# CKineticsDB - An Extensible and FAIR Datahub for Multiscale Modeling in Heterogeneous Catalysis

**Siddhant Lambor**, Sashank Kasiraju, Dionisios Vlachos
University of Delaware

**DOI:** 10.1021/acs.jcim.3c00123

**VIRTUAL KINETICS LAB**
ONLINE WORKSHOP
15th December 2023

# CKineticsDB - An Extensible and **FAIR** Datahub for Multiscale Modeling in Heterogeneous Catalysis

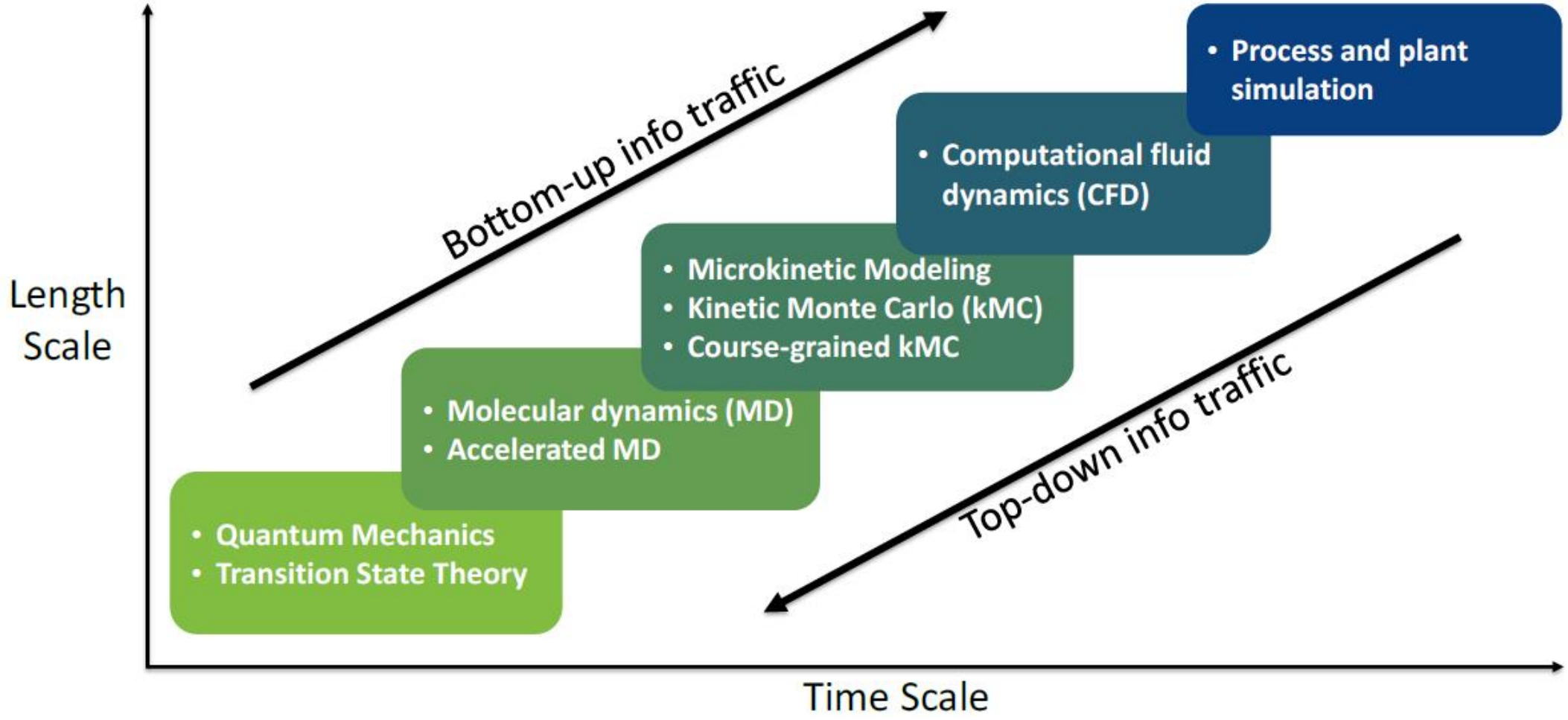
[1]

**Findable, Accessible, Interoperable, Reusable**

- Easily share and integrate data
- Make data interpretable by humans and by machines

1. Wilkinson, M. D. et. al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 2016, 3, 160018

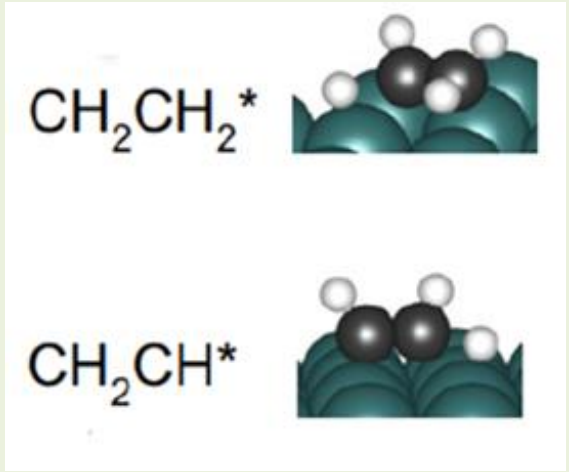
# Process Intensification and Material Discovery using Multiscale Modeling




Lym, J. (2020), PhD Thesis, University of Delaware

# Density Functional Theory (DFT) based Microkinetic Modeling (MKM)

## Density Functional Theory




Simulate the behavior of materials at the atomic level, to obtain bond length, electronic energies, vibrational frequencies, etc.

**Statistical mechanics** to calculate equilibrium constants for elementary steps

**Transition state theory** to calculate rate of reaction

DFT is computationally expensive and has a trade-off between accuracy and time for computation

## Microkinetic Modeling

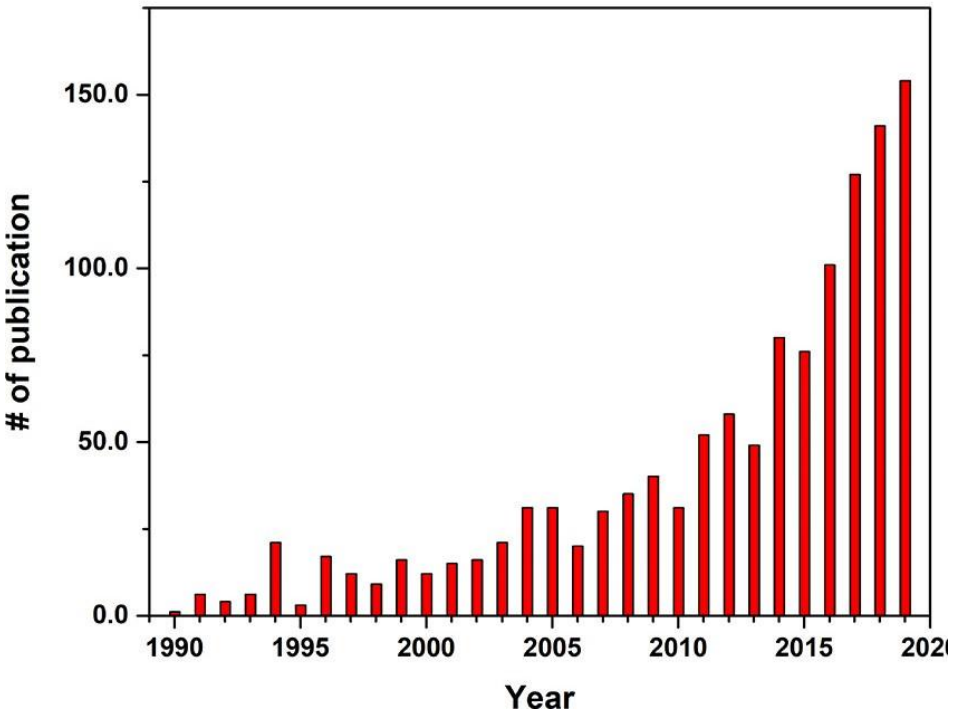
Reaction Mechanism

$$N(T) + *(S) \rightleftharpoons N(S) + *(T)$$
$$N(T) + *(S_L) \rightleftharpoons N(S_L) + *(T)$$
$$NH(T) + *(S) \rightleftharpoons NH(S) + *(T)$$
$$NH_2(T) + *(S) \rightleftharpoons NH_2(S) + *(T)$$
$$NH_3(T) + *(S) \rightleftharpoons NH_3(S) + *(T)$$

Solving equations representing the reaction mechanism to identify rate limiting steps, catalytic activity, predominant species, selectivity, and more

# No comprehensive databases for multiscale modeling in heterogeneous catalysis

Number of publications on microkinetic modeling




*Motagamwala et. al.*
*Chem. Rev.* 2021, 121, 2, 1049−1076

Local Clusters

Cloud Repositories

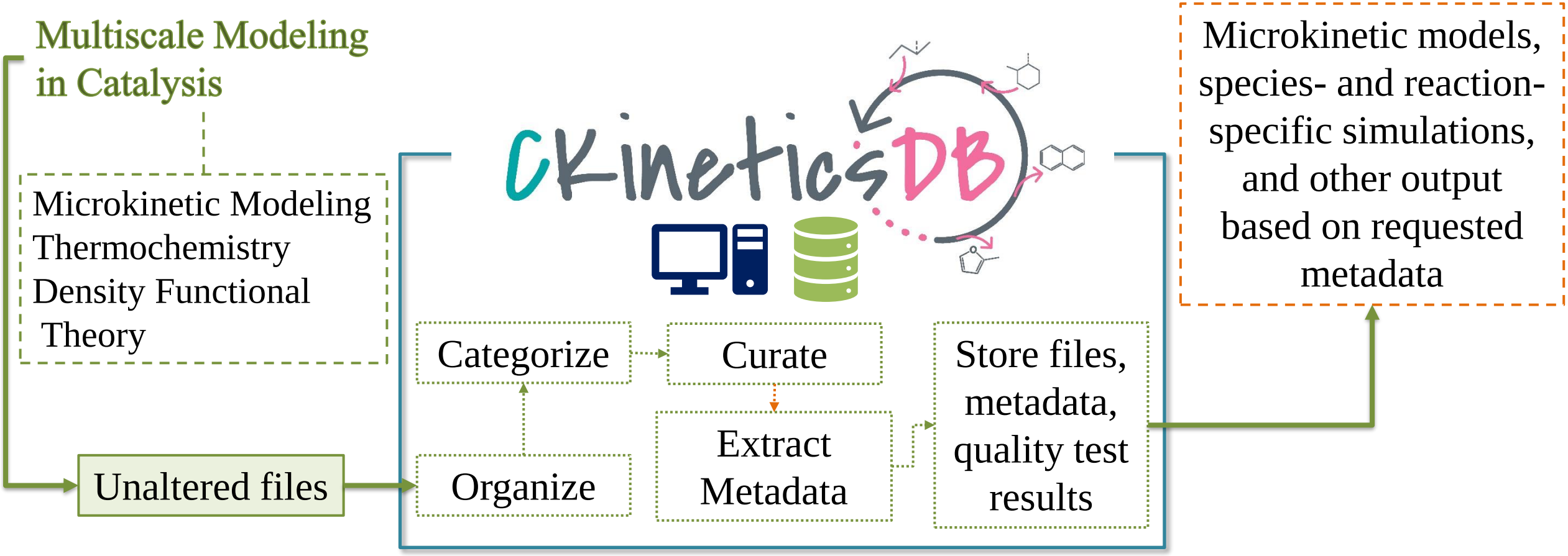
Portable Storage Devices

- Logistical hindrances to access data leading to redundancy
- Excessive demand for computational resources
- Advent of machine learning has increased the value of data

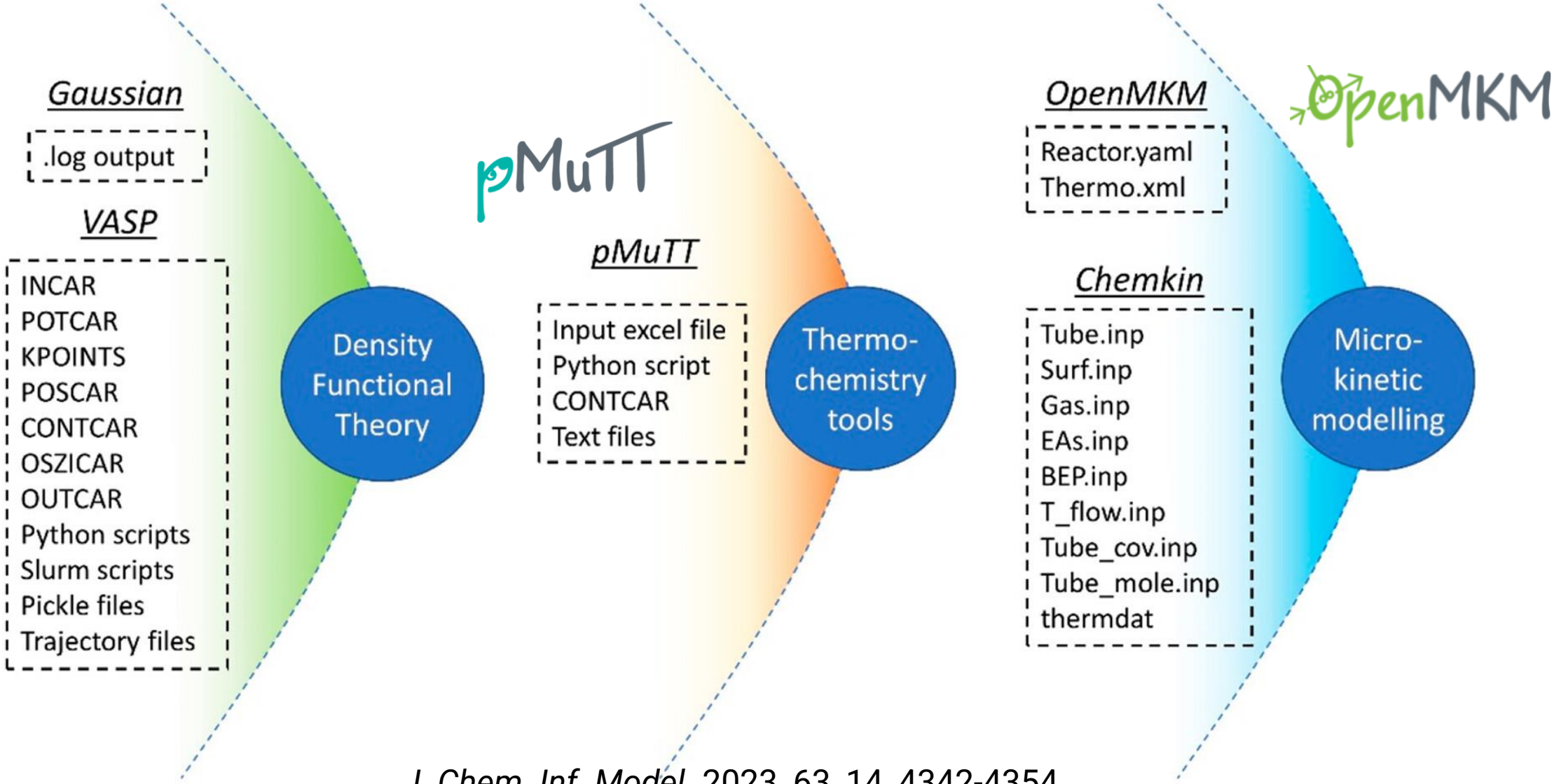
# Value in data management for multiscale modeling in catalysis

| Organize Data | Extract Information | Accelerate applications |
|---|---|---|
| Reaction Data | • Reaction mechanisms, microkinetic models<br>• Kinetic and thermochemical parameters | • Minimize DFT simulations<br>• Facilitate thermodynamic and kinetic studies<br>• Utilize chemically similar data for new mechanisms<br>• Develop multiscale software |
| Thermochemistry | • Scripts to process DFT output; data from NIST<br>• DFT energies and frequencies | |
| DFT calculations of molecules | • Input settings<br>• Catalyst specifications | |

# Chemical Kinetics Database



# CKineticsDB stores the simulation files involved in multiscale modeling




*J. Chem. Inf. Model.* 2023, 63, 14, 4342-4354

# CKineticsDB Infrastructure Components

## Database Management System

MongoDB is a, non-relational, document-oriented database

- Dynamic schema
- Expansion without downtime
- Online support
- Cloud-based solutions
- Vendor support

## Frontend Python Software

Integrations with Vlachos Group in-house tools

pMuTT [1]

DescMAP [2]

OpenMKM [3]

Develop scientific computing features

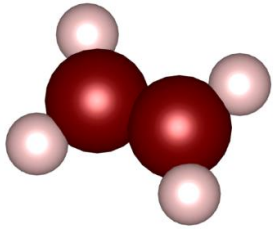


1. Jonathan Lym, et. al., Comp. Phy. Comm., 2020; 2. Xue Zong, et. al., Comp. Phy. Comm., 2023; 3. Bharat Medasani, et. al., J. Chem. Inf. Model. 2023

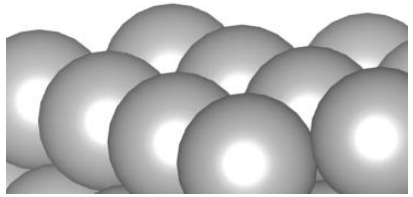
# Current Data Snapshot

## 14000+ DFT calculations

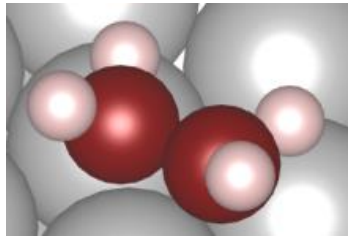
Gas Phase



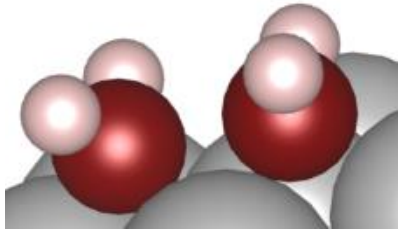
Bulk structures



Adsorbates



Transition states



## Catalysts

**Pure Metals**
Ag, Au, Cu, Ir, Ni,
Pd, Pt, Rh, Ru

**Zeolites**
H-BEA

**Metal oxides**
$Al_2O_3$, $ReO_x$, $TiO_2$,
$SiO_2$, $ZrO_2$

## Reaction Chemistries

Hydrogenolysis, dehydrogenation, hydroformylation, hydrodeoxygenation, C-O bond activation, and acylation; several catalyst facets and active center structures

# Ammonia Decomposition Mechanism Files

$$N(T) + *(S) \rightleftharpoons N(S) + *(T)$$

$$N(T) + *(S_L) \rightleftharpoons N(S_L) + *(T)$$

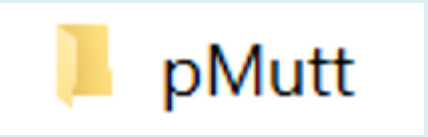
$$NH(T) + *(S) \rightleftharpoons NH(S) + *(T)$$

$$NH_2(T) + *(S) \rightleftharpoons NH_2(S) + *(T)$$

$$NH_3(T) + *(S) \rightleftharpoons NH_3(S) + *(T)$$

OpenMKM input files: reactor.yaml, thermo.yaml


Python Scripts, CONTCAR files, MS Excel file



# Ammonia Decomposition Mechanism Files

$$N(T) + *(S) \rightleftharpoons N(S) + *(T)$$

$$N(T) + *(S_L) \rightleftharpoons N(S_L) + *(T)$$

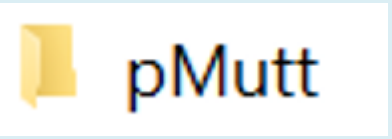
$$NH(T) + *(S) \rightleftharpoons NH(S) + *(T)$$

$$NH_2(T) + *(S) \rightleftharpoons NH_2(S) + *(T)$$

$$NH_3(T) + *(S) \rightleftharpoons NH_3(S) + *(T)$$

OpenMKM input files: reactor.yaml, thermo.yaml

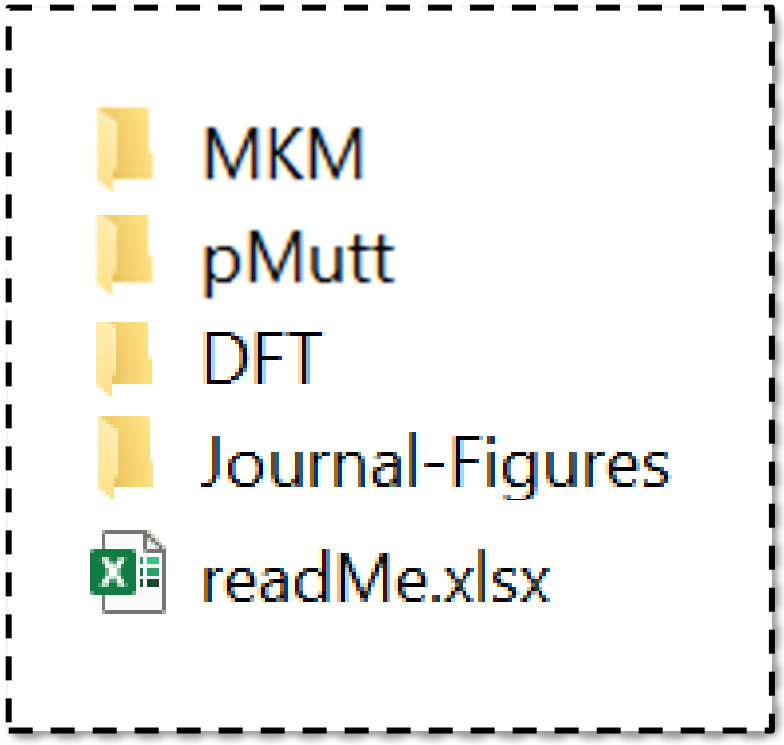

Python Scripts, CONTCAR files, MS Excel file




DFT simulation files

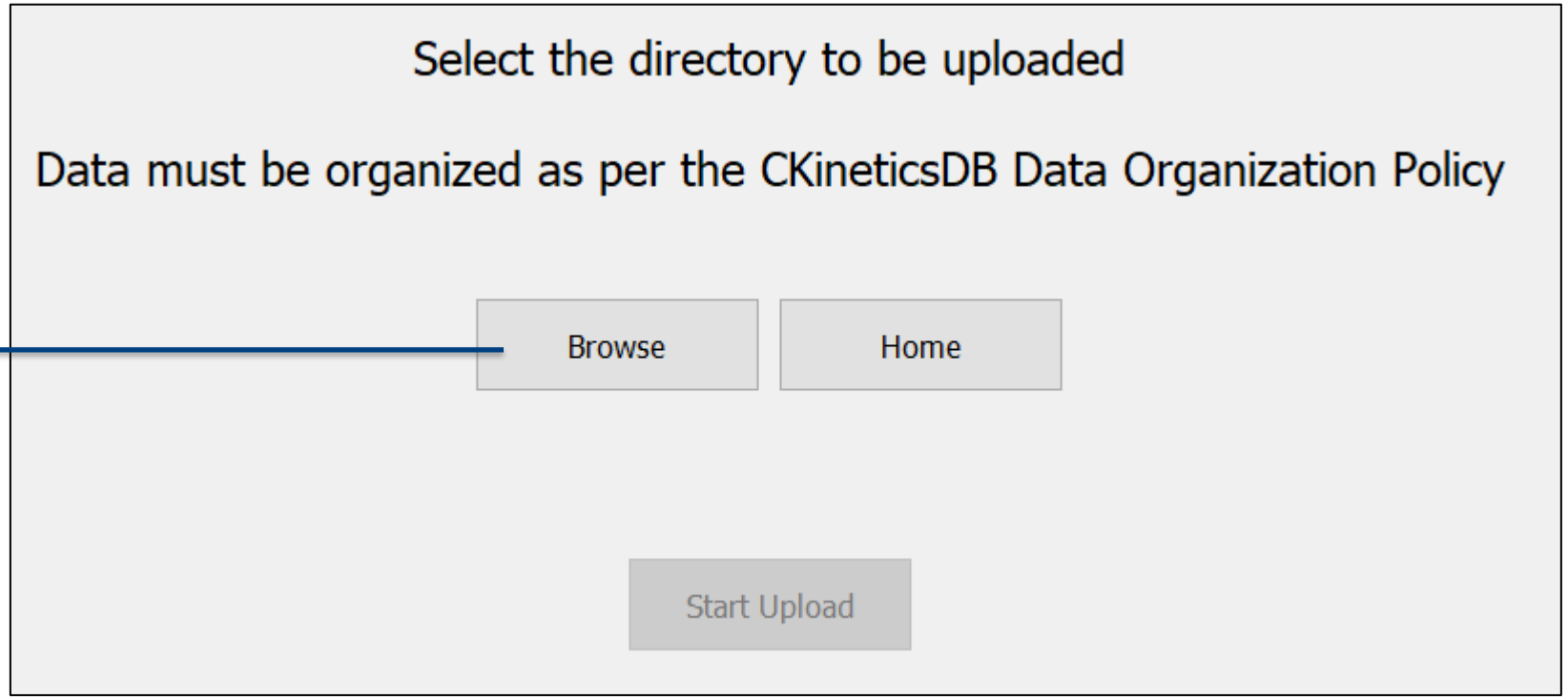
# Complete Dataset for a Published Microkinetic Model

Data Organization Policy




- Complete recipe for reproducing results shown in papers
- User-fed metadata for providing species nomenclature, software metadata, and research provenance
- Uniformity across all datasets
- Facilitating downstream software integration

# Uploading to CKineticsDB




vlab-ammonia-demo/

- DFT
- MKM
- pMutt
- Journal-Figures
- readMe.xlsx
- CKineticsDB_DFTassessment.pdf
- CKineticsDB_filesAssessment.xlsx

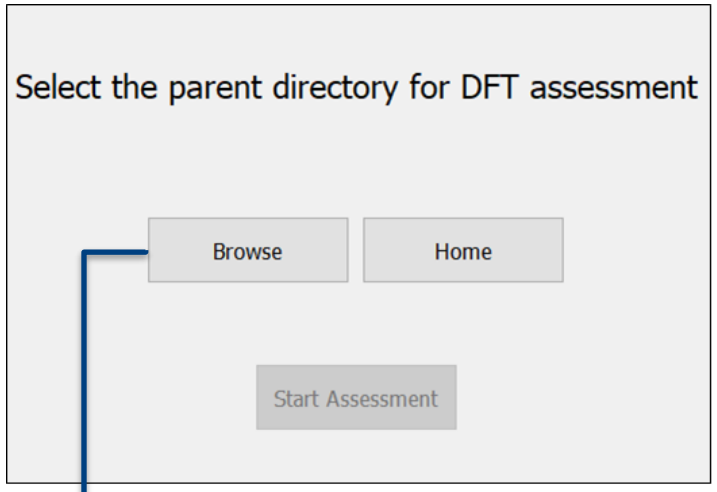
Data Quality Assessment

# Demo

- Data Upload, readMe
- Data Quality Assessment

# DFT Data Quality Assessment Workflow

Select the directory of calculations to be tested




```
Project/
└── DFT/
    ├── Calculation_Dir
    ├── Calculation_Dir
    ├── ...
    └── Calculation_Dir
```

Identify the type of calculation ------ Relaxation, Dimer, NEB, etc.

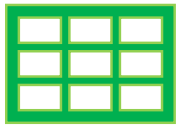
↓

Run tests based on the calculation ----- Convergence, KPOINTS, more

↓

Provide results and a dataset summary

Assessment results (PDF)



Missing files (MS Excel)

# DFT Data Quality Tests and Output

| Software | Calculation | Quality Test(s) |
|---|---|---|
| VASP | Ionic Relaxation | Convergence, Kpoints, Encut |
| VASP | Dimer | Convergence, Curvature, Kpoints, Encut |
| VASP | (Climbing - /) Nudged Elastic Band (inclusive of all images) | Convergence of the highest energy image, Kpoints, Encut |
| VASP | Individual NEB Image | Convergence |
| VASP | Frequency Analysis | Frequencies assessment, Kpoints, Encut |
| Gaussian | Optimization | Convergence |
| Gaussian | Frequency Analysis | Frequencies assessment |

Summary of a complete dataset's assessment

**Summary:**

**Total Number of Calculations: 239**

**Passed all tests: 168**

**Need to be reviewed: 71**

**Related Inconsistencies:**

**Ionic step information not available : 4**

**More than one imaginary frequencies : 10**

**No frequencies found in vibrational calculation : 56**

**No POSCAR file : 1**

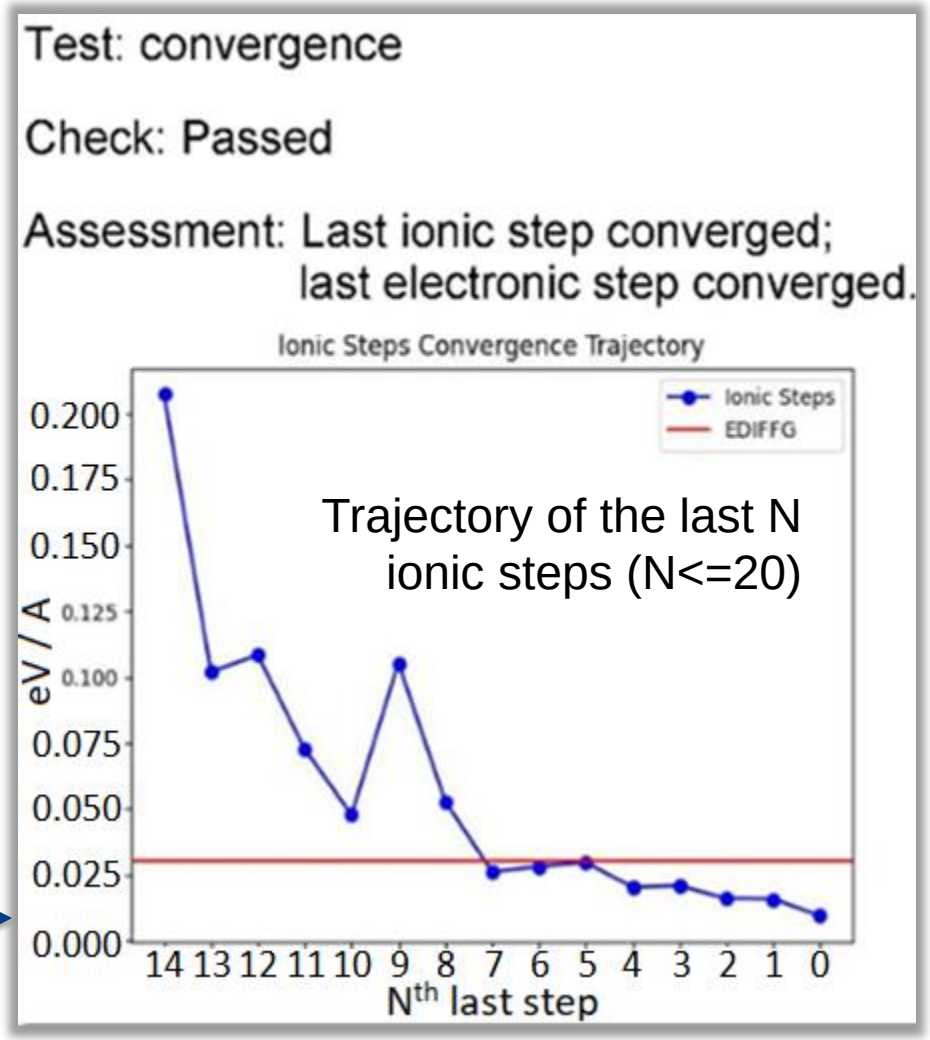
# Example of DFT Quality Test – Convergence (VASP)

Parse through DFT calculation files and extract information

**INCAR** – EDIFF, EDIFFG
**OUTCAR** – ionic steps and last electronic iteration

Check for the convergence metrics based on input settings and obtained output

Add test results to the assessment document

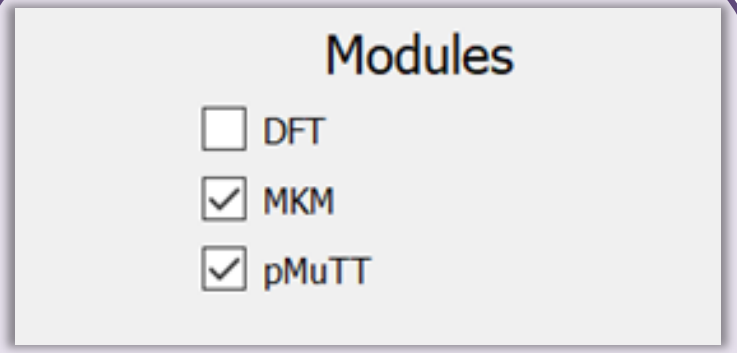


# Demo

- Downloading projects, reactions, molecules

# Downloading Parameters


- Download all data in JSON format
- Download data in original uploaded file format




- Either download the complete dataset or partially based on the ‘scale’ of interest


- Download the pMuTT data of the complete dataset or only of the selected reactions

# Reaction Mechanisms

| Chemistry | System | Reactions (species) |
|---|---|---|
| Heterogeneous catalysis | Ammonia Decomposition | 14 (26) |
| Heterogeneous catalysis | Propane aromatization | 19907 (5909) |
| Gas phase combustion | Hexadecane combustion[1] | 8130 (2116) |
| Metabolism | E.coli genome[2] | 2077 (1039) |
| Metal chemistry | Glycerol decomposition[3] | 3313 (537) |

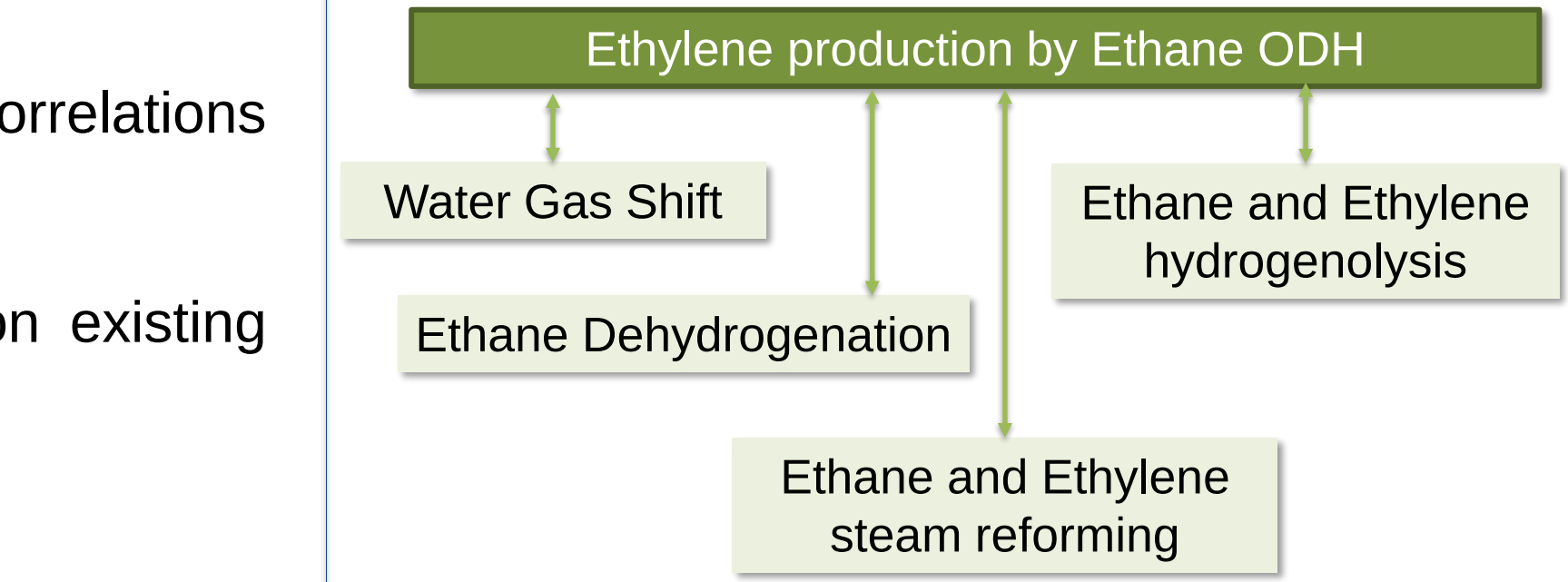
- Manual construction generally impractical
- Automated network generators
  - Generate all possible reactions and species of the network
  - Use reaction rules as basis – **"rule-based"**

[1] Combustion and Flame 2009, 156 (1), 181; [2] Mol Syst Biol. 2007, 3, 121; [3]Green Chem., 2014,16, 813

# Uses of Reaction and Multiscale data

- Recreate results of publications
- Develop models and correlations between scales
- Kinetic studies based on existing mechanisms

Create new microkinetic models from existing data




*GR Wittreich, HE Toraman, DG Vlachos, North American Catalysis Society Meeting, 2019*

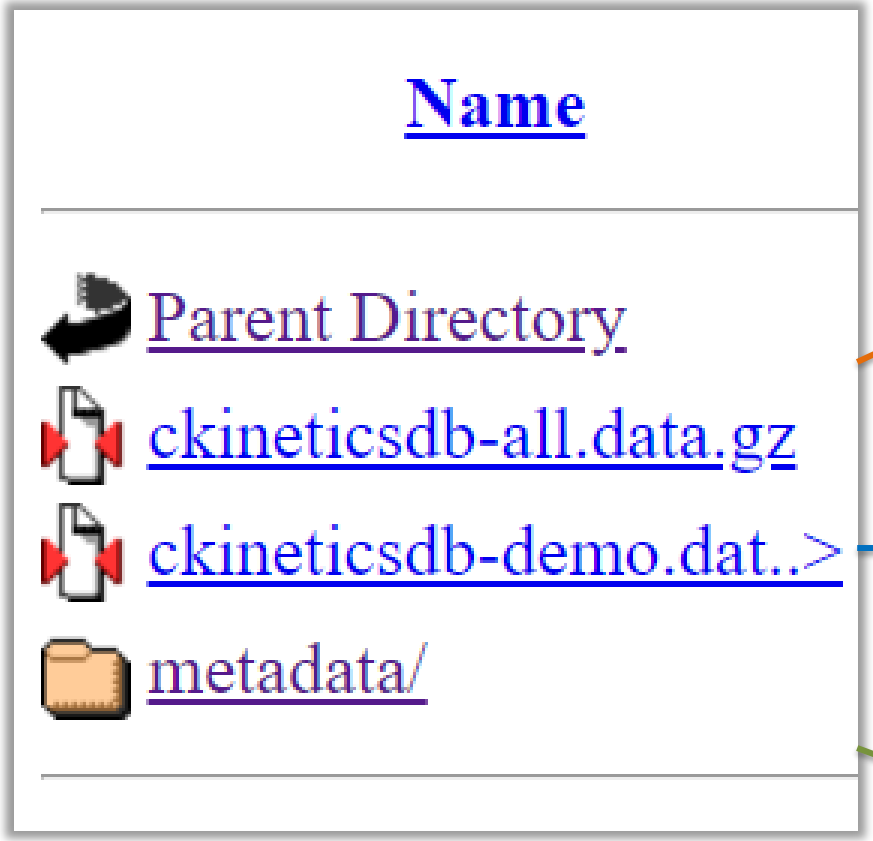
RAPID
Transforming Process Industries

# Demo

- Command line interface

# CKineticsDB Distribution

Available at: https://files.ccei.udel.edu/p/CKineticsDB/data/




Data associated with several publications of Vlachos group pertaining to microkinetic modeling

Demo containing only one dataset to test software setup

MS Excel and JSON files containing metadata of the complete dataset available above

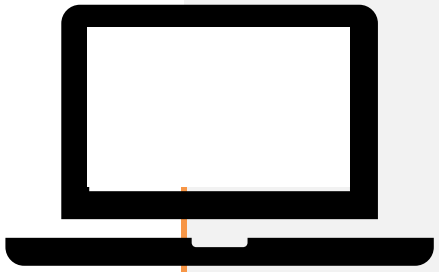
# CKineticsDB Data Workflow

https://files.ccei.udel.edu/p/CKineticsDB/data/

Compressed data snapshots generated by MongoDB


Download dataset

**User**

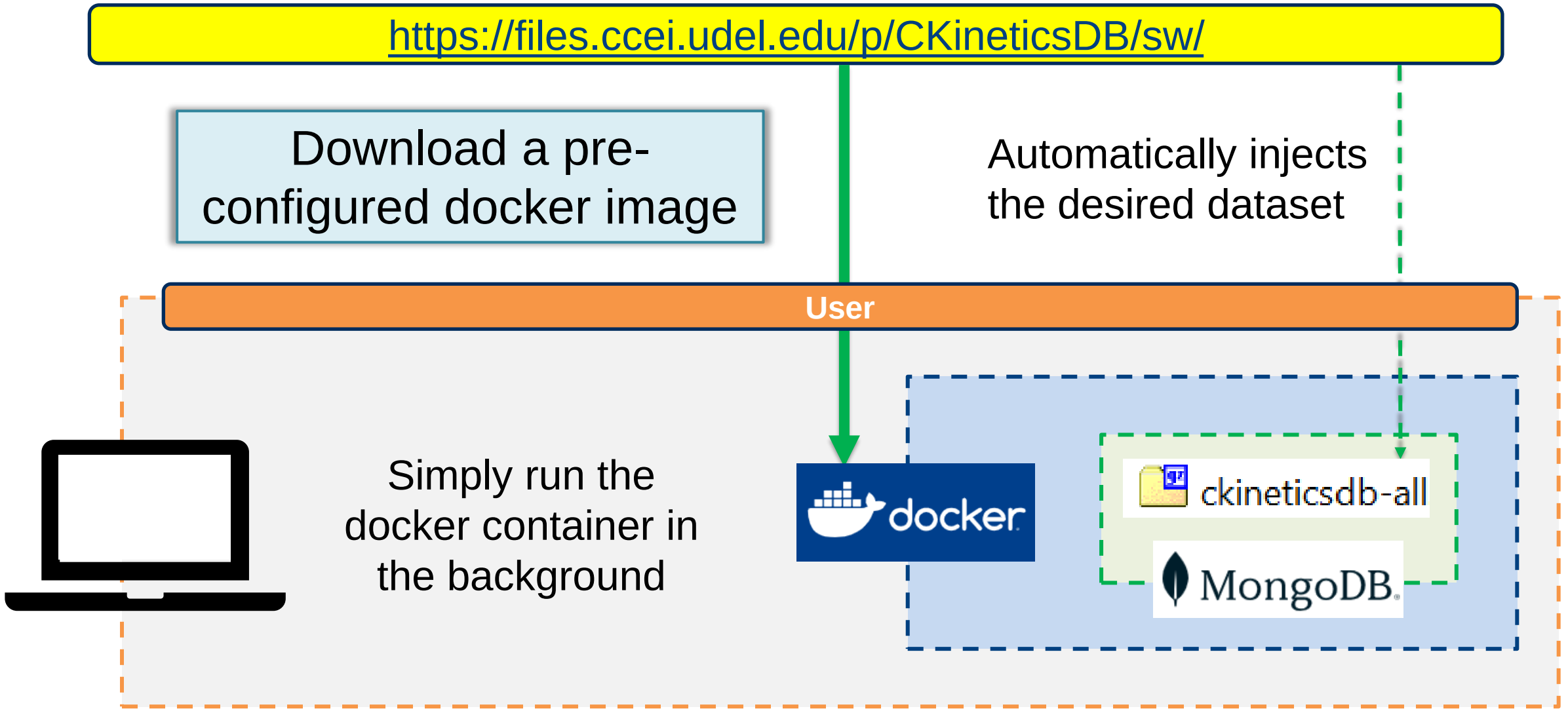


Run a MongoDB instance and unpack data

MongoDB.

ckineticsdb-all

# CKineticsDB Data Workflow

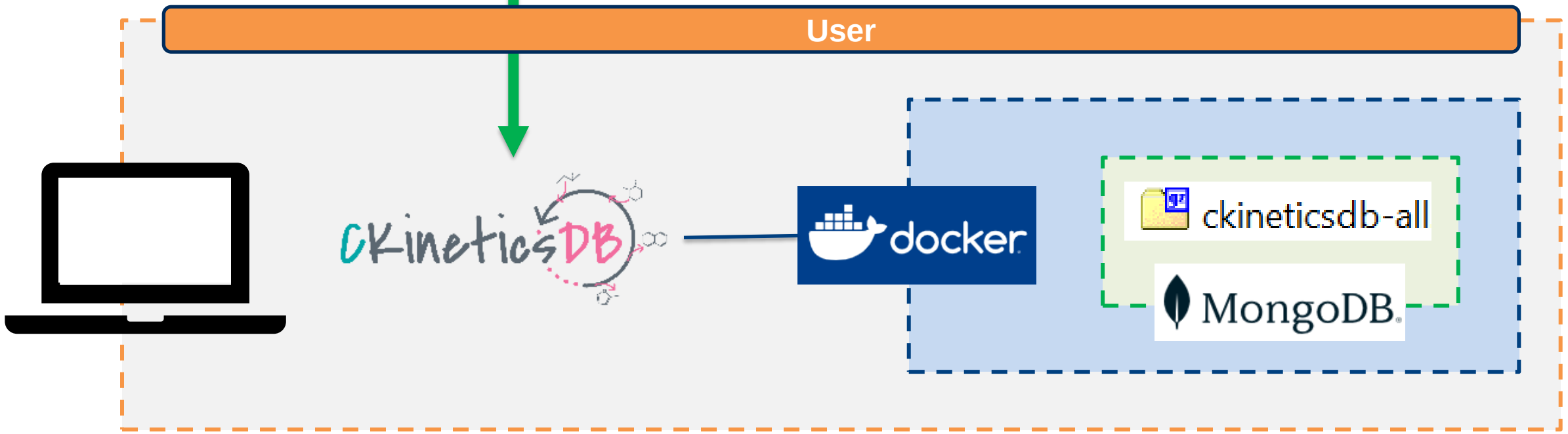


# CKineticsDB Data Workflow

https://files.ccei.udel.edu/p/CKineticsDB/sw/

Download and launch the desktop application

CKineticsDB automatically connects to the docker container



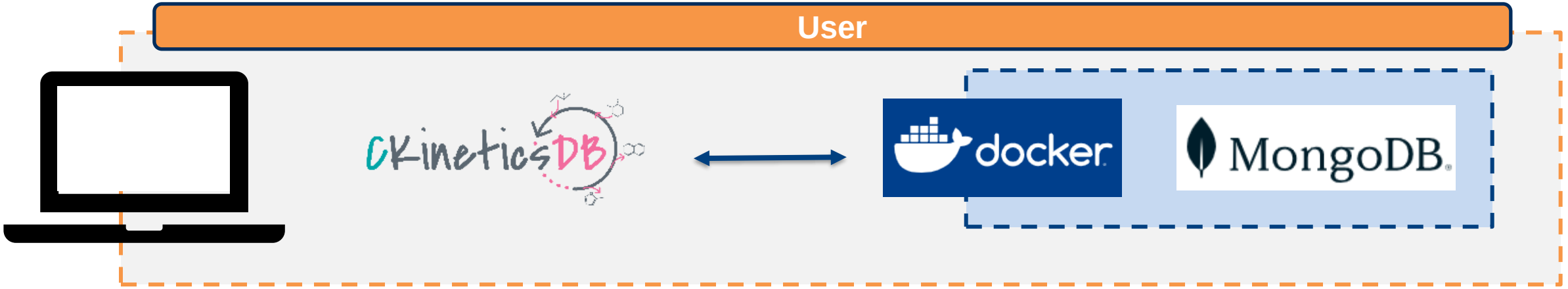
# Download CKineticsDB as a desktop application separate from the data

University of Delaware HPC

https://files.ccei.udel.edu/p/CKineticsDB/

Users don't need to -
- Learn MongoDB
- Run a local database server
- Worry about data persistence

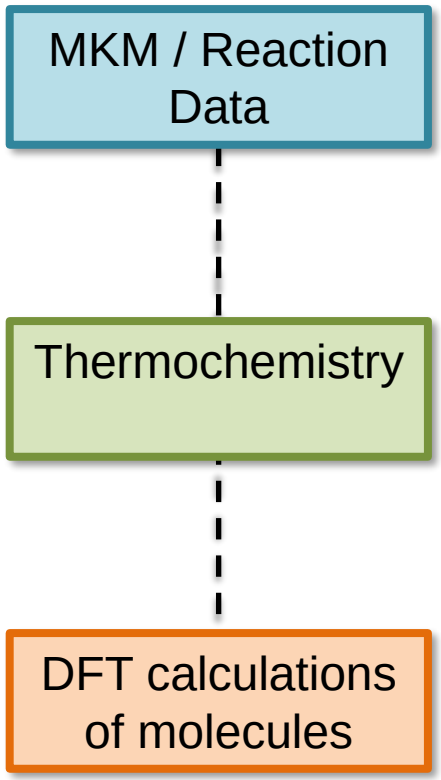
Users can -
- Connect CKineticsDB to any different database, local or remote
- Use CKineticsDB with their local data



# Summary

## Value from multiscale data management

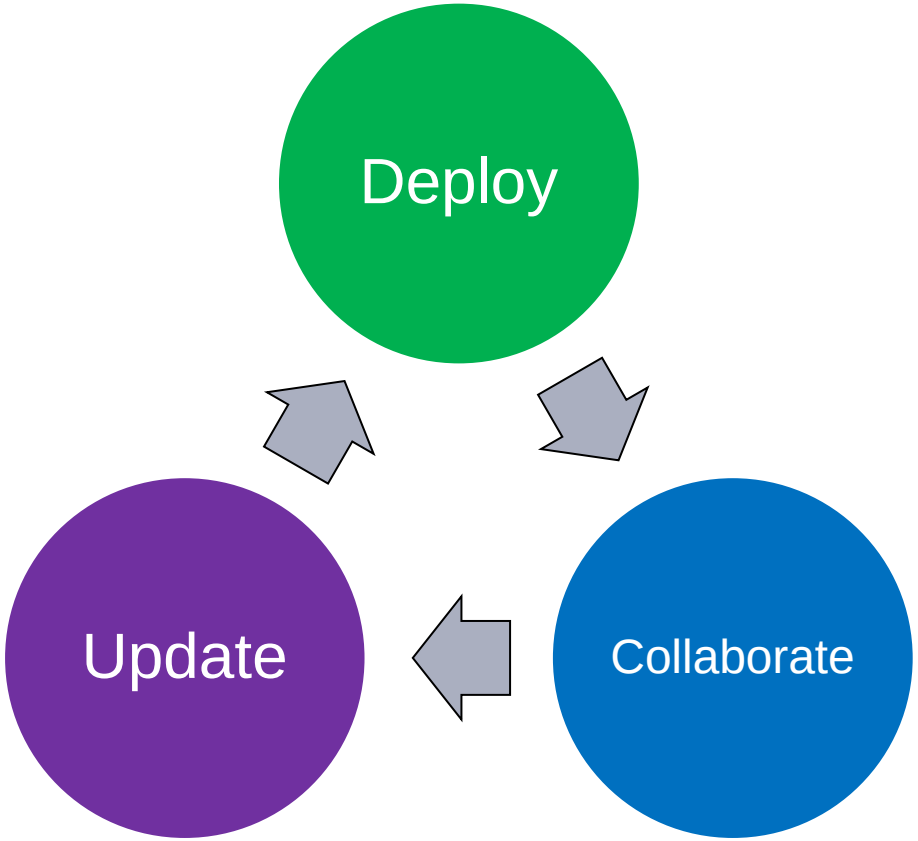
- Access reaction data and simulation files selectively from multiple scales
- Access scripts used to generate quantities from thermochemistry, kinetics, and complete MKMs
- Develop multiscale software and models
- Curate DFT calculations for computational diligence




## Software

- User friendly GUI and CLI
- Minimal learning curve
- Management of local data

# Collaboration and Future Development




- Update CKineticsDB for common needs of groups
- Cover more simulation software
- Build new data-based features
- Guide Onboarding

**Documentation**: https://github.com/VlachosGroup/ckineticsdb-documentation

# Acknowledgements

Dr. Jeffrey Frey, HPC
Kelly Walker, Logo

Department of Energy

RAPID
Transforming Process Industries

# Thank you

CKineticsDB