# SPARClink

Visualizing the impact of SPARC

Sanjay Soundarajan, Sachira Kuruppu, Monalisa Achalla, Jongchan Kim, Ashutosh Singh

# Defining the Problem

- SPARC datasets and protocols are not referenced in publications similar to research articles.
- Dataset and protocol identifiers are only cited within the body or under supplementary materials of the manuscripts.
- Manual computation is needed to accumulate the citation information of datasets, protocols, and their publications.
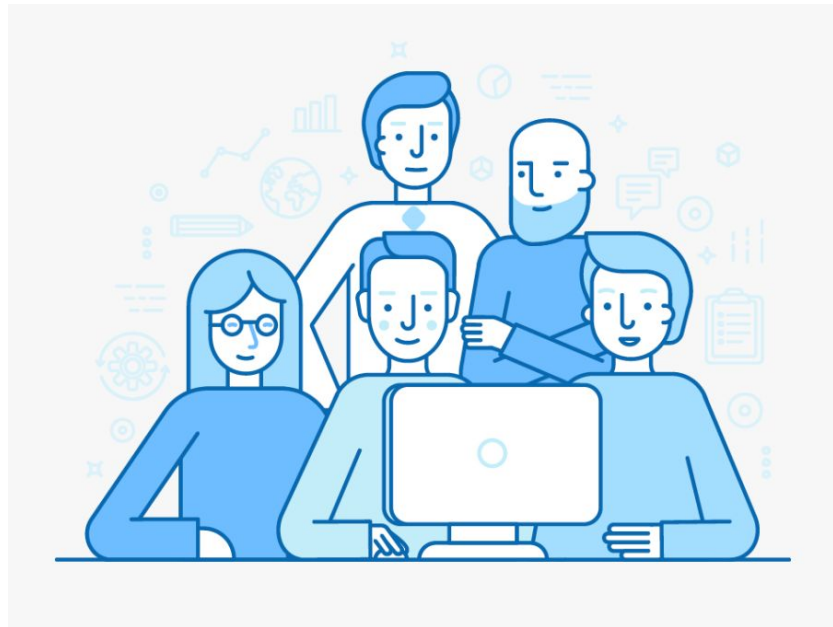- No simple visualizations for the average person
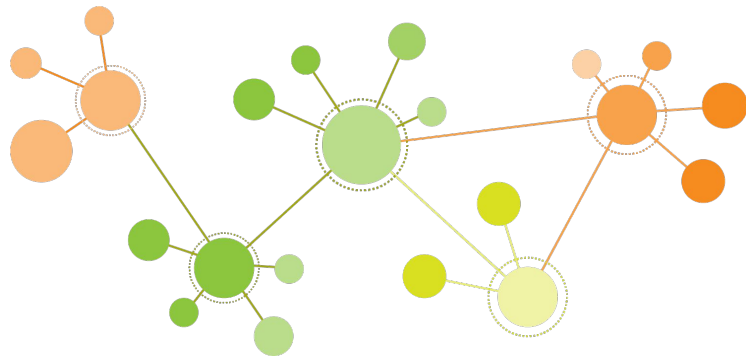
# Defining Impact of SPARC

We visualize the impact of SPARC as

- the number of datasets created
- the number of protocols created
- the number of papers published as part of NIH awards by SPARC.

How those datasets, protocols and publications are utilized by the wider scientific community.

# Our Solution

- Implement an automatic metadata extraction workflow
- Query for citation information using publicly accessible and open source tools and platforms.
- Perform machine learning based analysis on the results
- Provide a interactable visualization that any person is able to access
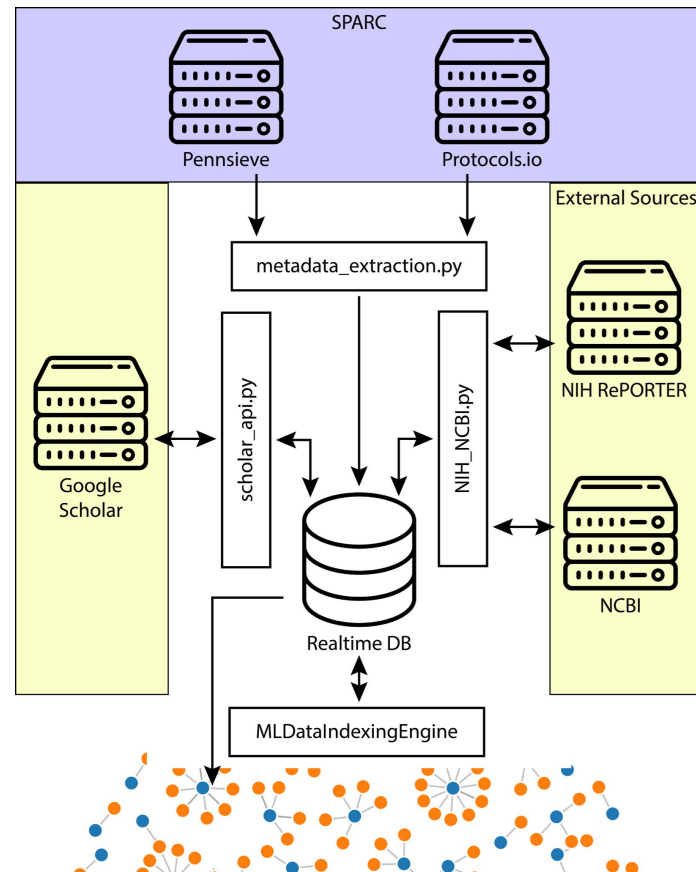
# Metadata Extraction

- Retrieve dataset metadata from Pennsieve.
- Retrieve protocols developed by SPARC from protocols.io
- Papers published as part of SPARC funding are retrieved from NIH RePORTER.
- Research articles that reference or mention these datasets, protocols, and papers are queried from:
  - NCBI (pubmed, pubmed central)
  - Google Scholar
- These items are linked to each other, and stored in a real-time database.
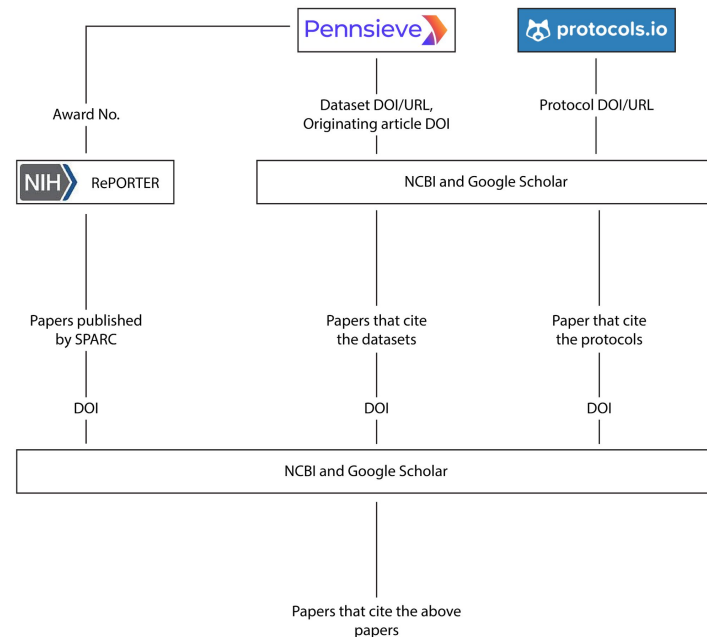
# ML Data Indexing & Visualization

Key-Features

- Spelling Recommendation
- Article recommendation based on their descriptor (e.g. Paper Title, dataset description)
- Keyword mining and visualisation through word cloud
- Click keywords for easy and fast search
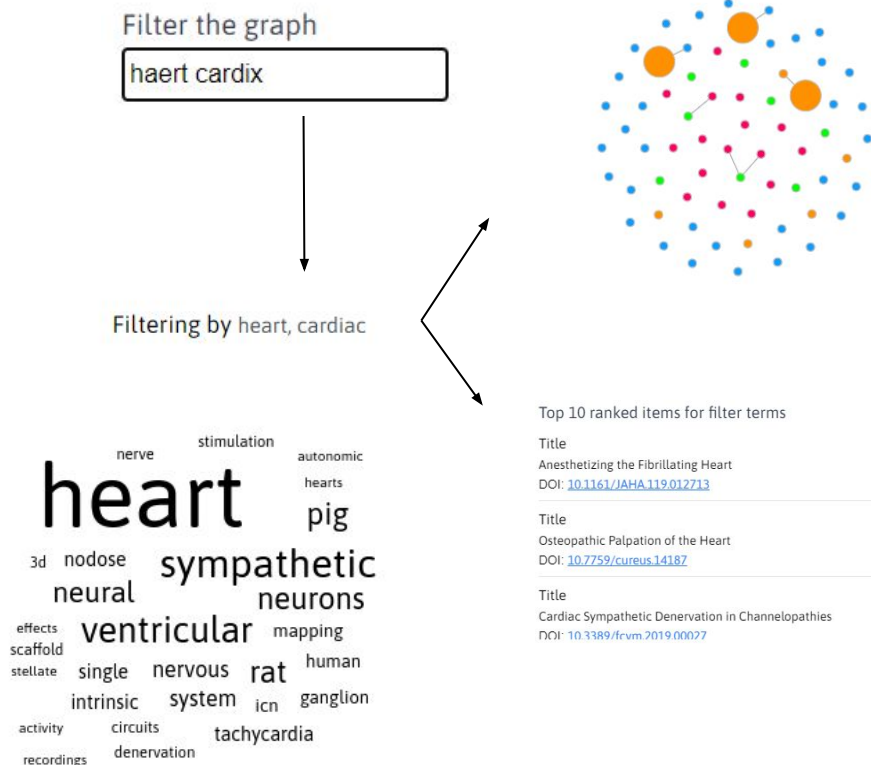
# Visualization

- Built with Vue.js, Tailwind CSS and D3.js
- AN interactive 'force based' undirected graph that showcases all citations across all SPARC material
- Pulls data from realtime database and ML Data Indexing Engine to add smart search and always accessible visualizations
- All ranked citations will be pulled from our backend API for a more accurate filter feature.
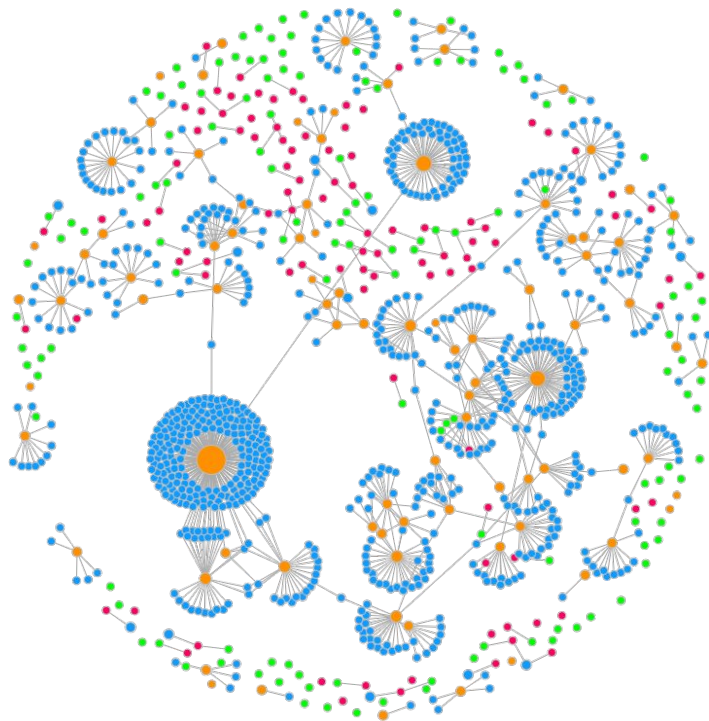
# Demo

bit.ly/sparclink

# Impact on SPARC research community

1. SPARClink can be used to visualize the impact of SPARC datasets, protocols, and publications on the wider scientific community.
2. Able to identify how datasets and protocols have been combined to answer other research problems in the biomedical domain.
3. Show key related resources based on query terms when trying to find helpful data sources
4. Fun to use visualization that makes searching for data an enjoyable experience.

# Future Work

1. Enable compatibility with SDS 2.0 for better metadata extraction
2. Redo the UI to be more accessible to assistive technologies and other accessibility improvements
3. Implement a more consistent backend for both the database and API with access to google scholar for even more suggestions.
4. Enable smarter learning algorithms to improve the returned suggestions for even more relevancy.
5. Compare the network structure of articles and datasets published by SPARC and outside SPARC to better understand the Impact of SPARC
6. Use SPARClink network for understanding the spread of articles across various subfields

A special thank you to the SPARC program and the DRC committee for giving us the opportunity to show the value and impact of SPARC data and their FAIRness as part of SPARClink

# Questions?
# Comments?
# Suggestions?