

University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England
Area Librarian e-Science Symposium

2018 e-Science Symposium

Apr 5th, 12:00 AM

Following the Trends in eScience: Updating the eScience Thesaurus

Tess Grynoch

University of Massachusetts Medical School

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium



Part of the [Scholarly Communication Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

Repository Citation

Grynoch, T. (2018). Following the Trends in eScience: Updating the eScience Thesaurus. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/cgtc-mc05>. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2018/posters/1

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.



Following the Trends in eScience: Updating the eScience Thesaurus 2.0



Tess Grynoch

Lamar Soutter Library, University of Massachusetts Medical School

Introduction

- In 2017, the New England eScience Portal for Librarians transitioned into the new, nation-wide NNLM RD3: Resources for Data-Driven Discovery site (nnlm.gov/data)
- A literature review of library-focused eScience articles was performed to update and transition the term pages of the eScience Thesaurus to the Data Thesaurus
- Using term mapping, what can these articles tell us about the trends in eScience research?

Methods

- In October 2016, replicated the search strategy used by Read, Creamer, Kafel, Vander Hart, & Martin (2013) to create the Thesaurus and limited the search to articles since 2013
 - 714 articles were found and citation information was imported into Mendeley
- Citation information was imported into VOSviewer for visualization and analysis. Software downloaded from: <http://www.vosviewer.com/>
- Term map was built from the accumulated titles and abstracts
- Term map parameters:
 - Binary counting (only counting if term is present in document, not adding extra counts for a term occurring multiple times per paper), Minimum number of occurrences of a term = 10
 - Applied term thesaurus to remove irrelevant terms and merge synonyms
 - Layout: Attraction = 2, Repulsion = 1 (default values)
 - Clustering: Resolution (determines the level of detail for the clusters, with a higher resolution generally resulting in more clusters) = 1, Minimum cluster size = 1, Merged small clusters
 - Weight of term circles by number of occurrences
 - Score for overlay visualization: Average publication year
 - Minimum line strength: 1, Maximum lines: 500

Results

- Of 9500 terms found by VOSviewer, 113 terms met the threshold of a minimum 10 occurrences and were not one of the irrelevant terms, and 4 clusters resulted
- Data, Science, and Research are central terms with Data being the most prevalent in the literature found, and spread across all years
- Terms that appear more often in the later papers include: Big data, Data quality, Open data, Relationship, Partnership
- Terms that appear more often in the earlier papers include: Information, Data curation, Trust, Institutional repository, Link, Privacy

Electronic Supplemental Files

<http://bit.ly/TrendseScience2018>

An electronic version of the poster and accompanying electronic files including a full list of terms used in the visualization, citation information used, and term thesaurus are available on eScholarship@UMMS



Visualizations

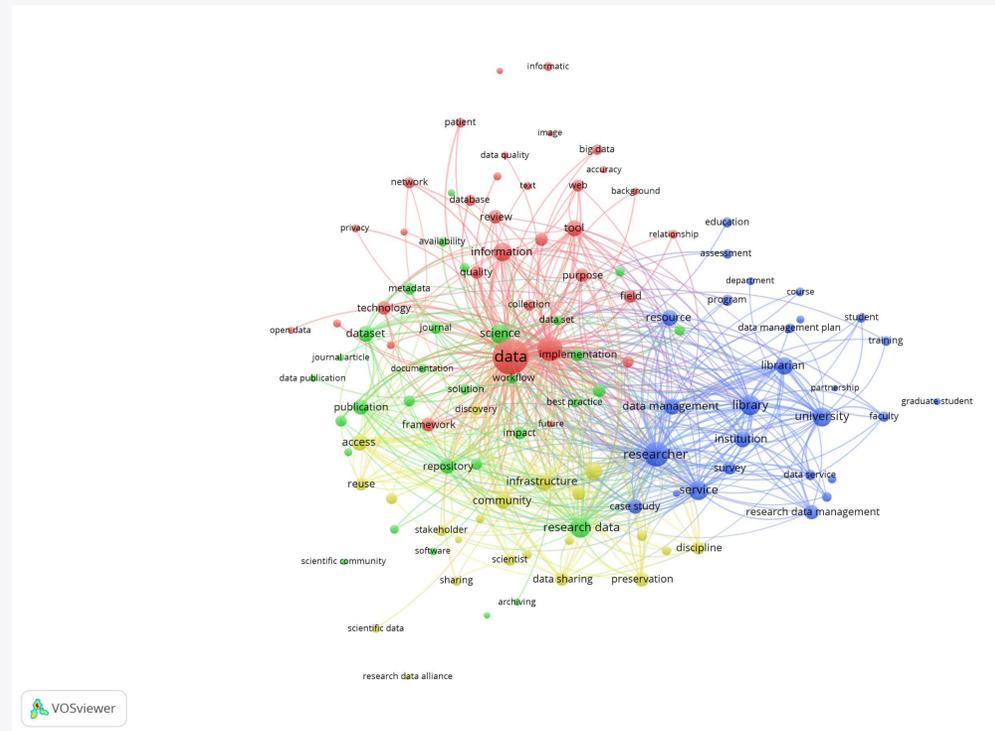


Figure 1. Term map of terms found in the 714 eScience articles from 2013-2016 where the colors depict the different clusters.

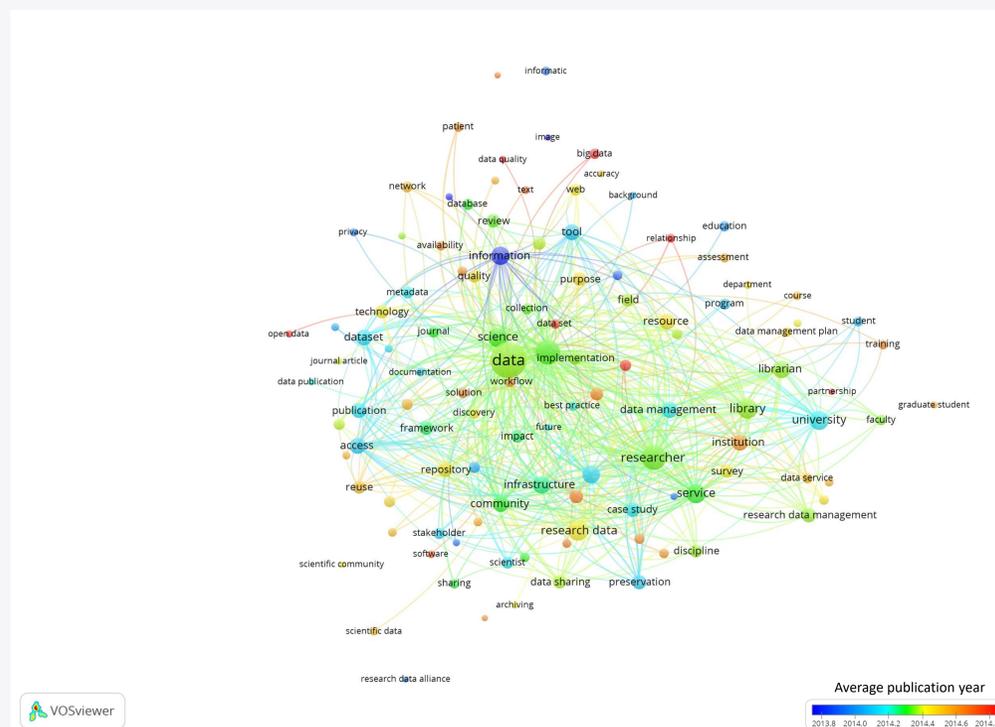
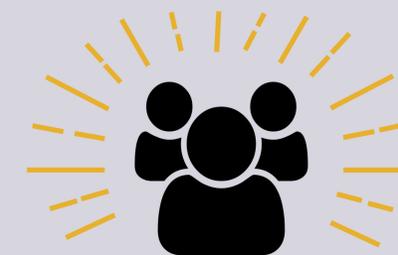


Figure 2. Term map of terms found in the 714 eScience articles from 2013-2016 with the average publication year overlaid as a color gradient.

Conclusions

- One term suspiciously missing: eScience or e-science
 - In 4 of the article titles and 7 abstracts (meets 10 occurrence minimum to be added to map)
 - Could have been merged into "Science", system might have through it was a typo (tried to prevent this through inclusion in term thesaurus)
- There are no major topic or cluster shifts within the 2013-2016 eScience articles
- Term mapping is a useful tool to see the shift in the publishing trends of a field, even in a small field such as library-focused eScience research, but would recommend a larger date range to gain a better understanding of the trends
- Due to the small dataset, the visualizations are a better snap shot of articles found than a trend prediction
- These data visualizations give a new take on existing bibliographic data
- The updated Thesaurus will be available at <https://nnlm.gov/data/data-thesaurus> coming April 2018

The Data Thesaurus is Currently Soliciting Feedback!



Is there a common data term that is often misused?

Want to share a critical resource with colleagues who are exploring the data field?

Now's your chance!

Check out the new Data Thesaurus at nnlm.gov/data/data-thesaurus and let us know what you think!

References

- Read, K., Creamer, A., Kafel, D., Vander Hart, R.J., & Martin, E.R. (2013). Building an eScience thesaurus for librarians: A collaboration between the National Network of Libraries of Medicine, New England Region and an Associate Fellow at the National Library of Medicine. *Journal of eScience Librarianship*, 2(2), 53-67. <http://dx.doi.org/10.7191/jeslib.2013.1049>
- Van Eck, N.J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ISSI Newsletter*, 7(3), 50-54. Available from: <http://www.issi-society.org/media/1131/newsletter27.pdf>

Contact

If you have questions and I am not standing in front of my poster, feel free to contact me:

- tess.grynoch@umassmed.edu
- [@Te_Gryn](https://twitter.com/Te_Gryn)

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>, or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

