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

Siddhant Lambor, Sashank Kasiraju, Dionisios Vlachos

University of Delaware

ACS CATL Conference, Fall 2023
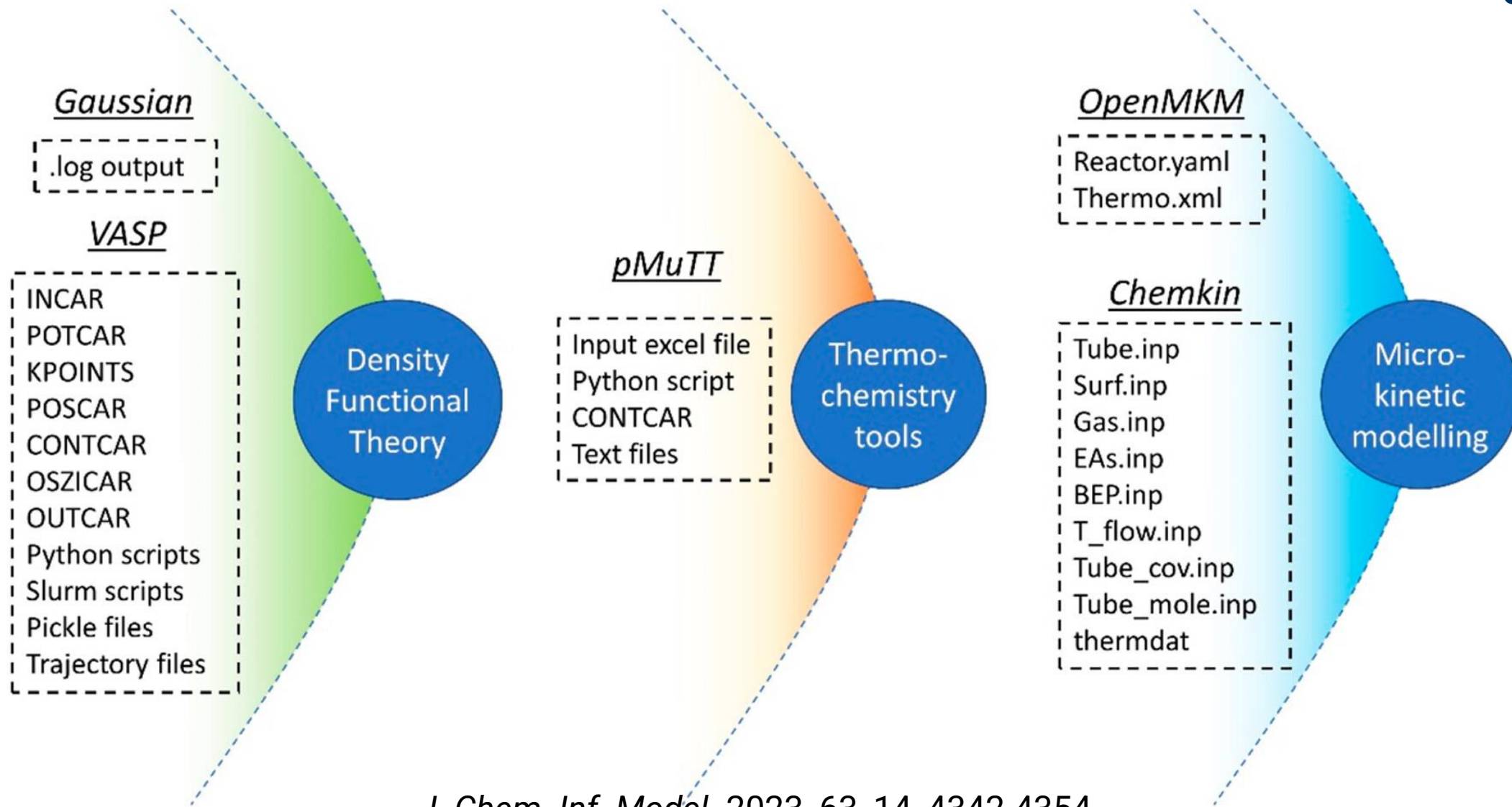
Open-Source Workshops

August 16, 2023

CKineticsDB: An Extensible and FAIR[1] Datahub for Multiscale Modeling in Heterogeneous Catalysis

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

# CKineticsDB stores the simulation files involved in multiscale modeling

Gaussian
- .log output

VASP
- INCAR
- POTCAR
- KPOINTS
- POSCAR
- CONTCAR
- OSZICAR
- OUTCAR
- Python scripts
- Slurm scripts
- Pickle files
- Trajectory files

Density Functional Theory

pMuTT
- Input excel file
- Python script
- CONTCAR
- Text files

Thermo-chemistry tools

OpenMKM
- Reactor.yaml
- Thermo.xml

Chemkin
- Tube.inp
- Surf.inp
- Gas.inp
- EAs.inp
- BEP.inp
- T_flow.inp
- Tube_cov.inp
- Tube_mole.inp
- thermdat

Micro-kinetic modelling

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

Chemical Kinetics Database

Multiscale Modeling in Catalysis

Microkinetic Modeling
Thermochemistry
Density Functional
 Theory

Unaltered files

Categorize

Curate

Organize

Extract Metadata

Store files, metadata, quality test results
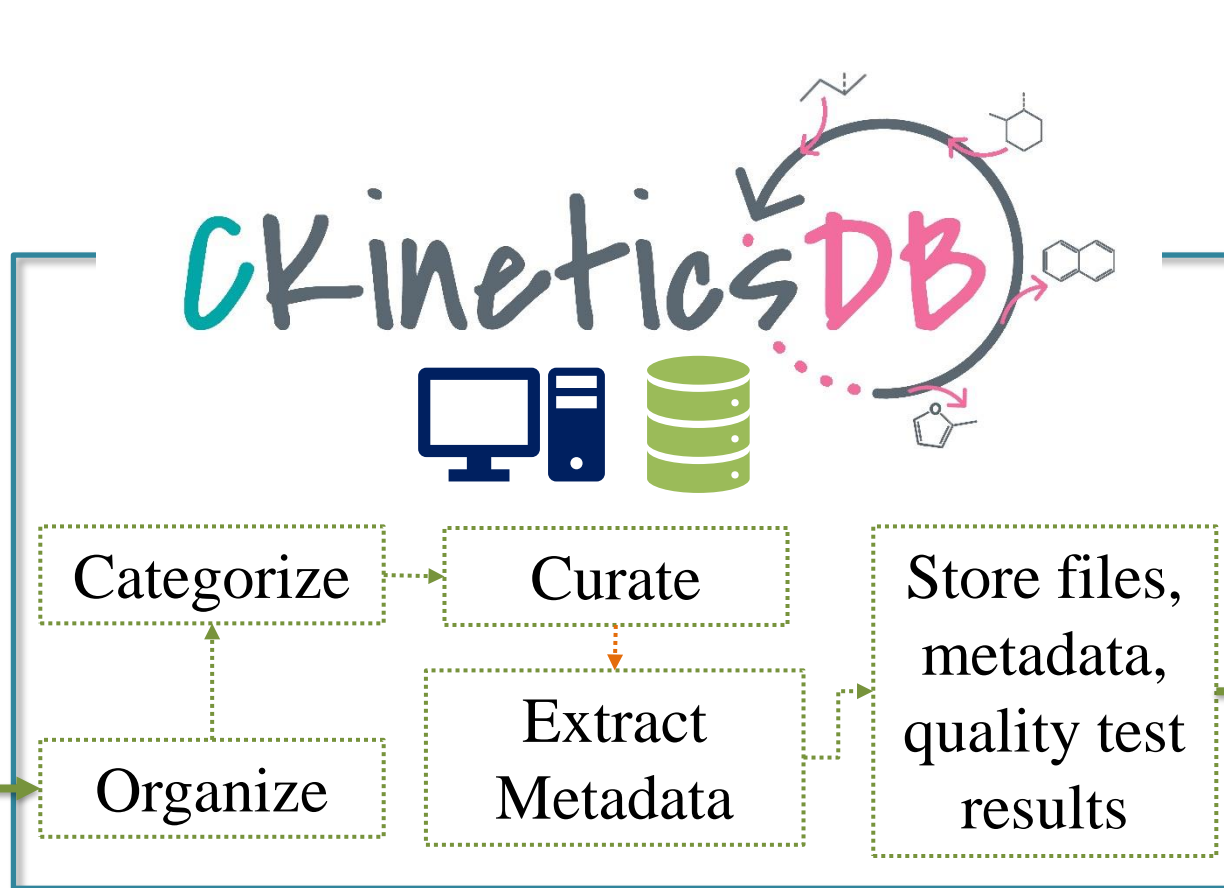
Microkinetic models, species- and reaction-specific simulations, and other output based on requested metadata

J. Chem. Inf. Model. 2023, 63, 14, 4342−4354

CKineticsDB top-level components

Database Management System

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

Frontend Python Software

Enables integrations with Vlachos Group in-house tools

Graphical User Interface

CKineticsDB

Select a task to begin

Quality   Download   Upload

UNIVERSITY OF DELAWARE ENERGY INSTITUTE   RAPID Transforming Process Industries   UNIVERSITY OF DELAWARE DELAWARE ENERGY INSTITUTE

Copyright © 2023, University of Delaware, All Rights Reserved.
Latest Release: 04-13-2023
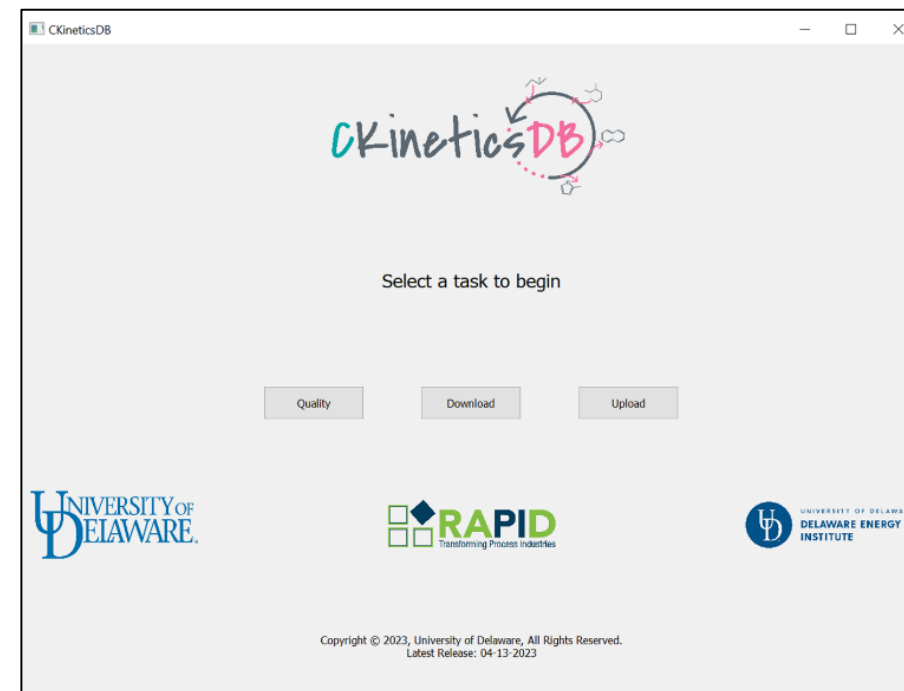
Command line interface

```
$ ckineticsdb download [OPTIONS]
$ ckineticsdb upload [OPTIONS]
$ ckineticsdb quality [OPTIONS]
```

Current data snapshot
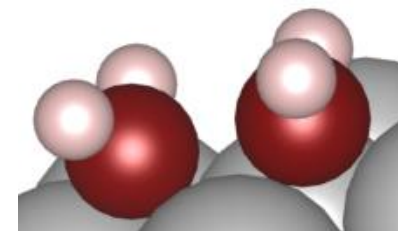
14000+ DFT calculations
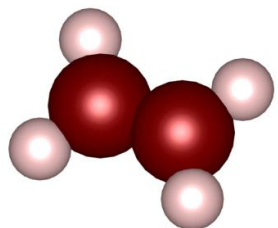
Gas Phase   Bulk structures   Adsorbates   Transition states
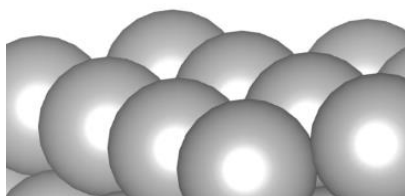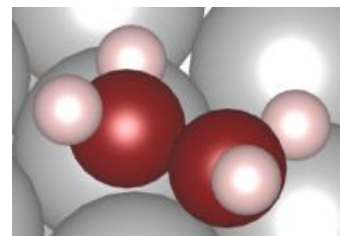
Catalysts

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

CKineticsDB data is openly shared online

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

Name
Parent Directory
ckineticsdb-all.data.gz
ckineticsdb-demo.dat..>
metadata/

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

ckineticsdb-all.data.gz
ckineticsdb-demo.dat..>
metadata/

Download dataset

User

Run a MongoDB instance and unpack data

MongoDB.
ckineticsdb-all

CKineticsDB Data Workflow

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

Download a pre-configured docker image

Automatically injects the desired dataset

User

Simply run the docker container in the background
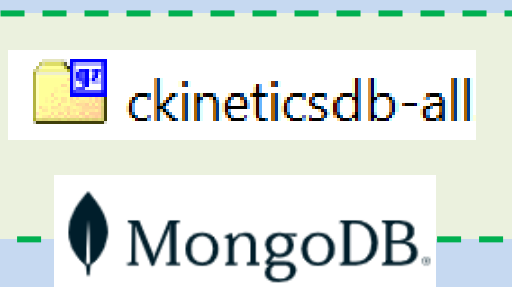
ckineticsdb-all

MongoDB

CKineticsDB Data Workflow

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

Download and launch the desktop application

CKineticsDB automatically connects to the docker container

User

Download CKineticsDB as a desktop application separate from the data

University of Delaware HPC

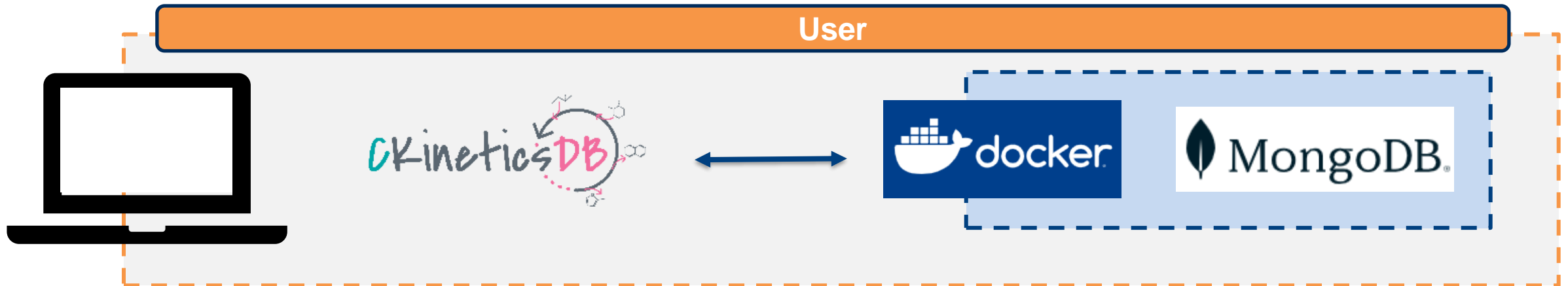https://files.ccei.udel.edu/p/CKineticsDB/

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

User

CKineticsDB Software Components

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

Name
↰ Parent Directory
📄 MAC_OS_application.zip
📄 Ubuntu_Application.zip
📄 Ubuntu_CLI_Applicati..>
📄 Windows_Application.zip
📄 ckineticsdb-database..>

Get access credentials by emailing to vkineticslab@udel.edu

Application.zip:
- Desktop application executable
- MongoDB server credentials file
- Template files and readMe file

-ckineticsdb-database.tar: Pre-configured docker image

Getting Started – Setting up the database

Detailed documentation at: https://github.com/VlachosGroup/ckineticsdb-documentation

1. Install Docker Desktop (from Docker website)

2.1. Download and load the Docker image:

Load the image: `> docker load --input <path_to>/ckineticsdb-database.tar`

2.2. Run a docker container and specify the data snapshot of choice to be injected

Options to inject different data snapshots – Default option

Running the container with no specific options will download and inject the demo data snapshot by default
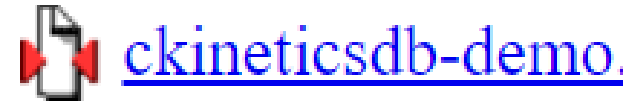
ckineticsdb-demo.

>>> *docker run --name ckineticsdb-db -p 27017:27017 ckineticsdb-database:latest*

Add a URL: Download and inject a specific data snapshot from the website

Add a local path: Inject a downloaded data snapshot locally

Syntax for each option is available in documentation

Demo

- Graphical user interface
- Parameters to browse and make selections in the GUI for downloading data

DFT data quality assessment

Select the directory of calculations to be tested

Select the parent directory for DFT assessment

Browse Home

Start Assessment

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)
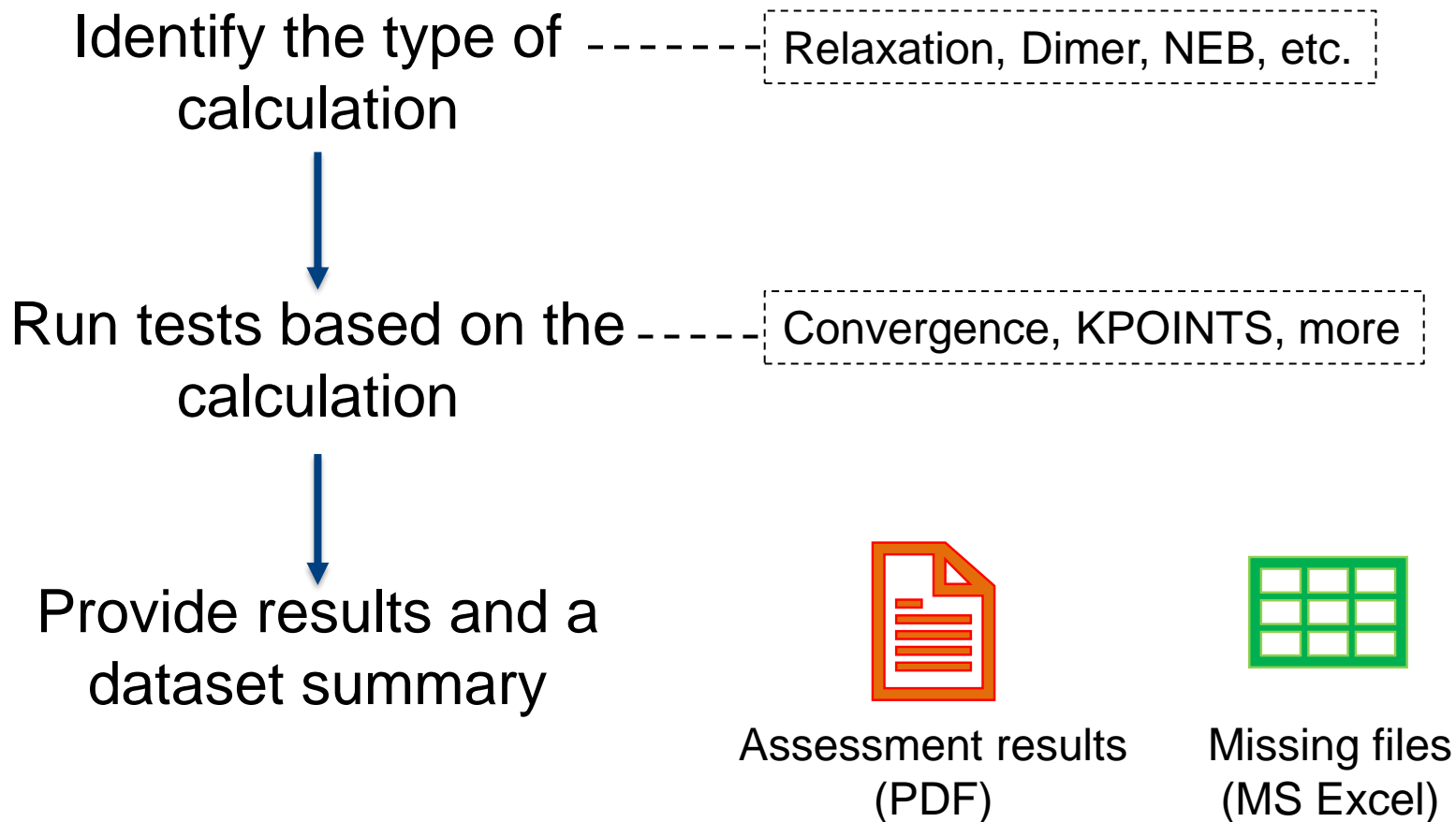
DFT Data Quality Tests and Output

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

Demo

- Data Quality Assessment Results

Conditions to upload a new dataset to CKineticsDB

Select the directory to be uploaded

| Browse | Home |

Start Upload

```
project_name/
├── DFT/
├── MKM/
├── pMuTT/
├── quality_assessment.xlsx
├── quality_assessment.pdf
└── readMe.xlsx
```

Organize the files as per the Data Organization Policy

↓

Run data quality assessment and include results

↓

Complete the readMe.xlsx file with materials information, software information, and other required details

↓

Upload

Collaboration and Future Development

Deploy

Collaborate
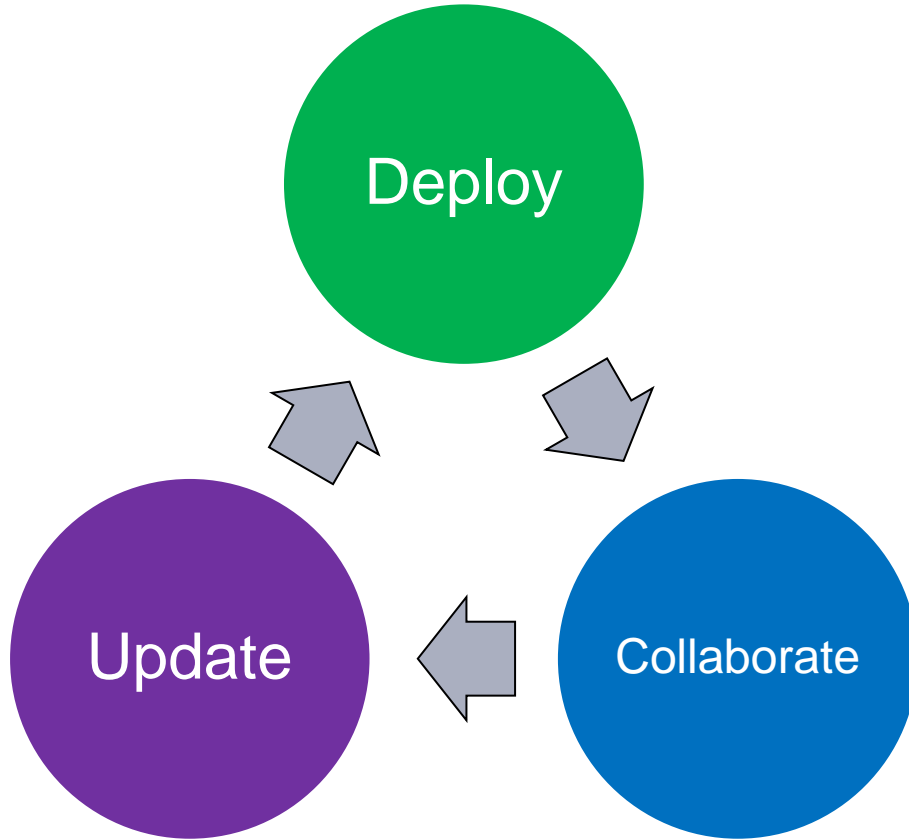
Update

- Update CKineticsDB for common needs of groups

- Cover more simulation software

- Build new data-based features

- Guide Onboarding

Acknowledgements

Dr. Jeffrey Frey, HPC
Kelly Walker, Logo

Department of
Energy

CKineticsDB Paper: https://doi.org/10.1021/acs.jcim.3c00123

Thank you

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