

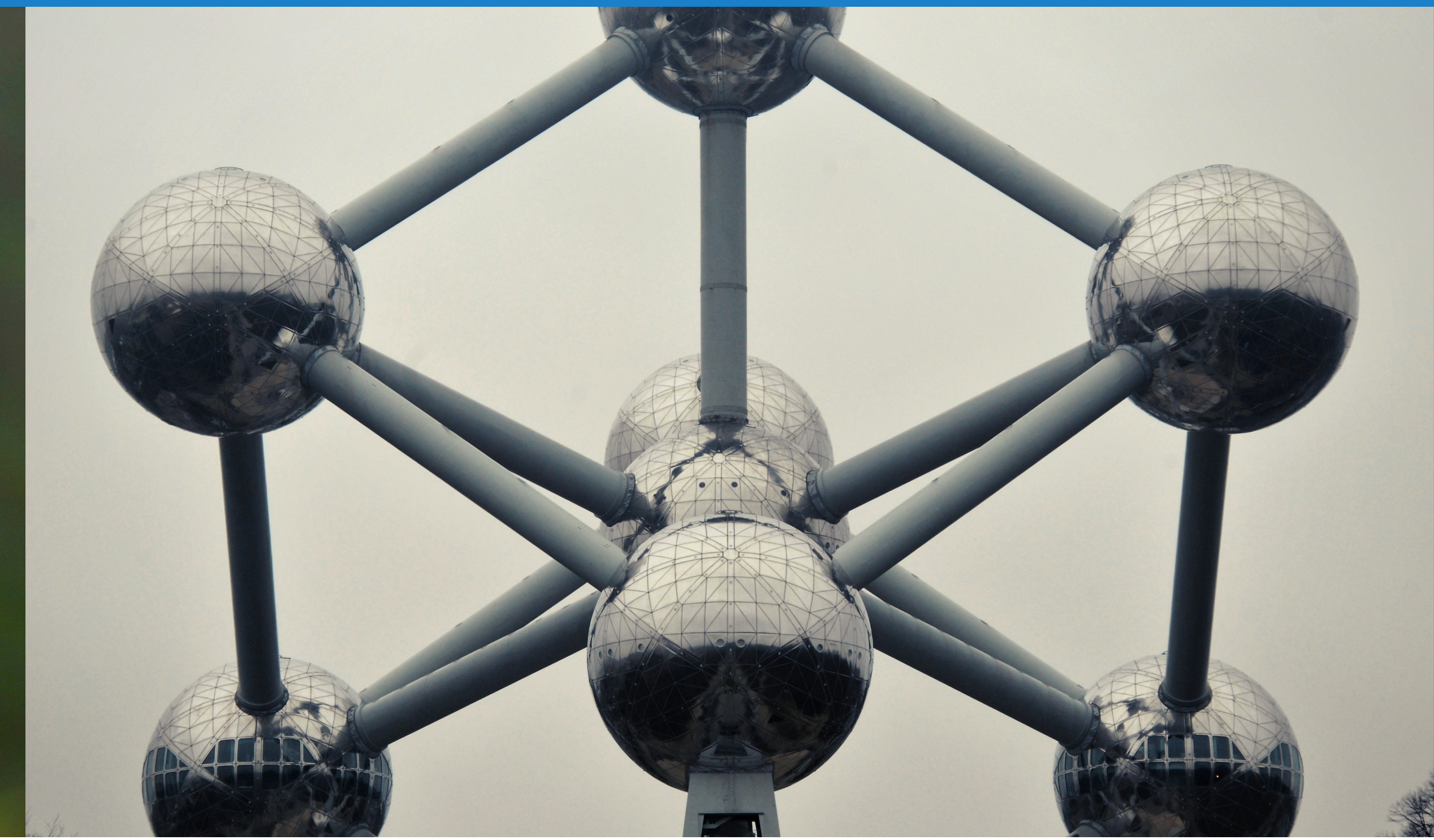
OPEN DISCOVERY FOR OPEN DATA:

An Open Source Search Tool for Institutional Research Data

Code4Lib 2019

Sara Mannheimer
Jason A. Clark
James Espeland
Kyle Hagerman

The MSU Dataset Search is an open source prototype that promotes discovery and reuse of institutional datasets through automatic metadata harvesting, unique metadata creation, and search engine optimization.



BACKGROUND

Academic libraries are increasingly participating in research data publication and preservation. However, in order to support research data repositories, we are either increasing spending by buying vendor solutions, or replicating work by building and managing individual instances of data repository software. As of 2018, there are nearly a thousand data repositories in the United States (see re3data), many of which provide services and policies that ensure their trustworthiness and suitability for institutional research data.

HOW IT WORKS

The MSU Dataset Search tool is based on an existing system at MSU Library that harvests citation information for published journal articles (Serman & Clark, 2017). MSU Dataset Search harvests metadata for MSU-affiliated research datasets using data repository feeds and APIs, which are parsed using PHP scripts. After undergoing automated deduplication and human curatorial review, metadata records are saved in a local database.

THE DATASET SEARCH

Small and mid-sized institutions can conserve their limited resources and increase discovery of local research datasets by directing researchers to a third-party repository, and then providing local access to institutional research datasets through a local search tool. MSU Dataset Search is an open source, scalable, sustainable, and standardized index that promotes discovery and reuse of institutional datasets while expending far fewer resources than those required for an institutional data repository.

IMPACTS & IMPLICATIONS

In the future, we hope to expand the project to include custom reporting, a submission interface for MSU researchers, and a JSON-LD API for developers. We also plan to share a package installer for local installations. With wider adoption, MSU Dataset Search could lead to three key impacts: (1) increased discovery and reuse of research data; (2) promotion of research data as a scholarly product; and (3) potentially leveraging economies of scale through community-wide implementation.