Organization: University of Rochester
Primary mentor: Sarah Siddiqui, STEM Librarian
Secondary mentor(s): Sarah Pugachev, Director of Science & Engineering Libraries and Research Initiatives

Project Title	Understanding Data Deposits
Description	With the NIH mandate for data sharing, University of Rochester Libraries have increased our support around data management and deposits. In addition to setting up a data repository (Figshare), we also encourage our researchers to seek out and use disciplinary repositories. We recognize that the need for data sharing will continue to grow as additional funders and publishers mandate it. The Libraries' would benefit from a stronger mechanism to understand where our researchers are depositing data. Our fellow will help us refine our data collection techniques using APIs and conduct reproducible analyses on the results.
Problems/ Research Questions	1. We need to gather data from a number of different sources. Some we can easily access (our data repository and Dimensions), for others we need to develop workflows to access (disciplinary repositories). The first task will be to discover how we can access information about UR researchers' deposits in these disciplinary repositories. We will provide data to start from Dimensions and our local Figshare repository. 2. After we have developed workflows for collecting data, we would like to understand researchers' behaviors: Who is depositing data? Where are they depositing it? How large are the datasets? What formats are

	submitted/supported?
Techniques	1. APIs scripting 2. Data analysis 3. Other relevant techniques as identified during discussions with the fellow
Tools/ Languages used	Python or R
Data	The fellow will be given two data sources to start and explore API sources for additional data points. **Description**: Dataset information from Dimensions filtered by UR researchers with information about data type, repository, date submitted, and researchers. **Data type**: CSV file **Description:** Data from our Data Repository, Figshare. Including information about deposits with dates, researchers, and file types. **Data type:** CSV file **Description:** Data from disciplinary repositories discussed during the project and harvested for co-analysis. **Data type:** Depends on repository
Outcome	We would like workflows that can be repeated by our libraries in future years. The outcome of these workflows and analyses performed by the fellow will be used to refine our data services at the University of Rochester
Milestone Timeline	**July-August**: LEADING Fellow bootcamp; Understand data provided; Explore other data sources **September-October**: Harvest data from additional sources using scripting **November**: Analyze collected data **December**: Share findings with larger team at UR
References	