DIBBs: Creating a Digital Environment for Data-Driven Science (DEEDS)

PI: AC Catlin Co-PIs: A. Alam, M. Sepúlveda, J. Francisco, K. Gallant
Institutions: Purdue University, University of Pennsylvania, University of Minnesota

Award #: 1724728

The DEEDS Challenge

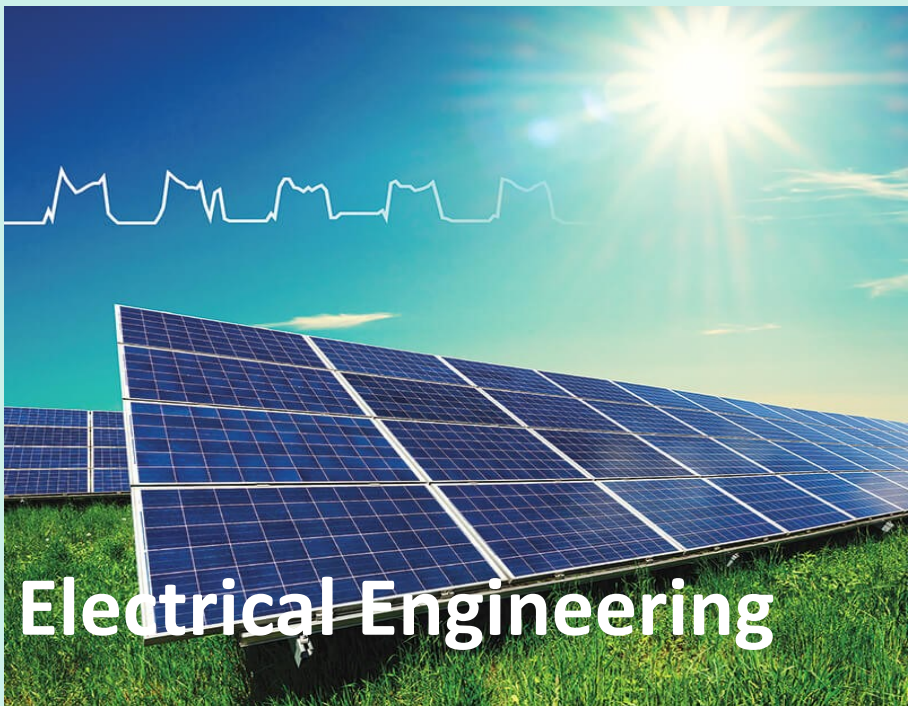Make our services so valuable, so indispensable that researchers from all science domains [*] will commit to using cyberinfrastructure to support their entire research process end-to-end – to benefit their own research effort and to promote more effective reuse and reinterpretation across scientific communities. With DEEDS, your data, computing and workflows come together in datasets that you build, use and share across your entire investigation lifecycle.

[*] including the vast number of scientists who have never used cyberinfrastructure of any kind to support their research.

DEEDS

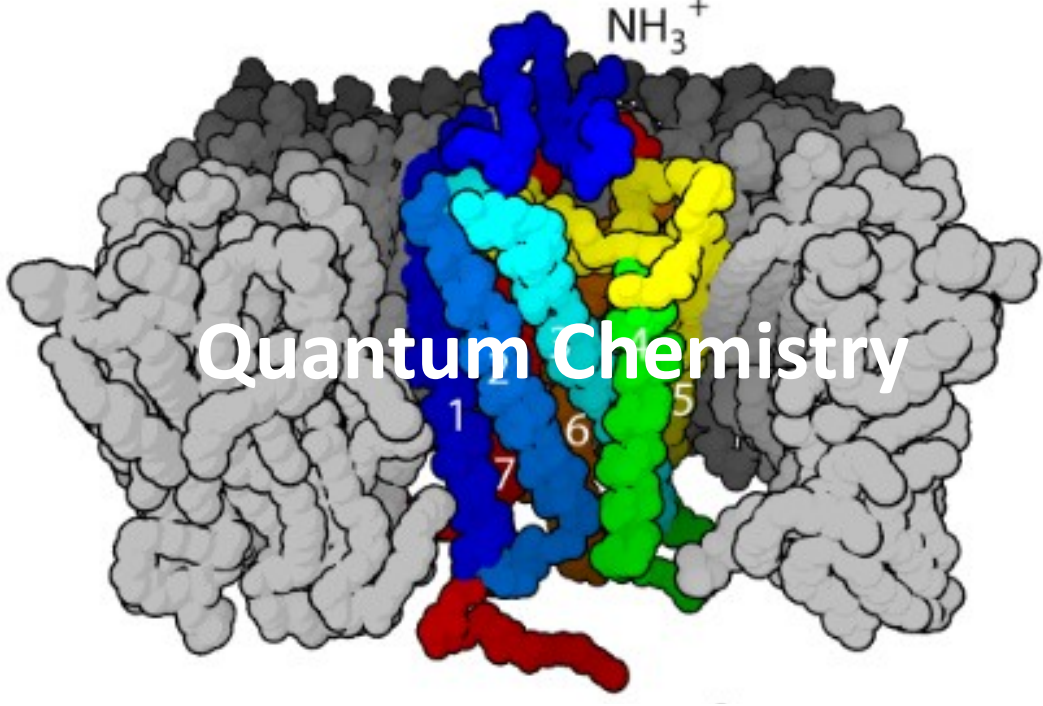### A Convergence Infrastructure

Innovation in data, computing, analysis & workflows that is effective across science domains

Electrical Engineering

diagnosis of solar PV system efficiency

file upload/classification, file import from duramat, jupyter curation tools, suns-vmp model research code & HPC execution– share, view, explore, disseminate

Quantum Chemistry

NH_3^+

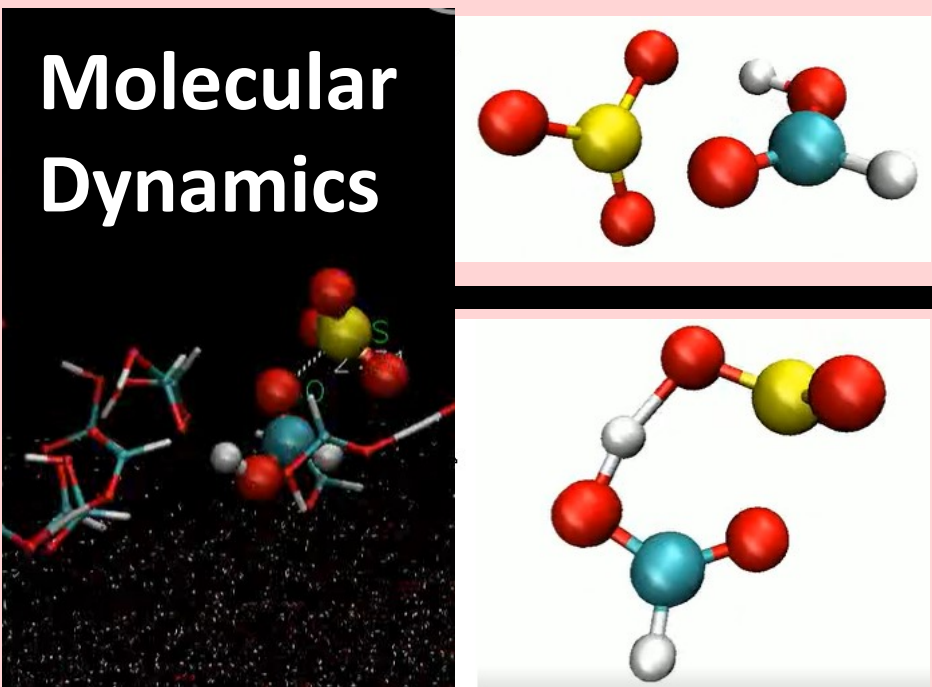vibrational theory of receptor protein activation

file upload/classification, Gaussian HPC execution, novel log post-processor for vibrational/thermochemical/charge data & depolarization plots– share, view, explore, disseminate

Environmental Science

PFAS toxicity reference values for aquatic exposure

complex data models as data tables with services (transformation of instrument output, validation) metadata, file upload/classification, ANOVA models, analytics– share, view, explore, disseminate

Molecular Dynamics

organic acid - SO3 reaction

file upload/classification, CP2K HPC execution, novel post-processor for distance & orientation computations, VMD videos– share, view, explore, disseminate

Global Ecology

tree species diversity in the Anthropocene

training & mapping datasheets to interactive data tables (11M rows), machine learning models, geotiff files & tile generation, spatial overlays & marker linkage to data tables– share, view, explore, disseminate
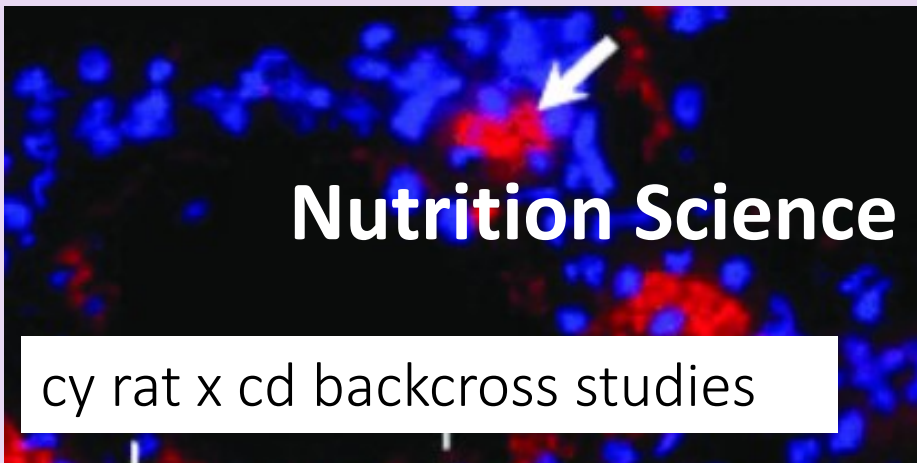
Civil Engineering

rural bridge health monitoring

NBI inspection datasheets to data tables with services (transformation, validation), file upload for 1M photos, figures, reports, analytics, jupyter trend analysis– share, view, explore, disseminate

Earthquake Engineering

building reconnaissance

Nutrition Science

cy rat x cd backcross studies

Biological Sciences

RNA seq muscle regeneration

Biological Engineering

engineering microbiomes

The DEEDS dashboard: collection, preservation, sharing, computing workflows, provenance, viewing, dissemination

Cases	Files	DataTables	Tools	Analytics	Explore
Project Organization	**File Repository**	**Complex Data Models**	**Execution & Workflows**	**Analysis & Visualization**	**Exploration & Publication**
Define	Upload	Define	Define or Import	R-based	DataView
Annotate	web, server, API	multi-dimensional, hierarchical, linked, cloned, propagated	research code, open source code, licensed packages, models ...	Data Frames	Technology
Metadata	Annotate			data tables, extract, merge, pivot, filter ...	view, navigate, search, filter, link, download ...
spatial, temporal, keywords, source, researchers, ...	Link	Update	Launch	R Toolkit	Visualization
	Metadata	by spreadsheet, interactive	select files, resources, auto-upload options ...	statistical computing, regression, time series ...	applications mapped to file classification
	user, timestamp, description, category ...	Metadata	Execute	Graphics	Exploration
	Classify	typing, labeling, applicability, customizable viewing	HPC facilities, local server, tracking ...	Reporting	linkage across all elements of dataset, customized data table viewer, hierarchy navigation, geospatial overlays
	reports, figures, media, data, geospatial, sensor, protocols, QAQC ...	Transform	Reproduce	Connect to Jupyter	FAIR Compliance
		Validate	compute workflow capture, traceability ...		reuse, reinterpretation

20,000 rural bridges in Nebraska

ecotoxicology file repository

berries & bone data tables for subject specimen measures

analyzing bacterial growth

exploring global tree biodiversity