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
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 NLMD 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 950 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


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@UMass

References


Contact

If you have questions and I am not standing in front of my poster, feel free to contact me: tess.grynoch@umassmed.edu

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!

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.