

Organization: Montana State University Library

Primary mentor: Kenning Arlitsch, Dean

<b>Title</b>	Analyzing the RAMP dataset to better understand content, use, and performance of institutional repositories.
<b>Description</b>	The Repository Analytics and Metrics Portal (RAMP) is a web service that has generated a massive dataset measuring use and performance across more than 50 institutional repositories (IR). This unique dataset offers the potential for a number of interesting data science projects, including analyzing the scholarly record across IR; demonstrating level of IR use; evaluating incentives to improving click-through rates; comparing RAMP download reports with vendor and IRUS download reports; automatically predicting disambiguated structured and ontological metadata; etc.
<b>Problems</b>	<ul style="list-style-type: none"> <li>• We do not yet know everything that can be done with this dataset and are open to interesting and creative ideas.</li> <li>• Dataset is large and must be parsed into useful segments.</li> </ul>
<b>Techniques</b>	Text mining; spreadsheets; machine learning; visualizations
<b>Data</b>	The RAMP dataset contains over 400 million rows in an ElasticSearch database. A production snapshot metadata record was created in Zenodo in December 2018 and the production dataset was loaded onto a Globus endpoint at Montana State for easy transferability. Monthly CSV file dumps are being loaded onto the Open Science Framework website.
<b>Outcome</b>	<ul style="list-style-type: none"> <li>• Analyses that inform papers and presentations.</li> <li>• Demonstrate the value proposition of IR and diagnose performance gaps.</li> </ul>