

Social Networks and Archival Context OpenRefine Plugin  
(Technical Report)

Archival Collections in the Digital Age  
(STS Research Paper)

An Undergraduate Thesis Portfolio  
Presented to the Faculty of the  
School of Engineering and Applied Science  
In Partial Fulfillment of the Requirements for the Degree  
Bachelor of Science in Computer Science

by

Sandra Johanna Gould

May 5, 2020

## **List of Contents**

1. Preface
2. Technical Report: Social Networks and Archival Context OpenRefine Plugin
3. STS Research Paper: Archival Collections in the Digital Age
4. Prospectus

## Preface

How can access to collections be improved? Libraries, archives, and similar institutions hold documents and other items for public use. However, many people in the general public don't know just how much they have available to them, or how to access it.

Social Networks and Archival Context (SNAC) is a database that helps researchers and archivists access archival material about historical persons. To populate the database, archivists must use SNAC's website interface, a slow and laborious task. To speed the process, the research team sought to develop a standalone plugin in OpenRefine, an open-source codebase written in Java, JavaScript, HTML, and CSS. A successful plugin may reduce data transfer time for hundreds of entries from hours to minutes. Upon completion of the project, the research team successfully created a stable plugin that allows its users to import external data, match it to existing data in SNAC, and push to SNAC. Future work on this project includes making the extension more user-friendly for its mostly non-technical user base and optimizing the plugin's performance, such as with machine learning or using cloud technologies.

How are library systems promoting the use of their collections and archival data? Libraries and archival institutions are investing in digital resources and are encouraging the public both on and offline to visit. They have digitized collections, provided ebooks, enhanced their online presence through social media and their own websites, provided internet access in their libraries, and renovated to include more collaborative and technology spaces. Besides the libraries and archives themselves, the American Library Association (ALA), the Federal Communications Commission (FCC), and the Institute of Museum and Library Services (IMLS) have all been leading the trend.