



**CADRE**  
Collaborative Archive & Data Research Environment

Microsoft®  
**Research**



# Open alpha release

---

Supported by U.S. Institute for Museum and Library Services

Version 0.1.0-alpha



# Program overview

- **The CADRE project**  
(Valentin Pentchev)
- **Demo 01: Data access and Reproducible Packages**
  - **Query-interface, Notebooks, Packages**  
(Matthew Hutchinson)
- **Demo 02: Networks and Visualizations**
  - **Query-interface, Notebooks**  
(Filipi Silva)
- **Upcoming CADRE events**  
(Xiaoran Yan)

# Demo Resources

- <https://cadre.iu.edu/>
- <https://cadre.iu.edu/gateway/>
- <https://iuni-cadre.github.io/AlphaDemo/>
- <https://github.com/iuni-cadre/AlphaDemoCode>

# Acknowledgments

We gratefully acknowledge that funding and data for this workshop was provided by:

- U.S. Institute for Museum and Library Services (IMLS)  
LG-70-18-0202



- Web of Science Group (Clarivate Analytics), Microsoft Research





**CADRE**  
Collaborative Archive & Data Research Environment

Microsoft®  
**Research**



# The CADRE project

---

Valentin Pentchev  
[vpentche@iu.edu](mailto:vpentche@iu.edu)



# CADRE Leadership



# Partners



-  University of Iowa Libraries
-  University of Michigan Libraries
-  Michigan State University Libraries
-  University of Minnesota Libraries
-  Ohio State University Libraries
-  Penn State University Libraries
-  Purdue University Libraries
-  Rutgers University Libraries



- Health Partners
- Pervasive Technology Institute
- Midwest Big Data Hub
- South Big Data Hub
- West Big Data Hub
- Microsoft Research
- Web of Science Group

# CADRE Project - The Dilemma

- Libraries cannot provide researchers with **sustainable, standardized access** to licensed datasets for text & data mining
- 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 resources, lack programming skills and would only be able to do so via a **graphical user interface**

# CADRE Project - The Solution

- CADRE is a cloud-based platform that will provide **secure access to library-licensed datasets** and open, non-consumptive datasets
- 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
- CADRE features a graphical **user interface**; standardized , multiple data formats; shared and custom **computational resources**; and a **space to share and store** queries, algorithms, derived data, results of analyses, workflows, and visualizations.

# CADRE Project - Indiana University Network Science Institute (IUNI)



- IUNI – <http://iuni.iu.edu> – a unique startup in an established academic institution
  - A cross-campus, transdisciplinary institute that brings together faculty who engage in network research from various scientific fields
- IUNI's mission
  - To strengthen the theories, methods, analytic tools, and practice of network science, and to foster collaborative, interdisciplinary network science approaches to understanding and improving the complex challenges of our world
- IUNI's Teams
  - A team of IT professional
  - A team of research scientists



# CADRE Project – Goals

## Identify Constituents' Needs

- Understanding **users' needs** and expectations
  - User stories, Product Owner Council meetings, **Communication**
- **Informatics/computer science researchers** and labs
  - **APIs, Notebooks**, Access to **Raw Data** and **Cloud Compute Resources**
- **Science of Science** community
  - **Interface Access to Databases** and Cloud Native Technologies
- **Library and research** community outside of computer science
  - **Web Interface** guiding Query Building and **suggesting the most appropriate backend technology** on a case by case basis

# CADRE Project – Goals

## Identify the Proper Technology

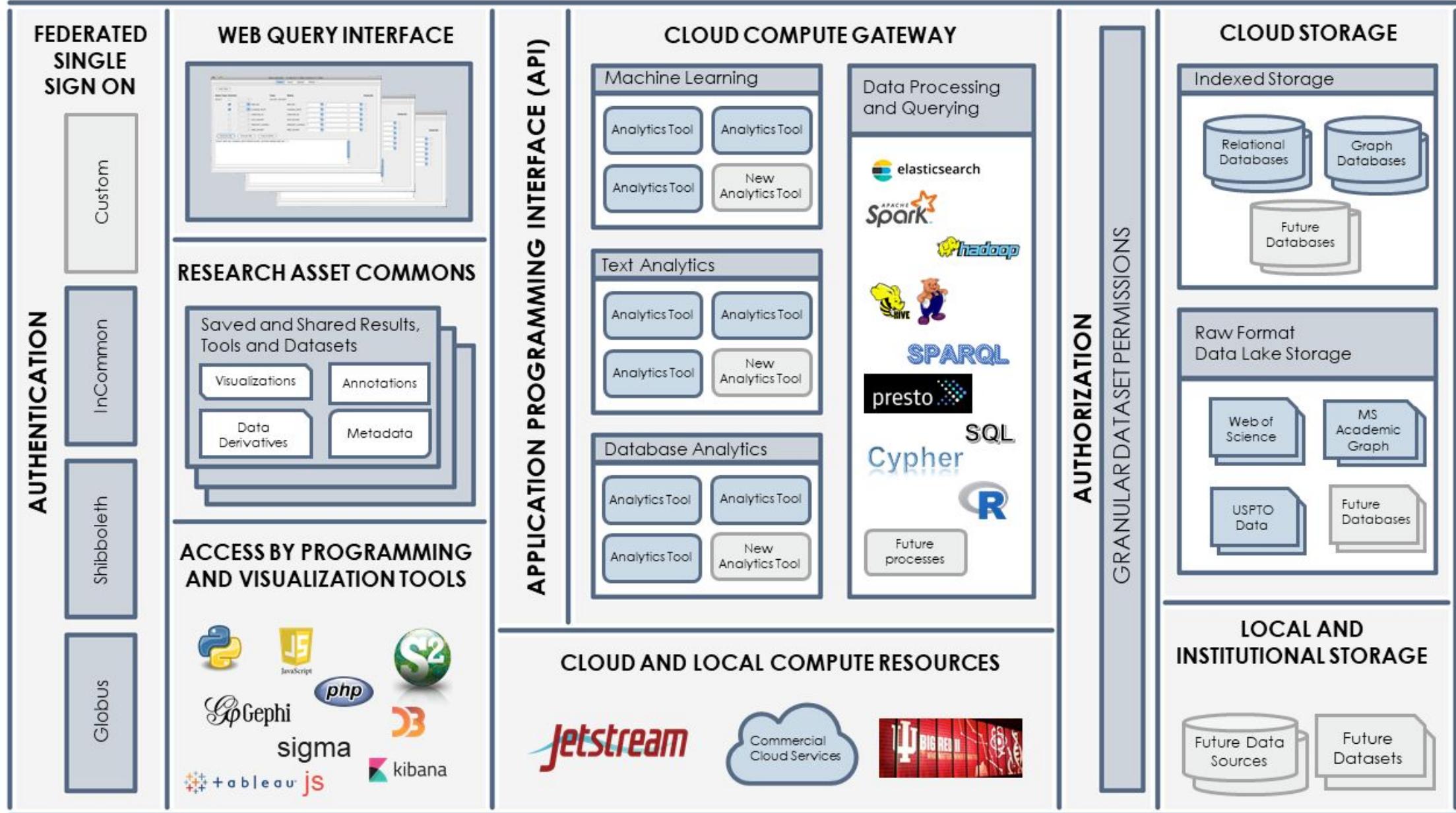
- **Raw data access**
  - Access to **XML, JSON, CSV** etc. files in their native form. **Containerized** tools and packages
  - Access to data using **cloud native technologies** like **U-SQL** and **Athena/Glue**
  - Access to cloud **distributed computing** using **Databricks** and **SPARK** on **HDInsight** and **EMR**
- **Database access**
  - Researching on currently available **cloud and serverless Relational Database** implementations for each dataset and query type
  - Researching on currently available **Graph Database** implementations for each dataset and query type. Currently evaluating **Neo4j, Tiger Graph, JanusGraph, cloud native and in-memory** alternatives
- **Web interface**
  - **Guided Query Building**
  - Ability to **suggest the most appropriate technology** on a case by case basis
  - **User control** over execution and use of resources

# CADRE Project – Goals

## Research Asset Commons

- **Federated Login**
  - Access from **any affiliated institution** using single sign on ( CILogon, inCommon, Shibboleth etc.)
  - **Restricted access to proprietary resources**, based on login credentials
- **Collaboration**
  - Ability to **save and share** with specific users, community or the public **metadata, queries, results, annotations, visualizations, algorithms, code, containers and virtual machines**
  - Community building and collaboration based around **same data access privileges** and goals
- **Reproducibility, Replicability, Provenance and Transparency**
  - Use of the same, **well documented original datasets**
  - **DOIs** identifying any and every data change or permutation
  - **Saved and shared** workflows a pipelines trough **Packages and Containers**
  - **Ability to publish using unique identifiers** leading back to Research Asset Commons

# SHARED BIGDATA-GATEWAY FOR RESEARCH LIBRARIES (SBD-GATEWAY)





Microsoft®  
Research



# Data access and reproducible packages

---

Matthew Hutchinson



# Demo 01

---

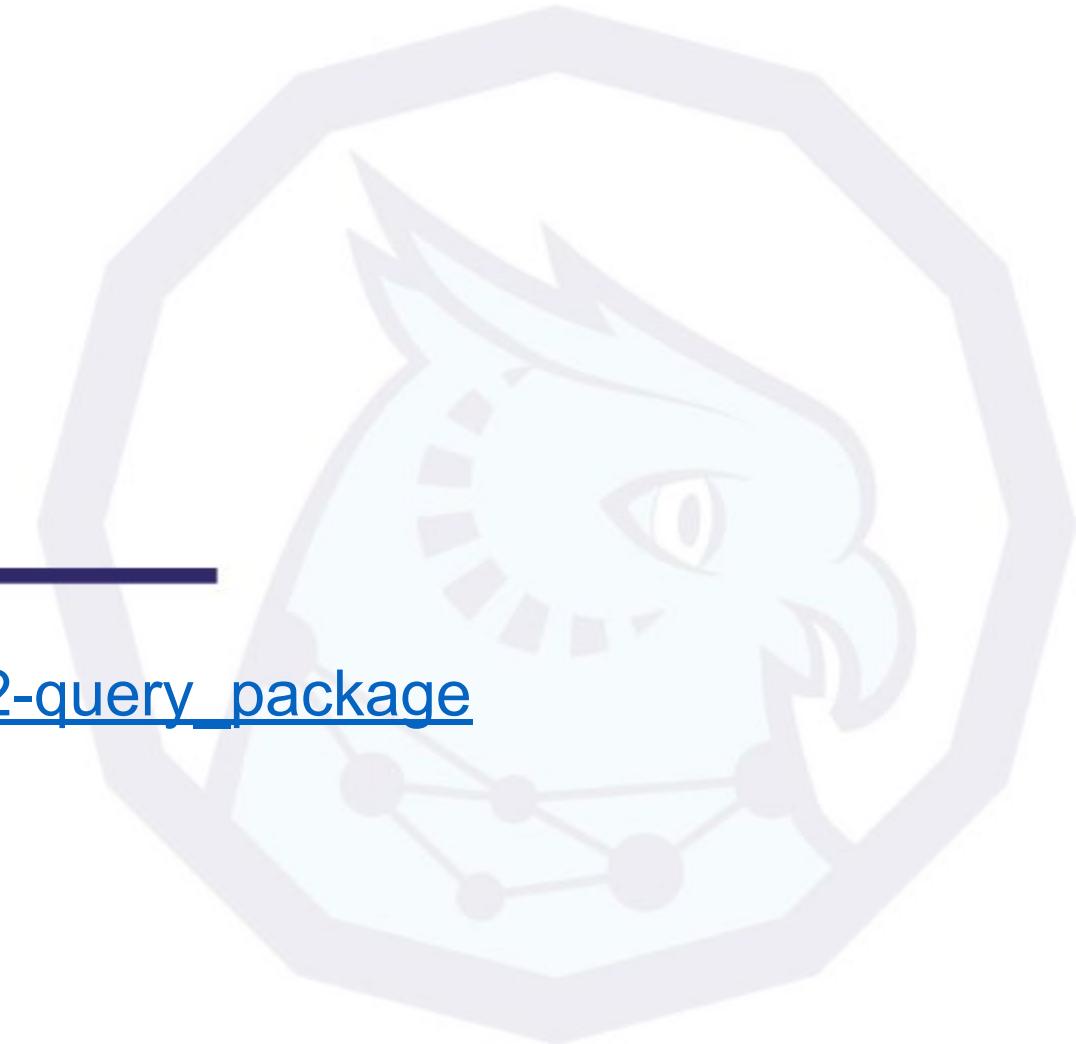
[https://iuni-cadre.github.io/AlphaDemo/02-query\\_package](https://iuni-cadre.github.io/AlphaDemo/02-query_package)



# Questions?

---

[https://iuni-cadre.github.io/AlphaDemo/02-query\\_package](https://iuni-cadre.github.io/AlphaDemo/02-query_package)





Microsoft®  
Research



# Networks and visualizations

---

Filipi Silva



# Demo 02

---

<https://iuni-cadre.github.io/AlphaDemo/03-visualization>



# Questions?

---

<https://iuni-cadre.github.io/AlphaDemo/03-visualization>





Microsoft®  
Research



# Outreach and upcoming events

---

Xiaoran Yan  
[yan30@iu.edu](mailto:yan30@iu.edu)



# CADRE's solution

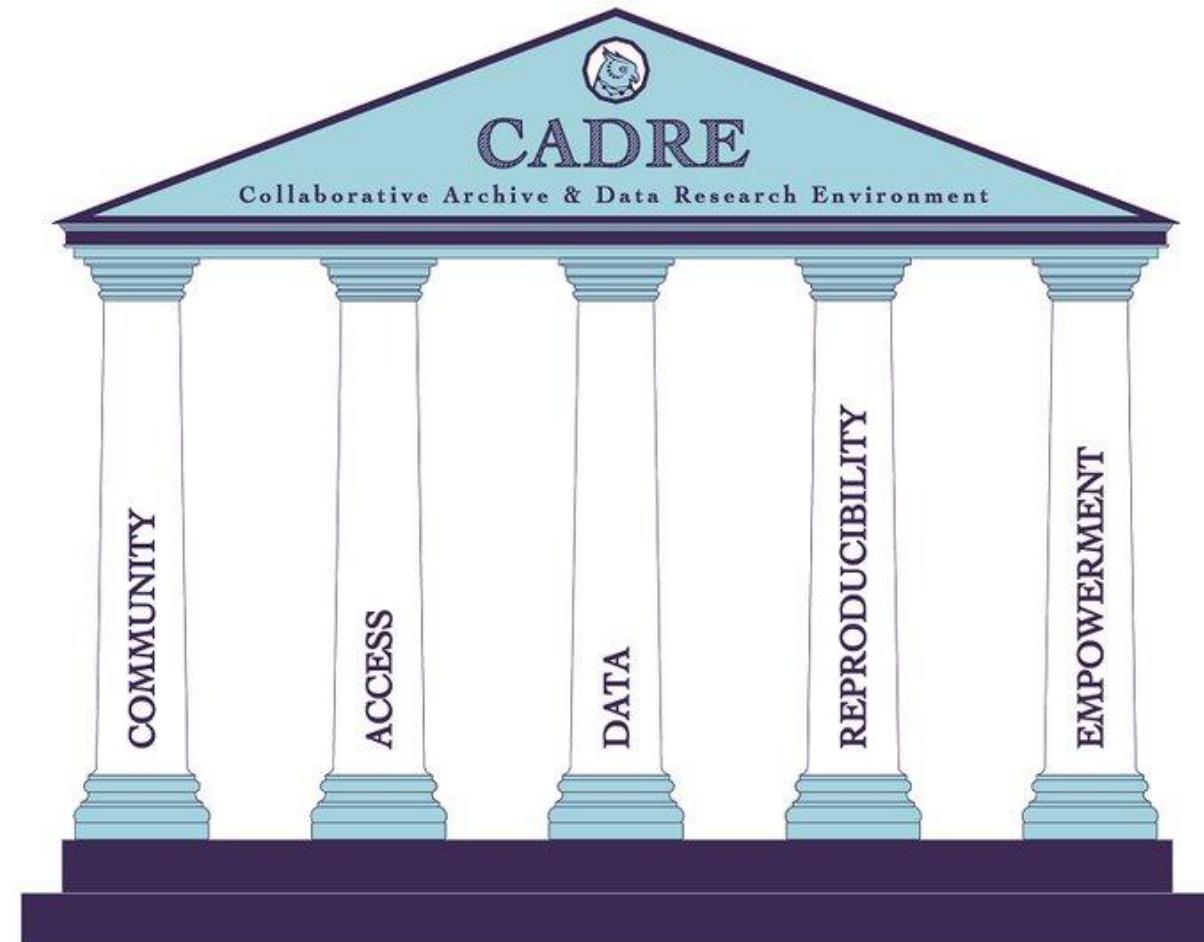


Access over 220 million scientific publications

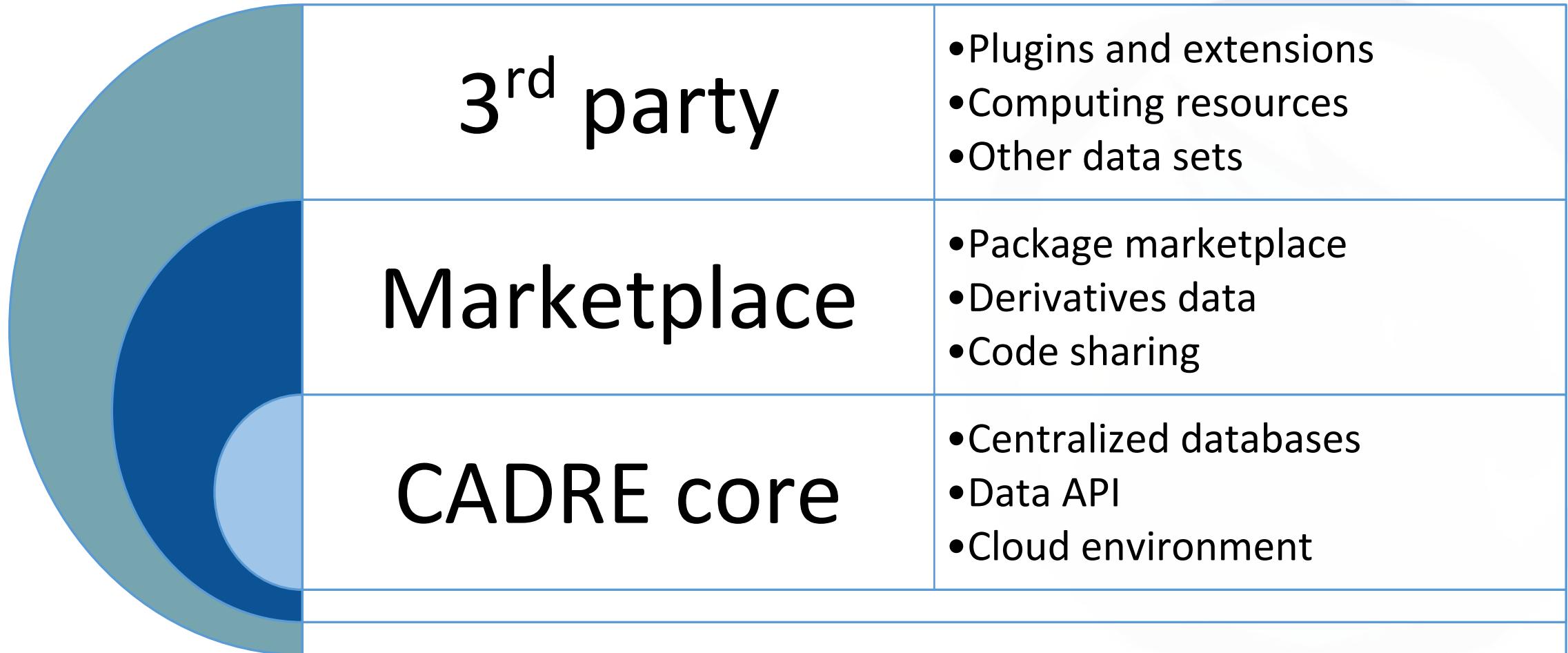
Effortlessly query data and analyze results

Reproduce research & leverage tools

# CADRE's community building effort



# The CADRE ecosystem



# CADRE events in 2020

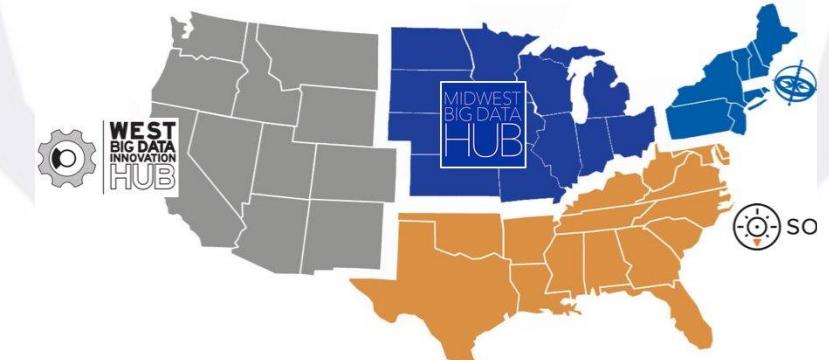
Feb. 2020



Mar. & Apr  
2020

May 2020

- CADRE webinar/seminar series start
- Fellowship team presentation
- Data provider presentation
- CADRE council meeting
- Fellow workshop at all hubs meeting



# CADRE fellowship project presentations

---

<https://cadre.iu.edu/work-with-us/cadre-fellowship>



# **Assessing the rise of China as a scientific nation**

Researchers: Caroline Wagner, Xiaojing Cai

Affiliation: The Ohio State University



# MCAP: Mapping Collaborations and Partnerships in SDG Research

Researchers: Jane Payumo, Devin Higgins, Scout Calvert, Guangming He

Affiliation: Michigan State 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



# CADRE milestones in 2020

- 
- Feb. 2020
- Version 0.1.0-alpha release
- May 2020
- Database update/migration
  - Package publication and interface update
- Oct. 2020
- Persistent beta version release
  - All major components ready
  - All three datasets ready

# Contact Us



<https://cadre.iu.edu>



[cadre@iu.edu](mailto:cadre@iu.edu)



@CADRE\_Project

# We need your Feedback

## Dashboard

Quick Start

Jupyter Notebook

Open a notebook environment to perform complex analytics on CADRE query results

Top Packages

demo02\_data\_package

By: CADRE Team

Created On: Fri, 07 Feb 2020 21:10:46 GMT

Tool:

demo02\_data

query2\_xnet\_package

By: CADRE Team

Created On: Fri, 07 Feb 2020 21:10:46 GMT

Tool:

query2\_xnet

[Send Feedback](#)



# Demo Resources

- <https://cadre.iu.edu/>
- <https://cadre.iu.edu/gateway/>
- <https://iuni-cadre.github.io/AlphaDemo/>
- <https://github.com/iuni-cadre/AlphaDemoCode>