Collaborative Archive & Data Research Environment (CADRE):

A Big Data Solution for Research Libraries

Valentin Pentchev[1], Robert Van Rennes[2], Jamie Wittenberg[3], Patricia L. Mabry[4], Xiaoran Yan[1]

[1]Indiana University Network Science Institute, [2]Big Ten Academic Alliance, [3] Indiana University Libraries, [4]HealthPartners Institute

CADRE's Initial Datasets

Web of Science: a leading commercial dataset with 63M papers and 1.2B citations.

Microsoft Academic Graph: an open bibliometric dataset containing 208M documents and 1.4B citations.

United States Patent and Trademark Office: 9M patent application documents.

Project Overview

The Collaborative Archive and Data Research Environment (CADRE) is a shared big data gateway for research libraries. This two-year project is being funded by the Institute for Museum and Library Services (IMLS) though a grant awarded to Indiana University.

The project was developed with a cloud-based, extendable cyberinfrastructure for sharing large academic library data resources with a growing community of scholars.

The gateway is currently seeded with open and licensed bibliographic datasets, (available in a variety of formats) and provides a suite of computational tools and a space for sharing and reusing analytic code and outputs.

The Big Ten Academic Alliance and nine of its member institutions are supporting the development of CADRE with additional support coming from the Web of Science Group, Microsoft Research, Jetstream as well as the Midwest, South, and West Big Data Hubs.

Contribute

We need USER STORIES for large scale bibliometric research.

Help us build CADRE infrastructure by submitting a user story: go.iu.edu/288v
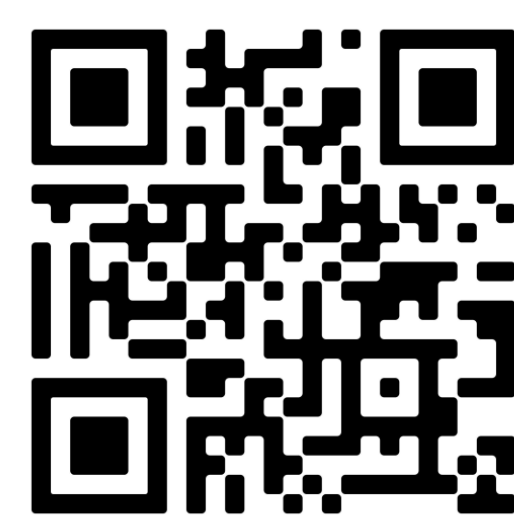
Become a CADRE Fellow and collaborate with our team: **cadre.iu.edu/work-with-us**

Design

Authentication: A federated security login system to utilize each institution's proprietary authentication system.

Research Asset Commons: A shared repository to save, store, & reproduce algorithms, data subsets, derived results, tools and methods.

Compute Gateway: A modular collection of tools, applications and technologies for research on cloud- based & local systems.

Cloud Storage: Raw data, relational and graph database storage connected to cloud and local compute resources.

Project Goals

Create a user community of libraries and researchers who can provide input to CADRE design and derive benefit from it.

Develop a sustainable funding model.

Deliver an initial national scale platform.

Release CADRE as an extensible, open-source platform, welcoming collaboration and future development

Engage the user community to identify and prioritize additional features, datasets and other improvements for CADRE.

Promote sharable and reproducible workflows, data derivatives and data standards.

Find new institutional partners and as well as Fellows to help build out CADRE.

Contact: cadre@iu.edu

Twitter: @CADRE_Project

Info: https://cadre.iu.edu

Summary

Academic libraries are challenged to provide sustainable, affordable, and standardized data and text mining cyberinfrastructure for large datasets.

CADRE is a cloud-based platform solution for making licensed, big data sets as well as open and non-consumptive data sets accessible with appropriate security, stewardship, and storage in place.

By sharing the cost of this solution across a large number of academic libraries, we will be able to provide a superior solution at a lower cost to participants. A free tier of basic services for public access will be offered.