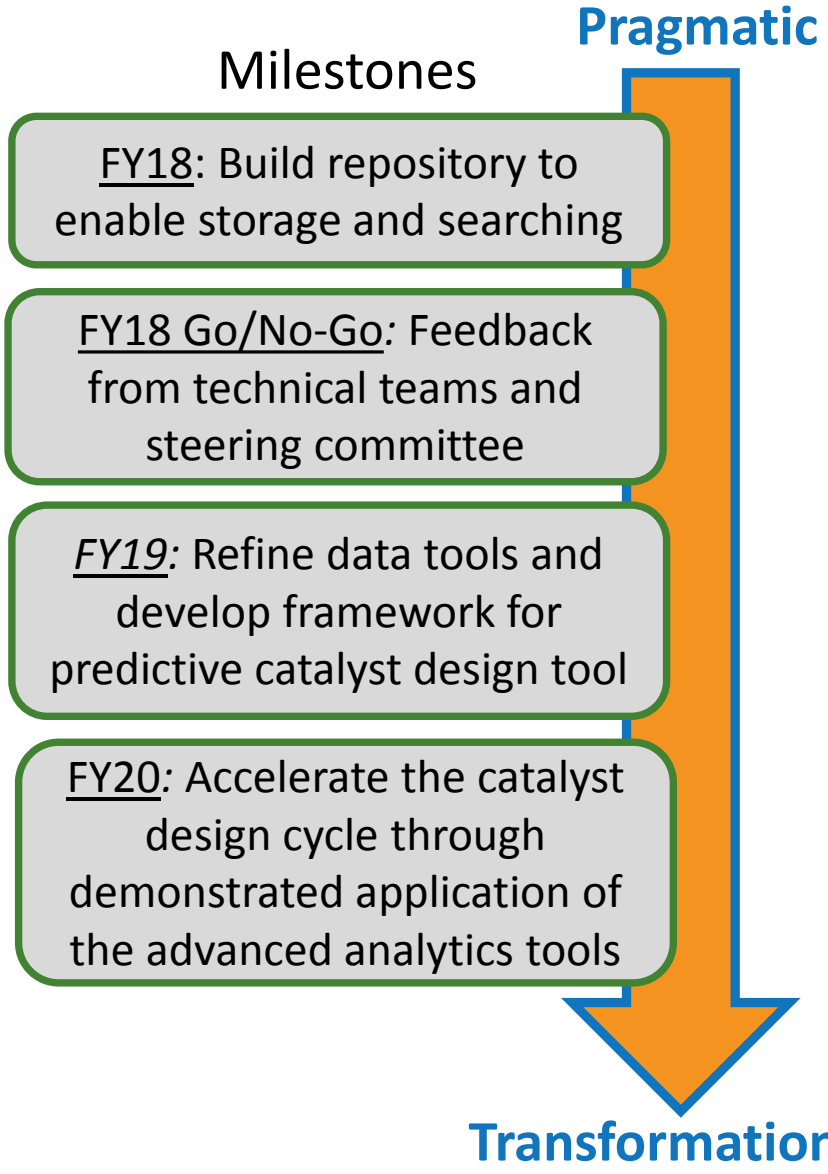
- Sharing Data
- Reducing Redundancy
- Accelerating Catalyst Design



Risk Mitigation

- Utilized by researchers
  - ➢ Integrate functionalities responsive to researcher needs
- Eliminates extra burdens
  - ➢ Develop plug-ins that upload large amounts of data with minimal researcher input

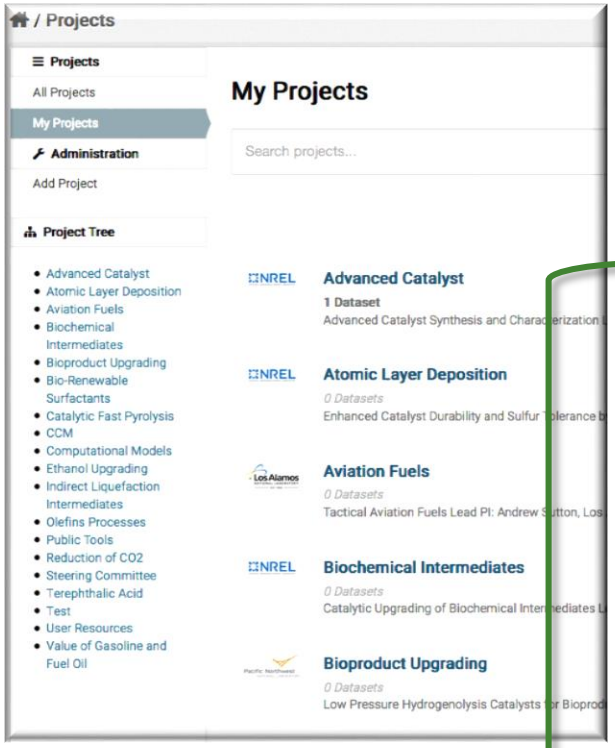
Technical Approach: Milestones and Success Indicators

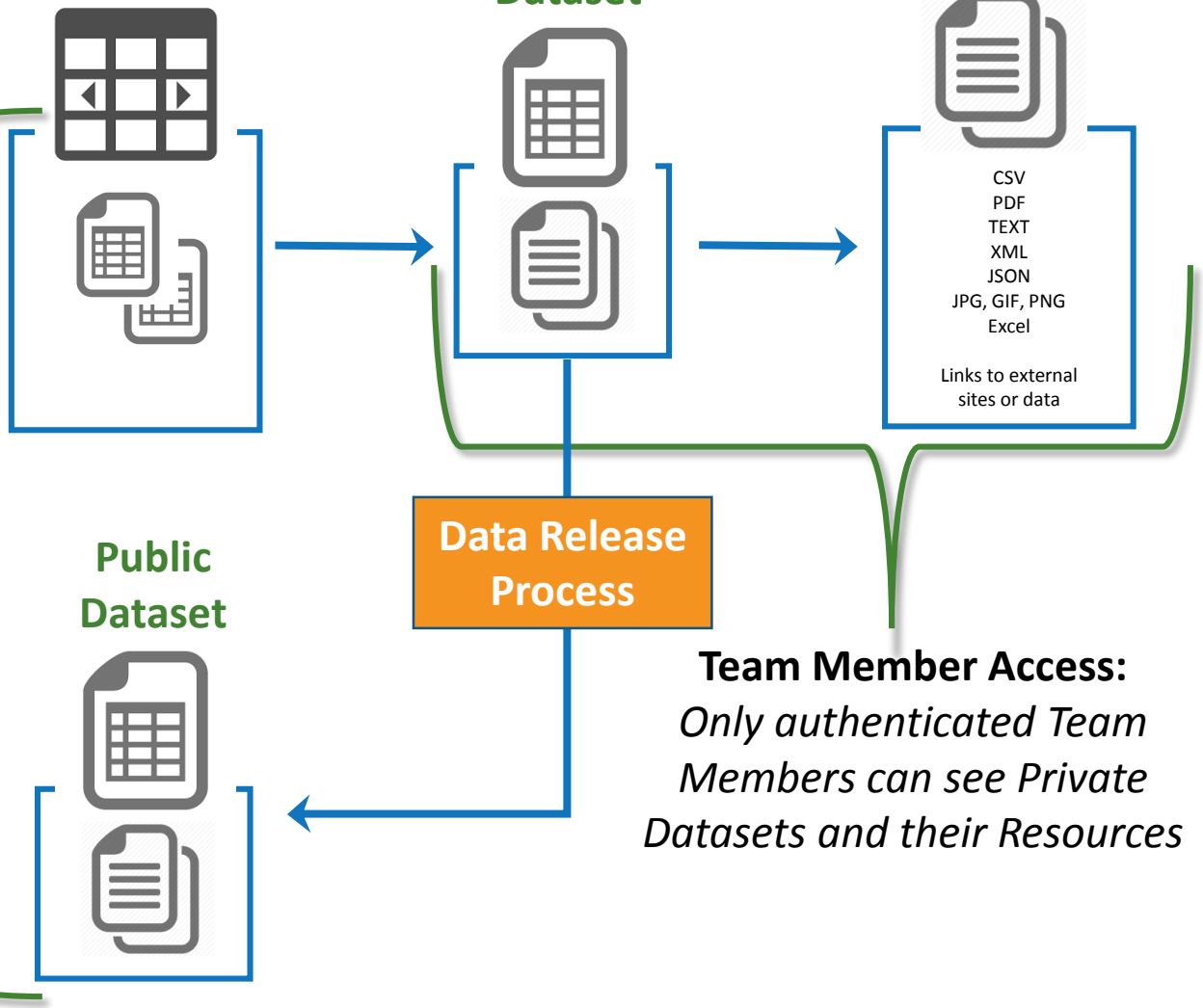


Critical Success Factors:

1. Data generated by researchers and industry partners can be **securely stored**, **easily searched**, and **quickly shared** with team members
2. Data is **processible** into formats that provide **key actionable information**
3. Data repository facilitates **dissemination** of ChemCatBio research through **controlled, public access**
4. Provide **predictive** capabilities that **accelerate** the **design of high-performing catalysts**

Progress: Centralized and Secure Data Repository



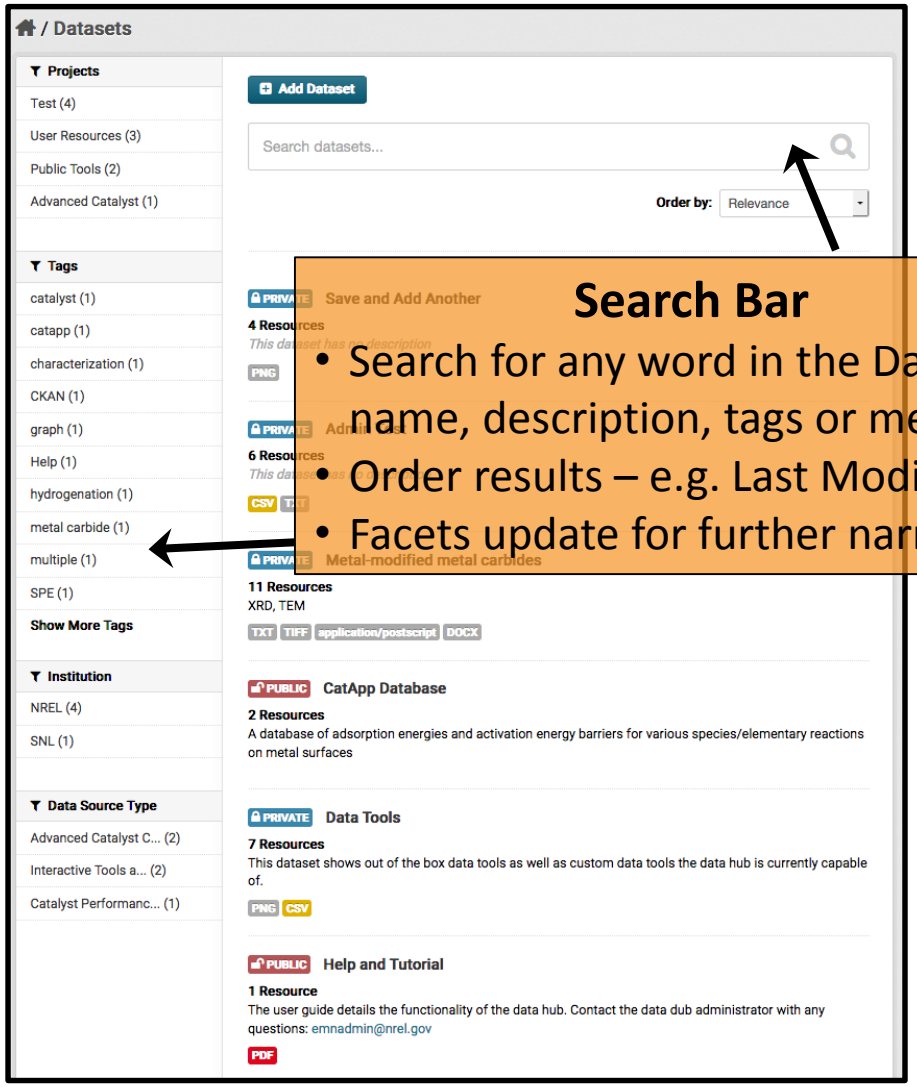


Public Access:
The public can see Project names, descriptions and abstracts. Public Datasets are also accessible.

Team Member Access:
Only authenticated Team Members can see Private Datasets and their Resources

Progress: Efficient User Data Search

Users can search across all accessible data

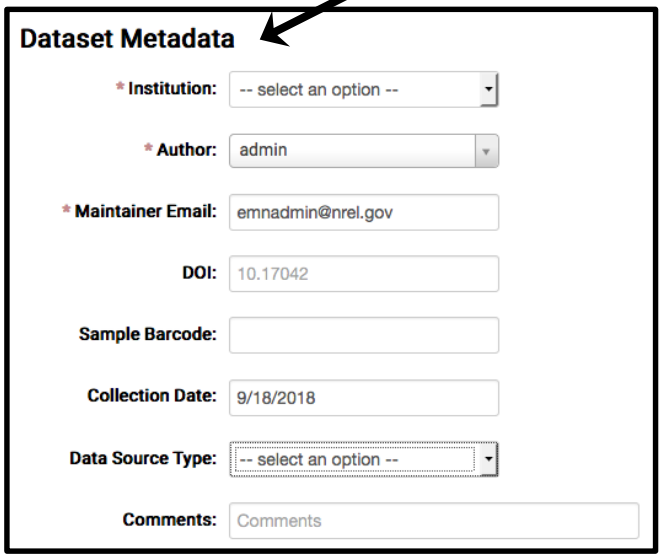


Search Bar

- Search for any word in the Dataset name, description, tags or metadata
- Order results – e.g. Last Modified
- Facets update for further narrowing

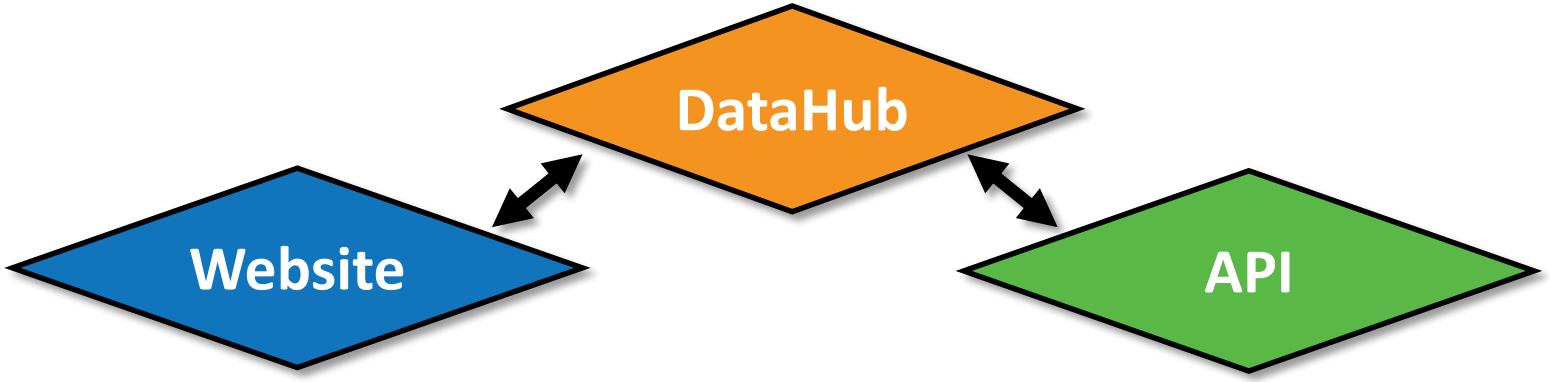
Metadata

- Entered/extracted at data upload
- Selected based on researcher input



Progress: Easy Access and Utilization of Data

Data Access and Application:

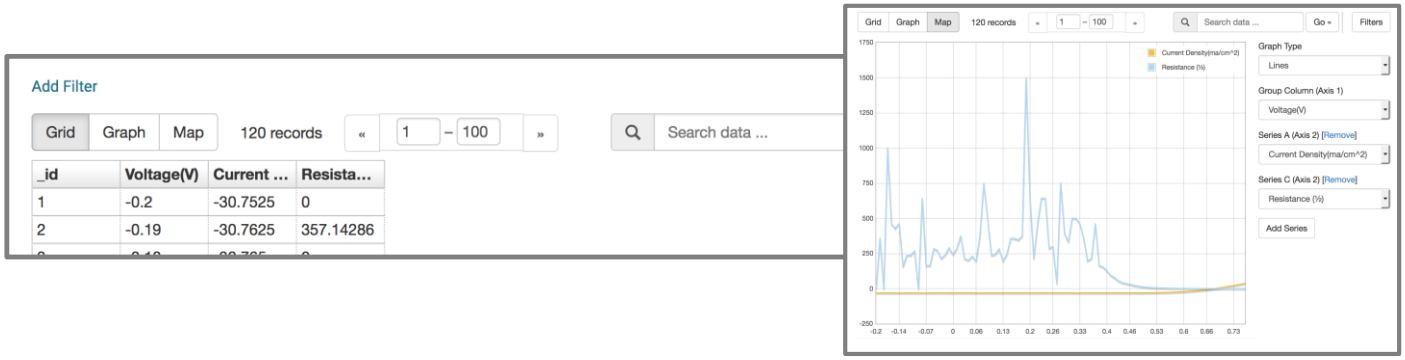


Provides user-friendly access for data and data tools

Enables building separate apps (e.g., Python notebooks) for visualization and analysis

Universal EMN Data Tool: Quick visualization, filtering, and searching

- Grid format or dynamic plotting of structured data (e.g., CSV files)

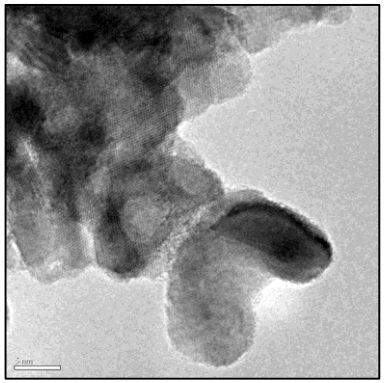


Progress: Fast Data Upload

Custom data tools address the **unique data needs** of ChemCatBio researchers

Automated Microscopy Metadata Extraction

- **Requested** by **ACSC** project team
- Enables **fast, batch data upload** and automatic extraction of instrument-specific metadata
- Enables **cross-institution collaboration** through advanced searching for image identification
- **Shared** across **all EMNs** through the EMN-Data Github repository



Original Image

Metadata Extraction Tool

Sample Name	**Ni_2P nanoparticle**
Acquisition Date	03/08/2017
Operator	Carrie Farberow
Microscope	FEI TEM
Magnification	2950000
…	

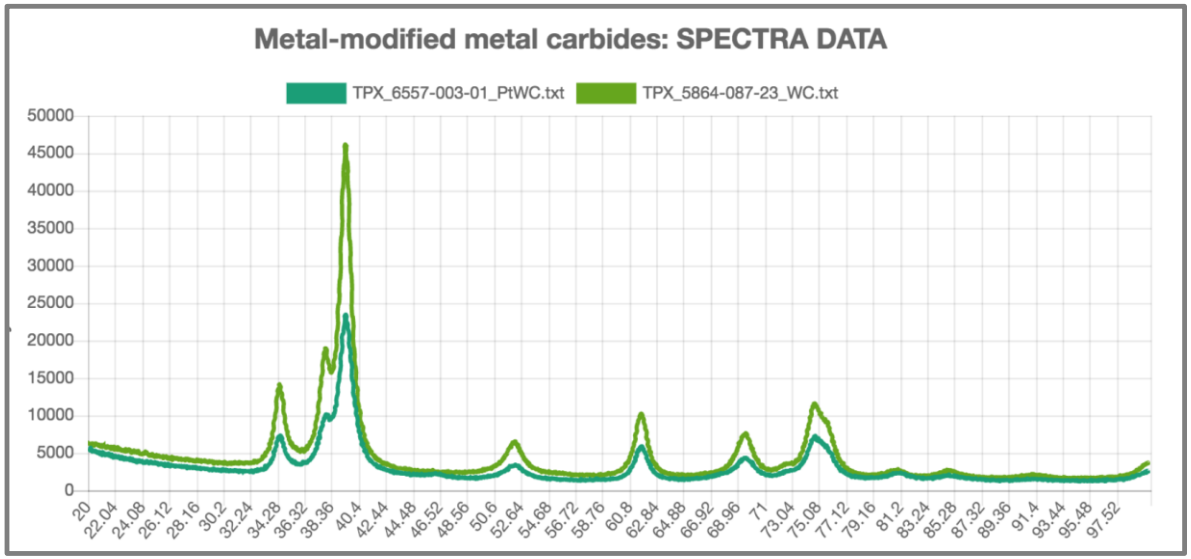
Extracted Metadata

Progress: Fast Data Visualization

Custom data tools address the **unique data needs** of ChemCatBio researchers

Multi-spectra Visualization

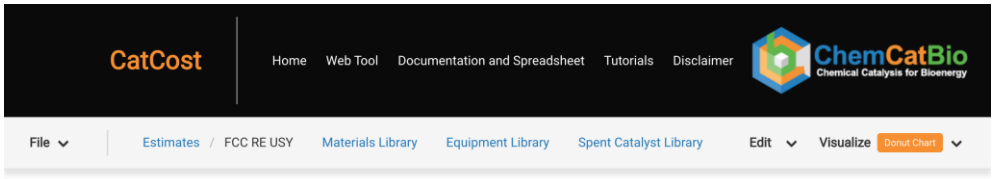
- **Requested** by **ACSC** project team
- Enables **rapid comparison** of data generated across institutions through **quick visualization** of **multiple spectral** data files in a **single plot**
- **Shared** across **all EMNs** through the EMN-Data Github repository

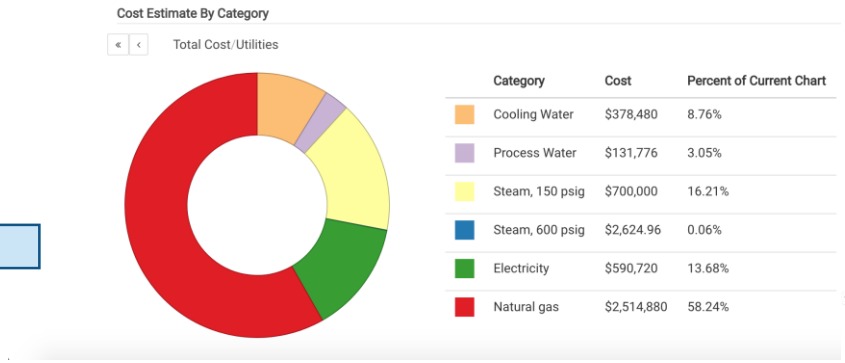


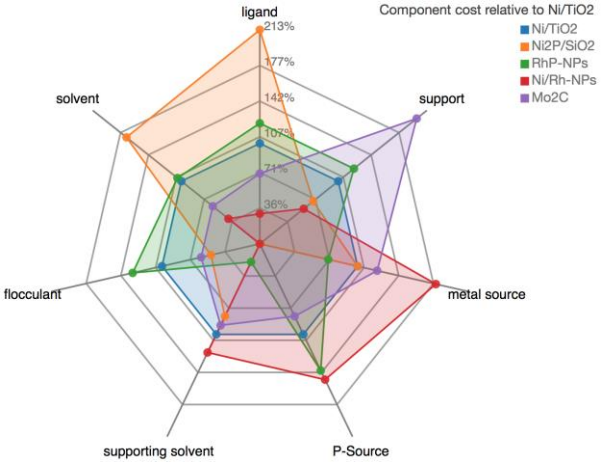
Progress: Facilitate Application of CatCost Tool

Integrated CatCost tool in DataHub for easy application in ChemCatBio research

- *Import materials libraries and estimates from the DataHub*

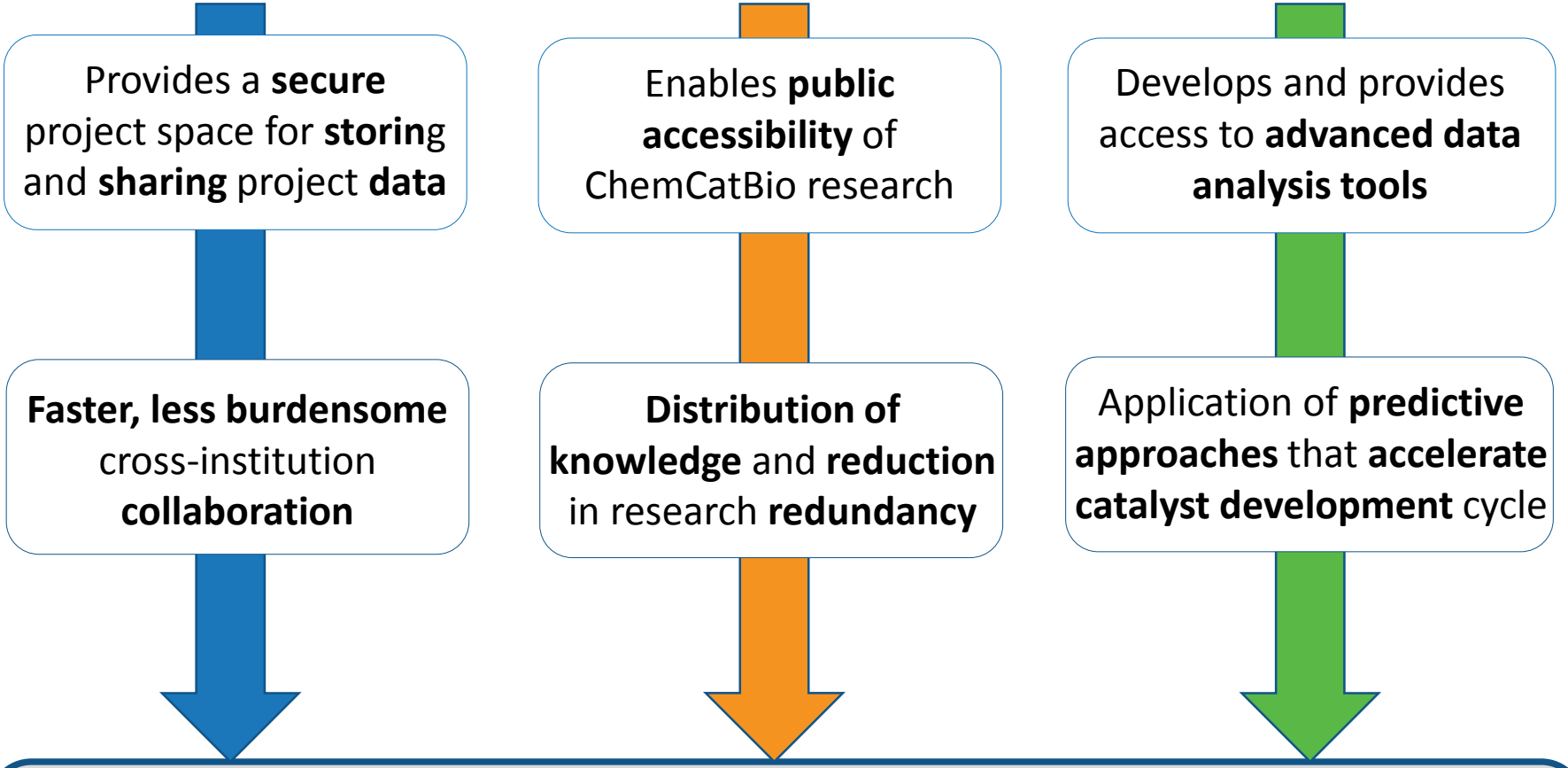






- *Export model outputs to the DataHub for archive and sharing*

Relevance: Faster, Less Costly Catalyst Development



BETO 2019 MYP **"Critical Barrier"**: **Reduce** the **time and cost** to develop active, selective, low-cost and durable **real-world catalysts** to produce **desired end-products from biomass and waste resources**.

Relevance: Broadly Applicable Catalyst R&D Tools

This project develops and provides access to broadly applicable and transformative catalyst R&D tools

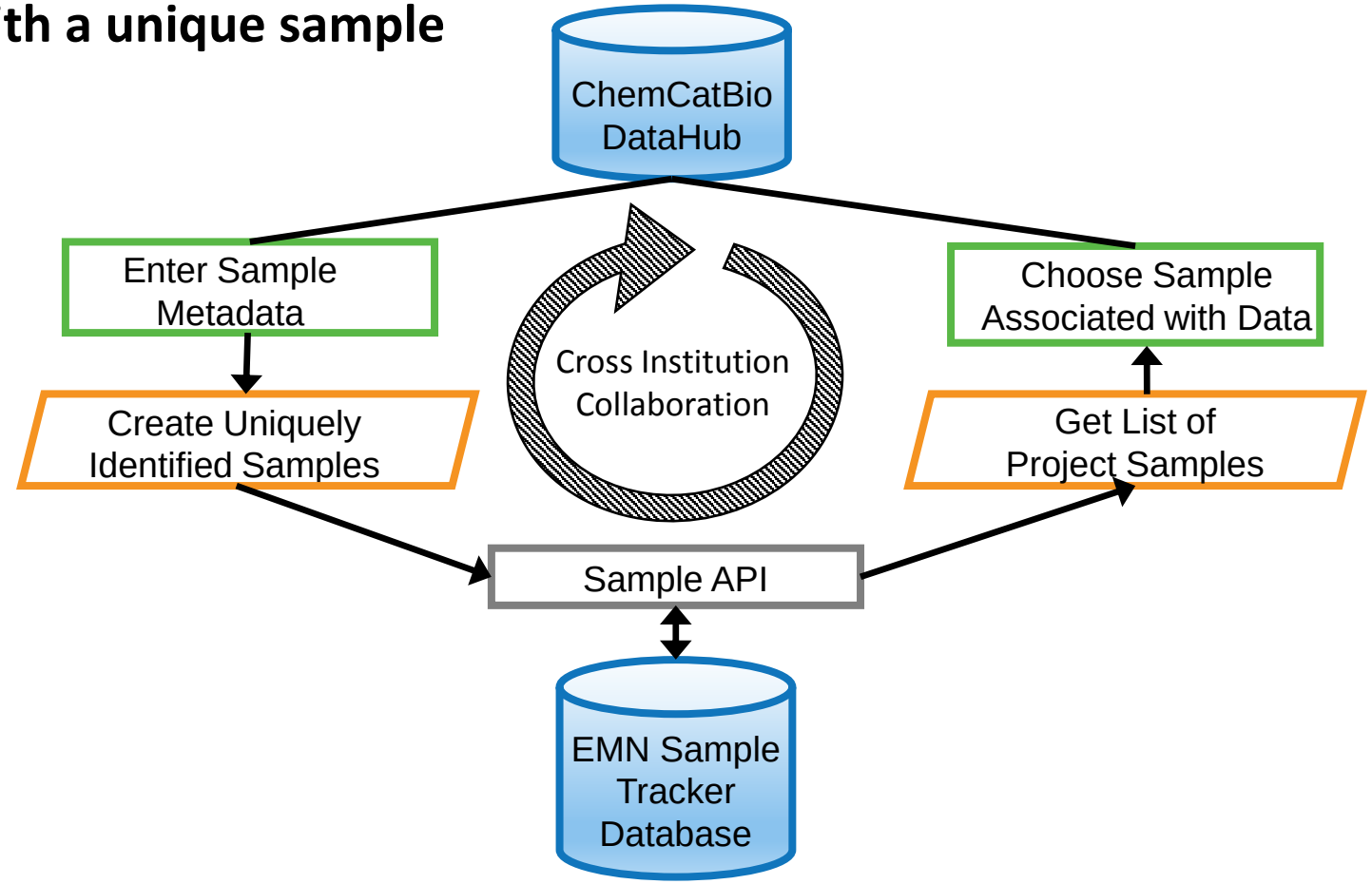
- Tools developed are **pathway-independent** and therefore **widely applicable** by industry for catalyst development
- **Public access** to both general and reaction-specific catalyst property data and associated analysis tools
- Centralized data storage and sharing location for ChemCatBio **DFA projects** creates immediate **value to industry partners**

Future Work: Sample Management

<u>Goal:</u> Researchers across institutions can share datasets and easily associate them with a unique sample

<u>Outcomes:</u>

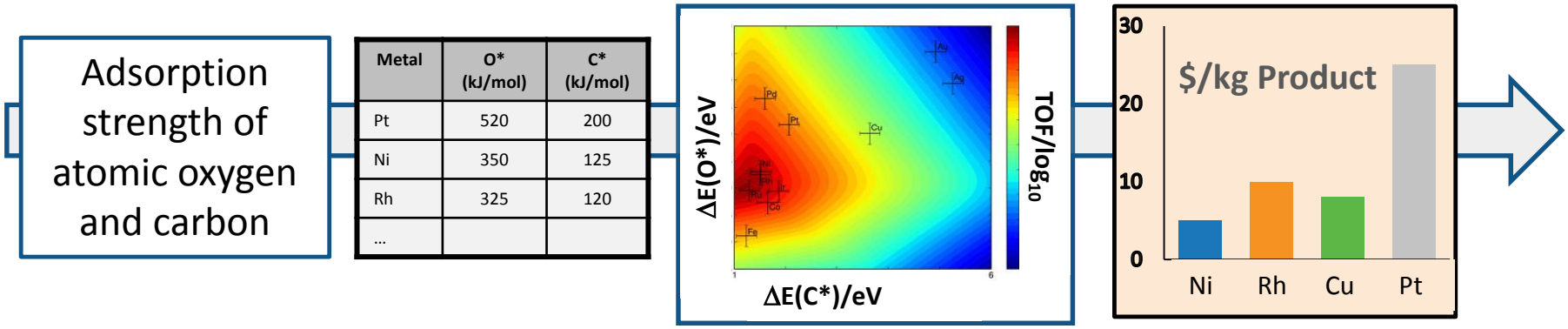
- ✓ **Reduces redundancy**
- ✓ **Increases knowledge gained**



Future Work: Catalyst Design Engine

Conventional State-of-the-Art Computational Catalyst Screening Process

Example: Methane Steam Reforming



Process Deficiencies

- Redundancy in descriptor value calculations
- Candidate materials often limited to expensive noble metals
- Difficult to consider economics for pre-commercial materials
- Lack of known descriptors for many complex reaction chemistries

Future Work: Catalyst Design Engine

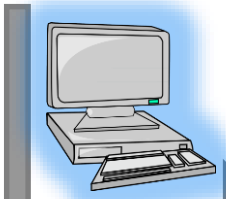
Integrated, broadly applicable, catalyst design tool that considers trade-off between performance and material cost to accelerate the design of high-performing, low-cost catalysts

Established Property-Performance Relationships

Computational Property Database

- Centralized, searchable compilation of published data
- Identification and fill-in of data gaps through CCPC collaboration
- Include integrated analysis and visualization tools

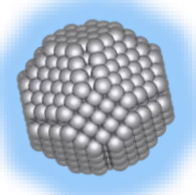
Leverage database and experimental collaborations for descriptor discovery



Predicted Performance of Diverse Materials

CatCost

- Predicted cost for pre-commercial materials
- Completed integration in DataHub in FY18



Optimized Low-cost, High-performing Catalysts for Targeted Synthesis and Testing

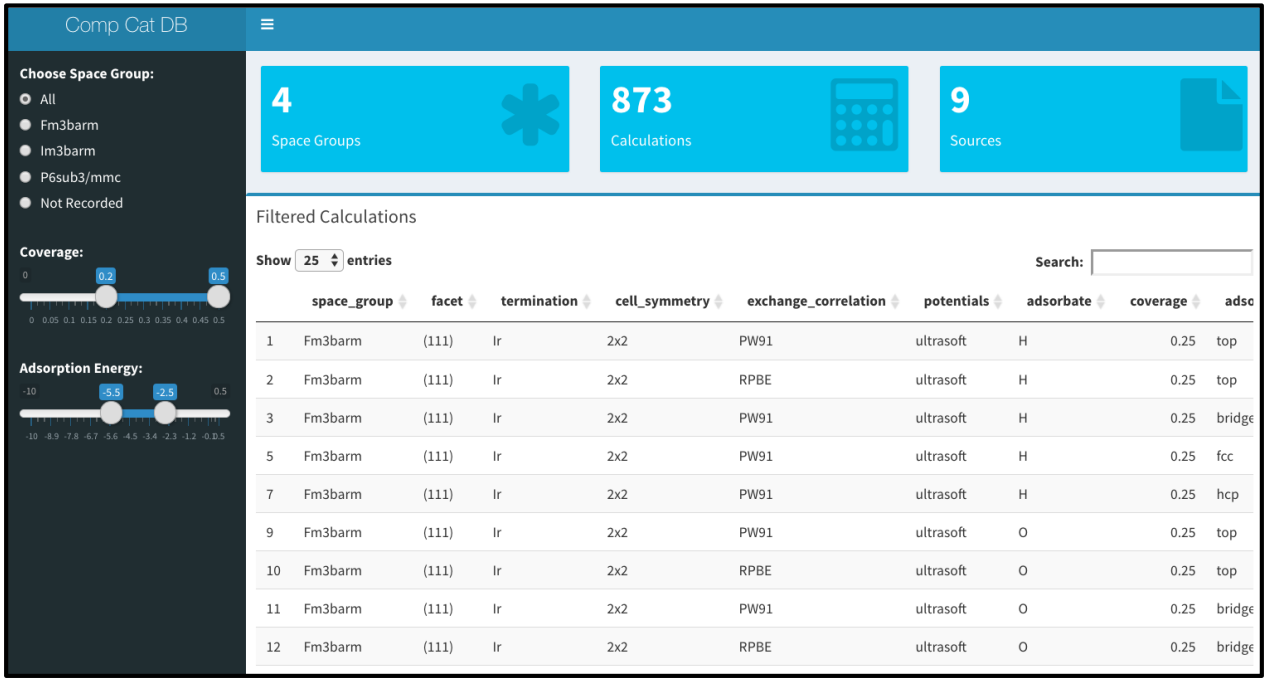
Future Work: Computational Catalyst Property Database

Progress to Date:

- Determined rules for data inclusion that maximize database utility while minimizing database compilation time
- Designed user interface framework
- Created prototype filter and search functions
- Populated database with initial set of 900+ calculations from literature

Next Steps:

- Expand data an order of magnitude (FY19Q2-Q4)
- Refine user interface (FY19Q2-Q3)
- Add analysis features, including visualization and plotting (FY20)



Summary

Project Goal: **Reduce** the **time and cost** for **ChemCatBio** and the **bioenergy industry** to develop **real-world catalysts** for the production of **desired end-products from biomass and waste resources** by providing:

1. Secure **data storage, sharing, and analysis** for cross-institution collaboration and **knowledge dissemination**
2. Development and application of **advanced analytics tools** that provide **predictive catalyst development** capabilities

Approach: **Researcher-centric** to ensure that the tools developed are **responsive** to researcher needs and **collaborative** with other **EMN data teams** to **eliminate redundancy**

Future work defined **based on researcher feedback** focuses on development of transformational tools including a **pathway-independent integrated catalyst design engine** that considers critical **trade-off** between **predicted performance** and catalyst **material cost** to **accelerate the design of optimized catalyst** formulations for a diversity of targeted end-products.

Acknowledgements

NREL EMN Data Team

Kristin Munch

Courtney Pailing

Nicholas Wunder

John Yarbrough

David Evenson

Josh Schaidle

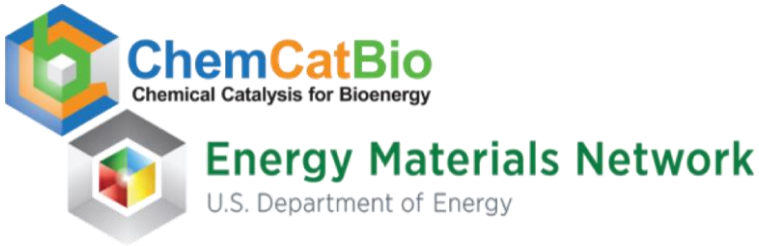
Matt Jankousky

Tuong Bui

ChemCatBio Researchers

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)



Thank you.

ChemCatBio
Chemical Catalysis for Bioenergy