



SOFTWARE TOOL ARTICLE

SPARCLink: an interactive tool to visualize the impact of the SPARC program [version 1; peer review: 2 approved, 1 approved with reservations]

Sanjay Soundarajan ¹, Sachira Kuruppu ², Ashutosh Singh³, Jongchan Kim⁴, Monalisa Achalla⁵

¹Fair Data Innovations Hub, California Medical Innovations Institute, San Diego, California, USA

²Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

³Electrical and Computer Engineering Department, Northeastern University, Boston, Massachusetts, USA

⁴Data Science, The George Washington University, Washington, District of Columbia, USA

⁵Clarkson Center for Complex Systems Science, Clarkson University, Post, Potsdam, New York, USA

V1 First published: 31 Jan 2022, 11:124
<https://doi.org/10.12688/f1000research.75071.1>
 Latest published: 31 Jan 2022, 11:124
<https://doi.org/10.12688/f1000research.75071.1>

Abstract

The National Institutes of Health (NIH) Stimulating Peripheral Activity to Relieve Conditions (SPARC) program seeks to accelerate the development of therapeutic devices that modulate electrical activity in nerves to improve organ function. SPARC-funded researchers are generating rich datasets from neuromodulation research that are curated and shared according to FAIR (Findable, Accessible, Interoperable, and Reusable) guidelines and are accessible to the public on the SPARC data portal. Keeping track of the utilization of these datasets within the larger research community is a feature that will benefit data-generating researchers in showcasing the impact of their SPARC outcomes. This will also allow the SPARC program to display the impact of the FAIR data curation and sharing practices that have been implemented. This manuscript provides the methods and outcomes of SPARCLink, our web tool for visualizing the impact of SPARC, which won the Second prize at the 2021 SPARC FAIR Codeathon. With SPARCLink, we built a system that automatically and continuously finds new published SPARC scientific outputs (datasets, publications, protocols) and the external resources referring to them. SPARC datasets and protocols are queried using publicly accessible REST application programming interfaces (APIs, provided by Pennsieve and Protocols.io) and stored in a publicly accessible database. Citation information for these resources is retrieved using the NIH reporter API and National Center for Biotechnology Information (NCBI) Entrez system. A novel knowledge graph-based structure was created to visualize the results of these queries and showcase the impact that the FAIR data principles can have on the research landscape when they are adopted by a consortium.

Open Peer Review

Approval Status ? ✓ ✓

	1	2	3
version 1 31 Jan 2022	? view	✓ view	✓ view

1. **Tao Zeng**, University of Chinese Academy of Sciences, Shanghai, China
2. **Karim Fouad**, University of Alberta, Edmonton, Canada
Abel Torres-Espin, University of Alberta, Edmonton, Canada
 University of California San Francisco, San Francisco, USA
3. **Angela Pinot de Moira** , University of Copenhagen, Copenhagen, Denmark
 Imperial College London, London, UK

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Visualization, machine-learning, citations, FAIR, data sharing



This article is included in the **Research on Research, Policy & Culture** gateway.



This article is included in the **Hackathons** collection.

Corresponding author: Sanjay Soundarajan (ssoundarajan@calmi2.org)

Author roles: **Soundarajan S:** Conceptualization, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation; **Kuruppu S:** Investigation, Methodology, Software; **Singh A:** Methodology, Software; **Kim J:** Software; **Achalla M:** Methodology, Writing – Original Draft Preparation

Competing interests: NIH SPARC is the primary funder for the publication of this research article.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2022 Soundarajan S *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Soundarajan S, Kuruppu S, Singh A *et al.* **SPARCLink: an interactive tool to visualize the impact of the SPARC program [version 1; peer review: 2 approved, 1 approved with reservations]** F1000Research 2022, **11**:124 <https://doi.org/10.12688/f1000research.75071.1>

First published: 31 Jan 2022, **11**:124 <https://doi.org/10.12688/f1000research.75071.1>

Introduction

The National Institutes of Health (NIH) Common Fund's Stimulating Peripheral Activity to Relieve Conditions (SPARC) program aims to transform our understanding of nerve-organ interactions with the intent of advancing bioelectronic medicine towards treatments that change lives.¹ The SPARC program employs a Findable, Accessible, Interoperable, and Reusable (FAIR) first approach for its datasets, protocols, and publications, hence enabling the data to be easily reused by research communities globally. The SPARC [data portal](#) can be used as the gateway to access fully curated datasets at any time.² Using the portal, researchers can search for data used in real-world experiments to verify or corroborate studies in device development. There is also potential for the data generated by the SPARC program to be useful outside the current field of study showcasing the benefits of multi-discipline data generation and sharing.³

All SPARC datasets are curated by the researchers according to the SPARC Data Standards (SDS), a data and metadata structure derived from the Brain Imaging Data Structure (BIDS).⁴ Several resources are made available to SPARC researchers for making their data FAIR, such as the cloud data platform [Pennsieve](#), the curated vocabulary selector and annotation platform [SciCrunch](#), the open-source computational modeling platform [o²S²PARC](#), the online microscopy image viewer [Biolumica](#), and the data curation software [SODA](#).⁴⁻⁶ The datasets submitted by researchers also follow an extensive curation process where teams from the SPARC Data Resource Center (DRC) examine the submitted data and work with the researchers to ensure all aspects of the FAIR data principles are being followed.^{4,6,7} Once these datasets are made public, access to them is provided through the Pennsieve Discover service and [sparc.science](#), the official access point of the SPARC Portal.⁸

While the submission and curation of data are simplified by such tools, one of the greater benefits of the FAIR guidelines is the ability to reuse data in other studies by other researchers around the world. However, a researcher who has submitted a dataset might not always be aware of the reuse of their original submitted data since current citation indexing tools, like Google Scholar, do not account for datasets. To address this shortcoming, we developed SPARCLink during the 2021 SPARC FAIR Codeathon (July 12th, 2021 – July 26th, 2021),⁹ a system that queries all external publications using open source tools and platforms and creates a database and visualizations of citations that are helpful to showcase the impact of the SPARC consortium. In this instance, we define impact as the frequency of citations of SPARC-funded resources. By using citations as the key measure in SPARCLink, we have created a method for showcasing the reuse of generated data and the benefits that FAIR data generation practices have on the overall scientific community. A visual representation of the reuse of data will allow both researchers and the general public to see the benefits of the concept of FAIR data and the immediate utilization of publicly funded datasets in advancing the field of bioelectronic medicine.

Methods

Our solution can broadly be categorized into four steps. The first step involves the backend extraction of data using various application programming interfaces (APIs). The second step is setting up and storing the extracted data on a real-time database. The third step involves using machine learning to improve user experience by developing context-sensitive word clouds and smart keyword searches in the portal. The final step is used to create an engaging visualization that users of the SPARCLink system will be able to interact with to view the extracted data. A visual representation of this workflow is shown in [Figure 1](#).

Extraction of data using APIs

We used the dataset information retrieved directly from the Pennsieve data storage platform by running the Pennsieve API to gather all publicly available SPARC datasets.¹⁰ The protocols stored on [Protocols.io](#) under the SPARC group were also queried via this method.¹¹ A list of public and published DOIs was created in our database with additional information regarding the study authors and descriptions.

We used NIH RePORTER to retrieve data about the papers published as part of SPARC funding. Research articles that reference or mention these datasets, protocols, and publications were queried from NCBI (PubMed, PubMed central) repositories using the search endpoint of their Python API.¹² [Figure 2](#) shows the overall flow of data between the APIs and resources queried to get the data. The NIH RePORTER API uses project number (also known as the award number) of NIH funding associated with SPARC datasets (this is provided by the author as additional metadata required when publishing a dataset) as an input to get details including a study identifier, name of the organization that received funding, country of the organization, amount of funding received and keywords of the project topic. The NCBI API uses an identifier for PubMed Central articles to retrieve information such as article name, journal name, year of publications, and authors.

Storing extracted data in a database

We used Google's Firebase real-time database to store all the information retrieved via the NIH RePORTER system. The data was stored in a JSON format with read access available to anyone via a dedicated URL. The data in this database was

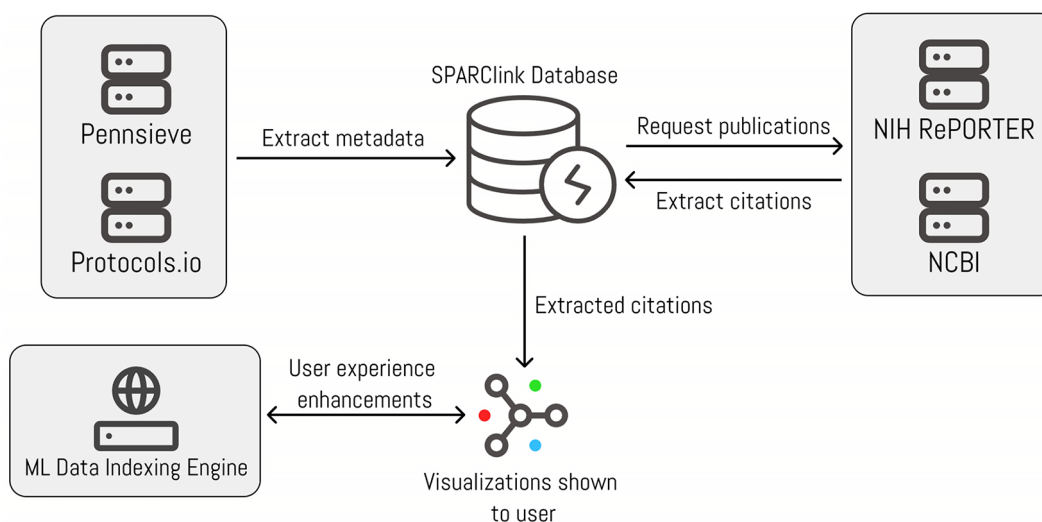


Figure 1. The flow of data between the submodules of SPARClick.

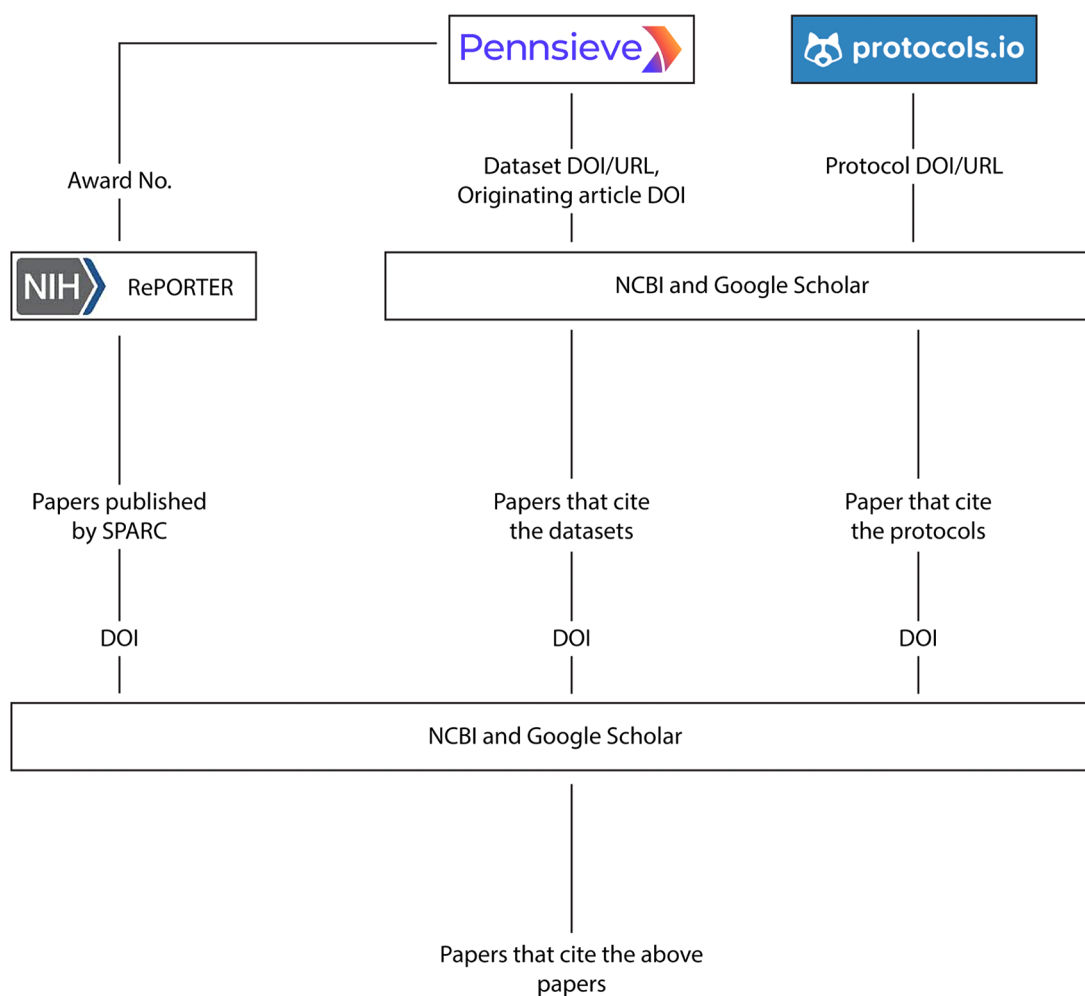


Figure 2. Methods implemented to gather citations of datasets, protocols, SPARC, and external publications.

split up into four separate sections labeled Awards, Datasets, Publications, and Protocols. All the entries within this database were given a unique identifier. These identifiers were used to link the data within the database to form a relational database. The links within the data represent the citations or use of resources within other publications. All publications within the database were uniquely identified as either SPARC-funded publications or non-SPARC publications (external publications that cite SPARC datasets and publications.)

Displaying the extracted data to the user

The front-end demo of the SPARCLink web page uses [Vue.js](#) to create a functional prototype of the SPARCLink system. An interactive force-based undirected graph visualization was created using the D3.js JavaScript library. The choice to represent the results through such a graph was motivated by the desire to show an intuitively understandable way of showing the connected nature of citations and data reuse. The website itself is hosted on Vercel as a static front end.¹³ On the webpage, the visualizations can be filtered by key terms or resource type to get a better understanding of the resources created using the SPARC program. A screenshot of the webpage is shown in [Figure 3](#).

Machine Learning Data Indexing Engine

To provide some additional functionality on the front-end demo of SPARCLink, we used machine learning algorithms to enhance the user experience. We called this function of the SPARCLink project the Machine Learning Data Indexing Engine.

We used the SymSpell algorithm present in the scikit learn package and trained it on the vocabulary built using the SPARCLink database.¹⁴ We used delete-only edit candidate generation for generating different combinations of spelling errors, and used both character-level embedding and word embedding for recommending the most probable correct spelling. The output of the spell correction algorithm was used to generate sentence-level embedding and was then compared with the embeddings of different descriptions of the items in the dataset. We obtained a ranking of all the items in the dataset based on their similarity with the searched string. The top 10 were chosen to be shown on the front end.

This module was also used to generate keywords using the [keyBERT](#) pretrained model.¹⁵ It generated the top 50 keywords associated with the whole document. It also made use of the Maximal marginal relevance algorithm to pick keywords that have a higher distance among them.¹⁶ This ensures diversity among the chosen keywords.

The engine also contains algorithms that learn vector embeddings of the descriptors of the elements present in the SPARCLink database. Based on these vector embeddings, the algorithms compute the similarity between the vector representation of each word in the vocabulary with the vector representing the whole dataset and find keywords that would describe the resource. A word cloud is generated based on the relevance of these results to further enhance the user experience.

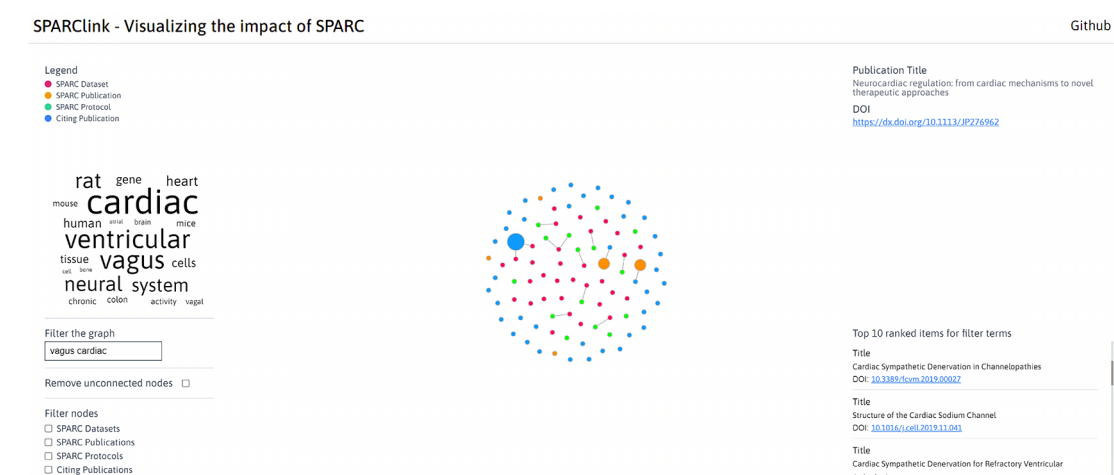


Figure 3. The design of the SPARCLink webpage where results from the machine learning module are shown alongside the visualizations of SPARCLink. The visualizations and the results in this figure have been filtered with the vagus and cardiac keywords.

Results

Using SPARCLink, researchers can aggregate all the resources created through the SPARC program and quantify their impact. The visualization created by the SPARCLink system is shown in [Figure 4](#). The nodes in the undirected graph signify a unique SPARC resource (publication, protocol, and dataset) and the edges in the graph signify the citations or references as found by SPARCLink. A well-connected graph of datasets and publications were observed but a significant number of protocols were seemingly distinct from the rest of the resources despite being pulled from the SPARC protocols.io group. This could be associated with protocols that are published on protocols.io but for which the associated datasets have not been made public yet.

The word map generated from the main dataset visualizations is shown in [Figure 5](#). The size of the word with respect to its neighbors corresponds to the frequency and significance of the word within all the searchable metadata that we have indexed. Selecting any of the words in this map will automatically filter the SPARCLink visualizations. Using a keyword filter on the graph will also prompt the top-ranking items for the keyword to be displayed on the side of the page. This ranking is shown as a scrollable list, as seen in [Figure 6](#). Both the word map and top-ranked recommendations are continuously updating themselves when new input terms are entered via the SPARCLink webpage.

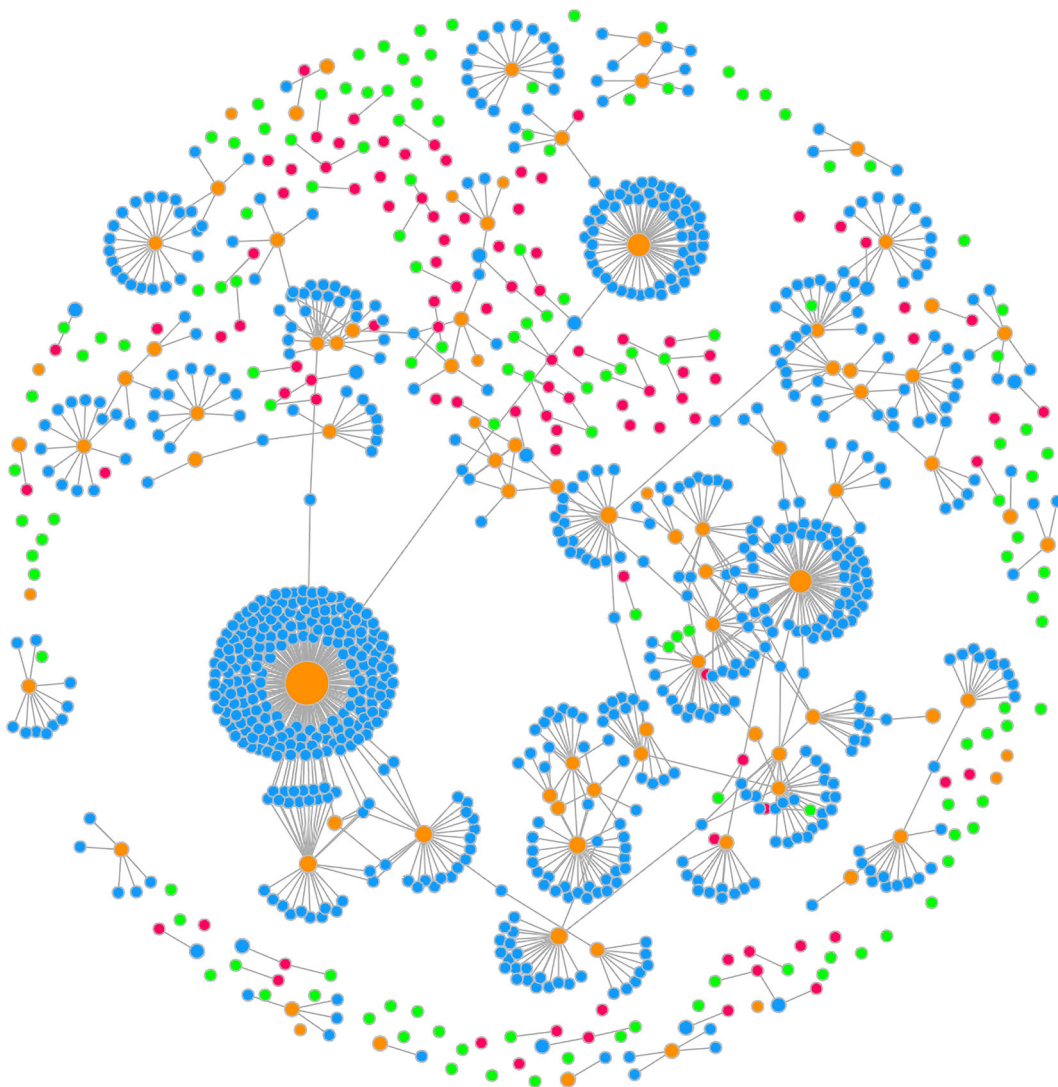


Figure 4. An interactive visualization created by SPARCLink showing the connected nature of all SPARC resources.

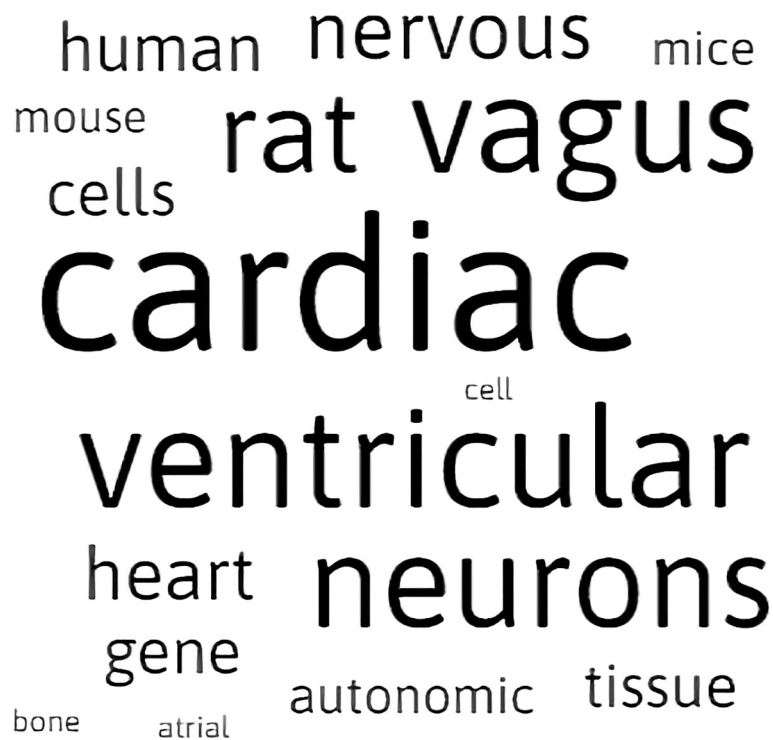


Figure 5. The word maps created by SPARCLink are a visual representation of the most significant words shown in the graph-based visualization.

Top 10 ranked items for filter terms

Title

Cardiac Sympathetic Denervation in Channelopathies

DOI: [10.3389/fcvm.2019.00027](https://doi.org/10.3389/fcvm.2019.00027)

Title

Structure of the Cardiac Sodium Channel

DOI: [10.1016/j.j.cell.2019.11.041](https://doi.org/10.1016/j.j.cell.2019.11.041)

Title

Acute hemodynamics of cardiac sympathetic denervation

DOI: [10.1016/i.ine.2020.06.006](https://doi.org/10.1016/i.ine.2020.06.006)

Figure 6. A list of resources that are recommended by SPARCLink when a search term filter is provided by the user.

Discussion and conclusions

Using FAIR standards can greatly improve the use of data across multiple disciplines and potentially lead to new and exciting discoveries in the field of biomedical science. The benefits of employing the FAIR data principles for data generation, curation, and sharing can, however, be hard to quantify for researchers or members of the general public. Using a system like SPARCLink, researchers at all levels can get up-to-date feedback on the use of their data and all the advantages that the FAIR standards provide to efforts in advancing biomedical science. In this work, we developed such a tool for the SPARC program to enable quantification of the reuse of the FAIR SPARC resources (datasets, manuscripts, protocols).

The primary challenge in accomplishing this task lies in the fact that the SPARC datasets and protocols are not referenced in the bibliography of research manuscripts, which is the common practice. Instead, the SPARC dataset and protocol identifiers or URLs are only mentioned in the text or under supplementary materials, which makes querying this information a challenging task. Furthermore, datasets created in the SPARC program can be embargoed for up to 12 months to allow researchers enough time to document and publish their findings. However, protocols are made public immediately since protocols.io does not have an option to embargo the open publishing of these protocols. This could also add to the sparse graphs and we can expect the connectedness of this graph to improve as time goes on.

In the future, we plan on adding the Google Scholar system as an additional resource for data extraction. This should improve the connectedness of our extracted data network as well. Additional filtering functions and performance improvements for very large numbers of nodes are also planned. Currently, the tool is hosted on an independent webpage, but we also aim to integrate it directly within the SPARC portal so that visitors can conveniently visualize the reuse and impact of the different SPARC-generated resources.

Data availability

At the time of publication, the SPARCLink system visualizations can be found at <https://sparclink.vercel.app> and are expected to be always online going forward. The backend system that queries all the publications is currently paused due to a lack of system resources. The code for SPARCLink has been developed to be accessible to anyone who wants to fork the repository from GitHub and run a local version of this project. Instructions on how to run the modules locally are also available in the GitHub repository. The database of currently extracted citation data can be queried via REST protocols using the links provided below. The machine learning data indexing engine is hosted on a web server provided by pythonanywhere.com and is publicly accessible via its API endpoints. This module is also available to be run in local configuration seamlessly.

Software availability

Source code available from: <https://github.com/fairdatahub/SPARCLink>

Archived source code as at time of publication: <https://doi.org/10.5281/zenodo.5550844>

License: MIT

Author endorsement

David Nickerson confirms that the author has an appropriate level of expertise to conduct this research and confirms that the submission is of an acceptable scientific standard. David Nickerson declares they were an organizer of the Hackathon in which the work described in this paper was performed. Affiliation: Auckland Bioengineering Institute, University of Auckland, New Zealand.

Acknowledgments

We would like to thank the NIH Common Fund's SPARC Program and the organizers of the 2021 SPARC FAIR Codeathon for their support during the development of this project.

References

1. National Institutes of Health: **Stimulating Peripheral Activity to Relieve Conditions (SPARC)**. 2014 [cited 2021 Oct 22]. [Reference Source](#)
2. National Institutes of Health: **SPARC Portal**. [cited 2021 Oct 22]. [Reference Source](#)
3. Quey R, Schiefer MA, Kiran A, et al.: **KnowMore: An Automated Knowledge Discovery Tool for the FAIR SPARC Datasets**. *bioRxiv*. 2021 [cited 2021 Oct 22]; p. 2021.08.08.455581. [Publisher Full Text](#)
4. Bandrowski A, Grethe JS, Pilko A, et al.: **SPARC Data Structure: Rationale and Design of a FAIR Standard for Biomedical Research Data**. *bioRxiv*. 2021 [cited 2021 Oct 22]; p. 2021.02.10.430563. [Publisher Full Text](#)
5. Patel B, Srivastava H, Aghasafari P, et al.: **SPARC: SODA, an interactive software for curating SPARC datasets**. *FASEB J*. 2020 Apr; **34**(S1): 1–1. [Publisher Full Text](#)
6. Osanlouy M, Bandrowski A, de Bono B, et al.: **The SPARC DRC: Building a Resource for the Autonomic Nervous System Community**. *Front Physiol*. 2021 Jun 24; **12**: 693735. [PubMed Abstract](#) | [Publisher Full Text](#)
7. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al.: **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci Data*. 2016 Mar 15; **3**: 160018. [PubMed Abstract](#) | [Publisher Full Text](#)
8. The University of Pennsylvania: **Pennsieve Discover**. [cited 2021 Oct 22]. [Reference Source](#)

9. SPARC: **2021 SPARC FAIR Codeathon**. **SPARC Portal**. [cited 2021 Oct 22].
[Reference Source](#)
10. The University of Pennsylvania: **Pennsieve API**. [cited 2021 Oct 22].
[Reference Source](#)
11. Protocols I: **Protocols.io for developers**. [cited 2021 Oct 22].
[Reference Source](#)
12. National Institutes of Health: **NIH RePORTER API**. [cited 2021 Oct 22].
[Reference Source](#)
13. Soundarajan S: **SPARClink Portal**. 2021 [cited 2021 Oct 22].
[Reference Source](#)
14. Garbe W: *SymSpell: SymSpell: 1 million times faster spelling correction & fuzzy search through Symmetric Delete spelling correction algorithm*. Github; [cited 2021 Oct 22].
[Reference Source](#)
15. Grootendorst M: *KeyBERT: Minimal keyword extraction with BERT*. Github; [cited 2021 Oct 22].
[Reference Source](#)
16. Carbinell J, Goldstein J: *The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries*. ACM SIGIR Forum; 2017; **51**: 209–210.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 1

Reviewer Report 16 February 2023

<https://doi.org/10.5256/f1000research.78888.r160993>

© 2023 Pinot de Moira A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Angela Pinot de Moira 

¹ Section of Epidemiology, University of Copenhagen, Copenhagen, Denmark

² National Heart and Lung Institute, Imperial College London, London, UK

This is a clearly written paper describing SPARCLink, a web tool that queries all external publications and creates a database and visualisations of publications that have utilised SPARC-funded resources, including datasets, publications and protocols. As well as documenting the impact of the SPARC consortium, the tool will be an invaluable resource for researchers utilising SPARC outputs for their research.

My main comment to the paper is regarding the aim to demonstrate the benefits of the FAIR principles. Currently, the paper mainly focuses on the retrieval of any publication utilising SPARC resources, i.e. including publications directly funded by SPARC. To demonstrate how the tool can be used to highlight/quantify the benefits of the FAIR principles, it would be useful to provide examples of this in the paper, i.e. demonstrate the reuse of SPARC data. If I have understood correctly, this is possible by filtering extracted publications on whether they are SPARC-funded or non-SPARC funded. It would be useful to provide a visualization with the non-SPARC funded filter applied, so that the extent of data reuse can be seen.

My second suggestion is to include a box of key terminology in the paper. There are a lot of terms that may not be understood by the reader and although links to definitions are provided, a glossary box may aid reading.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology, FAIR principles

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 26 January 2023

<https://doi.org/10.5256/f1000research.78888.r160964>

© 2023 Fouad K et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Karim Fouad

University of Alberta, Edmonton, Canada

Abel Torres-Espin

¹ University of Alberta, Edmonton, Canada

² University of California San Francisco, San Francisco, CA, USA

This is a very succinct article introducing a web tool for querying and visualizing the scientific output of the SPARC program and how it has been utilized by the research community. The process begins with extracting and storing related data via APIs. Then word clouds and key word searches allow users to refine their search to then visualize the relation between the SPARC output products. The product is a very useful tool to explore the ongoing impact and use of SPARC related research and products. Potentially the impact of the tool is not really on what the authors mention (i.e., show the impact of FAIR data sharing on the overall scientific community) and thus somewhat overstated. The visualization tool simply provides a summary of the reuse of SPARC related work products, but not a comparison to other approaches for data sharing that would allow for assessing the impact of FAIR. The strength lies in the demonstration of data reuse, specific impact of their data sets, etc. This is a highly valuable tool on so many levels including strategies for future research and general education of the value of data sharing.

The manuscript would benefit from a few clarifications and details especially in the figure legends. For example, the difference between a SPARC publication and a data set should be explained. Are

SPARC data sets not published?

Figure 3 show a lot of white space, and the legend would benefit from more detail. What does define the size of the nodes in the visualization? Did the authors consider directional links between the items in the form of a directed graph, which would simplify comprehension of what is shown? For example, the edge of the graph could show the direction from the node citing to the one receptor of a citation.

Figure 4 requires a legend for the different colors, an explanation for the different size of the nodes, and once again would benefit from directional edges. Lastly, not surprisingly the overwhelming links to publications mask the reuse of data sets. Maybe a different filter could be applied to show that important relation.

Minor suggestions:

- In the Abstract, consider adding 'electrical' in front of neuromodulation.
- 3d paragraph in introduction, consider not to refer (i.e., these tools) when starting a new paragraph.
- Please define "citations of SPARC-funded resources".

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroplasticity, data sharing

We confirm that we have read this submission and believe that we have an appropriate level

of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 27 June 2022

<https://doi.org/10.5256/f1000research.78888.r140292>

© 2022 Zeng T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Tao Zeng

AS Key Laboratory of Computational Biology, Bio-Med Big Data Center, University of Chinese Academy of Sciences, Shanghai, China

In the paper “SPARClink: an interactive tool to visualize the impact of the SPARC program”, authors aim to introduce a web tool SPARClink for visualizing the impact of SPARC, whose methods and outcomes support FAIR guidelines.

SPARClink should be a useful tool/software for supporting SPARC program and corresponding consortium. For the work and introduction in this paper, I have several suggestions:

1. SPARC program employs FAIR, so, it is necessary to introduce more about corresponding function of SPARClink serving each FAIR guideline, e.g. for “Reusable”, what data or protocol can be reused and how other researchers can obtain and reuse them.
2. Similar to other programs, the data produced in SPARC would have raw data, pre-processed data, or analyzed data, or summary data, etc. The organization or level of SPARC data is better to be clearly introduced, and supply detailed cases of how SPARClink can manage and share these data.
3. In current implementation, “interactive force-based undirected graph visualization” is simple, and the directed relation is better to consider, e.g. the paper and its public data, and the paper and its reused data, would have different relation directions. Also, in the abstract, the authors stated “a novel knowledge graph-based structure was created to visualize ...” Thus, the novelty of the network structure and representation should have highlights in revision.
4. For interactive visualization shown in Figure 4, or used in current web tool, the network has less information - it is better to directly show the information about each node or edge on network in web page in a scalable manner, especially for large knowledge network.
5. The word maps shown in Figure 5 would be a general function in many other tools, it is necessary to offer the interesting word or sentence association between key words and SPARC outcomes.
6. Indeed, for FAIR, the software SPARClink itself is suggested to supply Docker version, so other users can easily reuse such useful framework into their applications.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Machine learning and bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research