MAP Pilot
Machine Actionable Plans

University of California, Riverside

MAP Pilot Institution Project Plan

Description

Description of the pilot project, including the processes, tools, or changes that will be tested.

At UCR there are two aspects to our maDMSP pilot project: technical integration, and supporting social and communication workflows. Both center around the question, "How can we reduce the burden of data sharing and management activities on the researcher, research support staff, and other stakeholders?"

Technical Integration to Deploy Resources

The technical project partners the UCR Library, Research Computing, and Technical Operations in the Computer Science and Engineering (CSE) Department to integrate maDMSPs with an automated framework for deploying computing resources for research projects. Digital data, such as that in computer science research, is often generated computationally. Metadata related to this data which should be captured in a DMSP include required software, containers, and computational resources. Technical Operations in UCR's CSE department is already working on an automated workflow to deploy these resources based on IaC (Infrastructure as Code) approaches. We will examine the components required to create a pipeline to connect maDMSPs with this workflow in order to deploy computational resources which support research. We will first test deploying resources (containers) on Google Cloud Services, and then test the use and integration of AI/ML to identify relevant text beyond entity extraction for the deployment of these resources.

Another campus process and support mechanism which will be examined as part of this technical project are the DMSP materials and templates provided to PIs as part of proposal development. This aspect will address the question "if current DMSPs do not provide enough information to inform technical integration, how may we gather the information?" This may be through adapting DMSP boilerplate and template language, demonstrating the benefit to the researcher, gathering the information through another mechanism while pushing the info to a maDMSP, and/or other undiscovered methods.

Communication and Compliance Workflows

The second project supports social, communication, and compliance workflows, connecting many units on campus using maDMSPs. Initial feedback from stakeholder groups indicate that areas of possible integration are with our proposal management system to allow ingestion of DMSPs created locally, and to coordinate and track research outputs, including data-related and other publications. This project

would coordinate with key stakeholders to gain feedback on key points in the research cycle to implement automated notifications and report creation using the maDMSP as a connecting platform.

Objectives and Goals

Objectives and goals of the pilot project.

Technical

- The main goal of the technical project is to test the integration of maDMSPs with deploying computational resources to support data creation, management, and access in preparation for and during the active research phase of a research project.
- Starting with Google Cloud Services, we will identify the metadata required for automated deployment and access.
- Once identified, we will further identify detailed metadata required in order to safely use as a method of deployment, such as description, data security classification level, required dependencies, and other possible descriptors.
- We will cross reference these metadata with elements of data management plans, and maDMSPs to determine the use of the maDMSP as a source of truth and historic log of data management activities.

Communications & Compliance

- The main goal of the communication & compliance project is to test the ability to identify maDMSPs which relate to sensitive data, such as Intellectual Property, Human Subject data, and other data which need to be secured.
- We intend to test
    - How to identify such data,
    - What needs to be redacted for a maDMSP to be shared publicly,
    - How to notify campus personnel when appropriate, such as IT for security or RED for IP.

Pilot Team Composition

- Katherine "Kat" Koziar, Data Librarian, UCR Library
- Zhiwu Xie, Assistant University Librarian for Research & Technology
- Chuck Forsyth, Associate Director of Research Computing, UCR ITS
- Victor Hill, Director of Research Computing, Bourns College of Engineering
- Sherie Donahue, Proposal Development Officer, Research and Economic Development

This project is managed by California Digital Library (CDL) and the Association of Research Libraries (ARL), and funded by a National Leadership Grant from the US Institute of Museum and Library Services, LG-254861-OLS-23.