Empowering Data-Driven Research through an Open, Accessible Data Infrastructure

Jamie V. Wittenberg
Head of Scholarly Communication; Research Data Management Librarian
Indiana University
The Dilemma

Libraries cannot provide researchers with sustainable, standardized access to licensed datasets for text and data mining.
The Dilemma

It is cost-prohibitive for most individual libraries to develop and implement infrastructure to provide access to licensed big data sets and large or unwieldy open data sets.
Many researchers who could benefit from text and data mining library-acquired & large open resources are only be able to do so via a graphical user interface
The Solution

CADRE is a cloud-based platform that provides secure access to library-licensed datasets and open, non-consumptive datasets.
The Datasets

Web of Science
Leading commercial bibliometric dataset

Microsoft Academic Graph
Leading open bibliometric dataset
The Solution

By sharing the cost of this solution across a large number of academic libraries, we are able to provide a superior solution at a lower cost to members, as well as a free service tier for non-members.
Project Partners

Indiana University Libraries

IU Network Science Institute

Big Ten Academic Alliance

This project was made possible in part by the Institute of Museum and Library Services LG-70-18-0202.
The Solution

CADRE features a graphical user interface; custom computational resources; and a space to share and store queries, algorithms, derived data, results, workflows, and visualizations.
The Solution

CADRE enables users to query across collections of datasets.
CADRE Project Leadership
The Platform

Secure, cloud-based platform

CADRE is a cloud-based platform that provides secure access to library-licensed datasets and open, non-consumptive datasets.

Affordable, sustainable service

By sharing the cost of this solution across a large number of academic libraries, we are able to provide a superior solution at a lower cost to members, as well as affordable tiered memberships to non-members.

Standardized access to data

Access standardized data & work from shared standards to ensure reproducibility.

Seeded with:
- Microsoft Academic Graph (208 million docs & 1.4 billion citations)
- Web of Science (73 million papers & 1.3 billion citations)

User-friendly GUI & other tools

CADRE enables users to query across datasets with a user-friendly GUI query-builder. Researchers can also leverage what other users create, including data-analysis tools, query results, and more.
How it works

- Log in to the platform from your affiliated institution using in-house credentials
- Access Microsoft Academic Graph and Web of Science datasets – over 220 million papers & 1.7 billion citations
- Use the user-friendly GUI query-builder to query big datasets in seconds
- Build your own data-analysis and visualization tools in Jupyter Notebook -- or leverage the data-analysis tools others create
- Reproduce queries, tools, derived data, research results, workflows, & visualizations—and utilize what others share in the platform

Sign in through federated login

Access major datasets

Query with GUI query-builder

Analyze results with shared tools

Reproduce your work
CADRE has opened a fellowship program for those working on coronavirus-related research called the RCSC Program. We are now accepting proposals to join the program.

Learn how to apply at: cadre.iu.edu/rcsc
As an RCSC researcher, you can…

- Access & query across the COVID-19 Open Research Dataset (CORD-19), Web of Science, and Microsoft Academic Graph on CADRE
- Use CADRE’s GUI querybuilder, Jupyter Notebook coding environment, and cloud–computing resources
- Receive intensive technical support for your work
- Present your research in our fellows webinar series
Utilizing Data Citation for Aggregating, Contextualizing, and Engaging with Research Data in STEM Education Research

Researchers: Michael Witt, Loran Carleton Parker, Ann Bessenbacher
Affiliation: Purdue University
Understanding citation impact of scientific publications through ego-centered citation networks

Researchers: Yi Bu, Chao Min, Ying Ding
Affiliation: Indiana University Bloomington and Nanjing University
MCAP: Mapping Collaborations and Partnerships in SDG Research

Researchers: Jane Payumo, Devin Higgins, Scout Calvert, Guangming He, Anusha Manjunatha
Affiliation: Michigan State University
Measuring and Modeling the Dynamics of Science Using the CADRE Platform

Researchers: Russell Funk, Michael Park, Thomas Gebhart, Britta Glennon, Julia Lane, Raviv Murciano-Goroff, Matthew Ross, Jina Lee, Erin Leahey

Affiliation: University of Minnesota, University of Pennsylvania, New York University, Boston University, University of Arizona
The global network of air links and scientific collaboration – a quasi-experimental analysis

Researchers: Katy Börner, Adam Ploszaj, Lisel Record, Bruce Herr II
Affiliation: Indiana University Bloomington and University of Warsaw
Comparative analysis of legacy and emerging journals in mathematical biology

Researchers: Marisa Conte, Samuel Hansen, Scott Martin, Santiago Schnell
Affiliation: University of Michigan and University of Michigan Medical School
Systematic over-time study of the similarities and differences in research across mathematics and the sciences

Researcher: Samuel Hansen
Affiliation: University of Michigan
Assessing the rise of China as a scientific nation

Researchers: Caroline Wagner, Xiaojing Cai
Affiliation: Ohio State University
Fellows Webinar: Mapping Collaborations and Partnerships in SDG Research

April 29, 2020 - April 29, 2020
Our fellowship team from Michigan State University will present the research they developed using CADRE in this webinar.

See Details

Fellows Webinar: Understanding Citation Impact of Scientific Publications Through Ego-Centered Citation Networks

June 17, 2020 - June 17, 2020
Our fellowship team from Indiana University Bloomington and Nanjing University will present the research they developed using CADRE in this webinar.

See Details
What’s Next?

- Sustainability and governance model
- Expand CADRE membership
- Add new datasets
Contact us

https://cadre.iu.edu

cadre@iu.edu

@CADRE_Project