

Controlled Vocabularies for Scientific Data: Users and Desired Functionalities

Yue Zhang, Adrian Ogletree, Jane Greenberg
Drexel University
Metadata Research Center
30 N. 33rd St.
Philadelphia, PA 19104
{yue.zhang, aogletree, janeg}@drexel.edu

Chelcie Rowell
Wake Forest University
Z. Smith Reynolds Library
PO Box 7777
Winston-Salem, NC 27109
rowellcj@wfu.edu

ABSTRACT

Controlled vocabularies have great applicability for organizing and providing access to scientific data. This paper presents research examining the controlled vocabulary use and desired application features specific to scientific data. A survey was conducted, gathering data from U.S. DataNet participants and other data stakeholder communities. Results indicate that participants see controlled vocabularies as valued tools. Participants also reported an interest in repository technology that provides access to multiple vocabularies for scientific data.

Keywords

Controlled vocabularies, ontology, scientific data, data repository.

INTRODUCTION

Controlled vocabularies are semantic systems that are useful for organizing and accessing resources—and supporting semantic interoperability among object descriptions and repositories. Cost, usability, and other challenges surface when trying to work with more than one vocabulary in a single system. Semantic web technologies, such as linked data and the Simple Knowledge Organization System (SKOS), provide mechanisms for addressing these challenges (Greenberg et al., 2011). In fact, these developments provide a platform for enabling access to multiple controlled vocabularies in a dynamic way, as demonstrated by the HIVE (Helping Interdisciplinary Vocabulary Engineering) technology, which has been implemented in the Long Term Ecological Research (LTER) Network (LTER, 2012) and the DataNet Federation Consortium (DFC) (Conway et al., 2013).

{This is the space reserved for copyright notices.}

ASIST 2015, November 6-10, 2015, St. Louis, MO, USA.

[Author Retains Copyright. Insert personal or institutional copyright notice here.]

The HIVE implementation in these two settings (LTER and DFC) prompts us to raise important questions about the state of controlled vocabularies and their application for scientific data. We have found that the HIVE approach has been successful in leveraging SKOS vocabularies and machine learning algorithms for indexing scientific research data. Despite this progress, knowledge about the current state of controlled vocabularies across scientific repositories is limited. Moreover, there is an absence of research addressing desired system functionalities supporting controlled vocabulary applications for scientific data. Research addressing these topics can improve HIVE technology—and more broadly, controlled vocabulary applications for scientific data among various stakeholders (data contributors, curators, administrators, developers, and system users).

The research reported on in this paper considers these challenges, and examines controlled vocabulary needs across a broad swath of data repository stakeholders. Specific goals were to:

- Discover who is using controlled vocabularies and group users based on their roles.
- Compare and contrast how different types of users are using controlled vocabularies.
- Evaluate the desired features and functionalities requested by users.

The following literature review covers relevant research on controlled vocabularies, including research addressing the benefits of using controlled vocabularies as low-level semantic ontologies. Then, we discuss the research objectives, methods, and results. Lastly, we present a discussion of the results in the context of other relevant work and conclusions, highlighting key findings and noting next steps.

LITERATURE REVIEW

Controlled vocabularies are tree-structured knowledge organization systems (KOS) that provide conceptual relationships among entities and properties of concepts (Baker et al., 2013). They are knowledge structures that

allow users to organize and find information, and understand the scope of a discipline or domain being represented. Controlled vocabularies generally encode hierarchical, associative and equivalent relationships, and can be viewed as low-level semantic ontologies. As semantic ontologies, they can aid in information retrieval of documents—including research data.

Controlled vocabularies for information retrieval and interoperability

Many search and retrieval tools use ontology-supported semantics (controlled vocabularies), including query expansion with semantic relationships to extract relevant information from large corpora of documents. Supported by domain-based ontologies, information extraction tools can provide solutions to the vagueness of natural language and abstract concepts by incorporating background knowledge and semantic relations into the information retrieval process (Nagypál, 2005). A conventional search engine cannot analyze users' search queries to their full extent; however, with the support of an ontology, the query can be analyzed syntactically and semantically (Wache, 2001). Meanwhile, ontology-based retrieval methods provide shared knowledge for expressing information about the content of documents and implicit relations among concepts (Decker, 1999).

Ontologies can also be used to formulate semantic queries and to deliver accurate information (Decker, 1999). Therefore, the retrieval results are more relevant to the search query through ontology-based query expansion (Wache, 2001). Compared to conventional information retrieval tools, ontology-based semantic retrieval methods are able to use the semantics of content to guide query formation, provide answers with well-defined syntax and semantics for further processing, and reveal implicit information of background knowledge (Shah, 2002).

Controlled vocabularies also aid in achieving interoperability among metadata content and systems. These semantic systems are generated to select an authorized term for a concept and reference variant forms and words. When multiple information systems use the same controlled vocabulary, they support semantic interoperability. This practice is historically common in library catalogs and has great potential for data repositories, where their application is not as common.

Controlled vocabularies for scientific data

Conducting a literature review on the application of controlled vocabularies for scientific data is challenging, given the wide diversity and scope of items that qualify as data. Another challenge is that there are various scientific communities—each demonstrating different levels of understanding what is a controlled vocabulary. For example, biologists understand scientific taxonomy, which is a form of controlled vocabulary, although chemists' understanding may be more specific to the elements on the periodic table and or their variable names on a spreadsheet.

With the move toward open data and NSF data sharing policies (NSF, 2011; OSTP, 2013), it seems timely to consider the role of controlled vocabularies for managing scientific data and helping it to be accessible in the best possible way. There have been several notable registry-like efforts over the last near-decade that have made controlled vocabularies accessible for scientific data and related resources. Two examples include the National Center for Biological Ontologies (NCBO) bioportal (NCBO, 2015), and BioSharing (BioSharing, 2015), both of which provide access to many controlled terminologies. These developments and similar undertaking have been successful in providing access to vocabularies. Even so, access to multiple controlled vocabularies (including ontologies) in a single repository is limited, thus impacting data organization, search, access, and interoperability. This predicament underscores the need to explore this topic in greater detail. Specifically, should and how can repositories support access to multiple, appropriate, controlled vocabularies? The fact that scientific data also have many facets, not only topical, but also geographical, temporal, spatial and others, further speaks to their applicability for improving the management of digital scientific data. These noted limitations informed the goals and objectives that shaped this study, and they are reviewed in the next section of this paper.

GOALS AND OBJECTIVES

The goals of this study were to address the challenges noted with controlled vocabularies for data. Specifically, we targeted two research areas: the current state of controlled vocabularies and the potential improvement of their features and functions.

Research Questions:

- Who is using the semantic ontologies?
- How they are using the semantic ontologies?
- What functions do they want in the future?

It is necessary to sort users of controlled vocabularies into groups by how they use the controlled vocabularies. This organization contributed to our understanding of *what users would like*; *what people who work on repositories would like*; and *whether or not there are mutual goals*. We defined various user groups so that we could examine the use of and desired functions among different user groups. Finally, we addressed the future needs of functionalities for controlled vocabularies.

METHODS

We distributed a survey to NSF DataNet partners to collect and analyze original data.

Data Collection

The data was collected through a web survey conducted by the Metadata Research Center at UNC-Chapel Hill (now at Drexel). The survey link was distributed to the DataNet

community. The web survey targeted DataNet partners affiliated with DataONE (Data Observation Network for Earth), DFC, SEAD (Sustainable Environment Actionable Data), and TerraPop (Terra Populus). Scientists, curators, administrators, and repository developers who were involved in the deposit or management of scientific research data in repositories could participate in this study. The survey was open for response from May 17, 2013 to July 15, 2013. In total, there were 243 participants involved in the survey who answered questions about their duties, usage of data repositories and controlled vocabularies, and future desired tools of controlled vocabularies. The data is published and citable at figshare (Rowell et al., 2015).

Data Analysis

The data was collected and analyzed using descriptive statistics with SPSS. The analysis starts by identifying the primary controlled vocabularies used by participants. We analyze participants' answers about their current use of controlled vocabularies generally, and then group users by their work duties and analyze different groups' use of functions related to controlled vocabularies. We also find participants' future desire for functionality of controlled vocabularies, compare with their current use, and compare different groups' results.

RESULTS

The results present the state of controlled vocabularies and future desired functionalities.

Which vocabularies are being used?

The most widely used controlled vocabulary is LCSH, a finding that is consistent with other studies in metadata-creation practices (Ma, 2007; Park, 2010). This vocabulary is followed by several domain-specific vocabularies, such as TGN and ITIS. This is reasonable because DataONE provides scientific data archiving for ecological and environmental data produced by scientists worldwide. Other controlled vocabularies that are used less frequently are MeSH, NBII and EnvThes. AGROVOC and ERIC are vocabularies that are used. Meanwhile, 38.6% of users stated that they used none of the listed controlled vocabularies, implying there is a lack of available controlled vocabularies for some specific domains.

Who is using them and how?

In the survey, participants were asked to describe how they worked with a data repository in the last 12 months. Then they were classified into 4 groups: data contributors, data curators, administrators and repository developers. Data contributors are users who deposited research data with a data repository. Data curators are users who manage research data deposited with a data repository. Administrators are researchers who Served as a PI, co-PI, or full-time employee of an NSF DataNet Partner. Developers are users who developed systems, software, or other infrastructure to support a data repository. Participants were also able to describe their duty in an open question,

which will be analyzed later. Participants could select multiple duties and they could be identified with more than one role.

Of the total 243 participants, 32.5% participants recognized themselves as data contributors, 46.5% participants worked as data curators, 11.1% participants were administrators and 60.1% participants were repository developers. Therefore, the majority of participants are repository developers, data curators and data contributors. 31.3% participants also described their duty in the open question.

In the four groups of users, data contributors, data curators and repository developers work with online controlled vocabularies mostly. For the analysis, we focused on these three groups and investigated their use of features of controlled vocabularies.

| Functions currently used/supported | Yes | No | Don't Know |
|--|--------|--------|------------|
| Entering free text | 76.80% | 18.30% | 4.90% |
| Selecting from a single controlled vocabulary | 67.60% | 24.60% | 7.70% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 44.40% | 43.00% | 12.70% |
| Annotating a subject term selected from a controlled vocabulary | 24.60% | 59.20% | 16.20% |
| Using software to generate suggested subject terms selected from a controlled vocabulary | 26.10% | 64.80% | 9.20% |

Table 1. Functions that are currently supported

In total, there were 142 participants answering questions about current use of functions of controlled vocabularies. In general, 109(76.8%) participants had their data repositories supporting functions to enter free text, and 96(67.6%) had data repositories supporting functions to select subject term from a single controlled vocabulary. 63(44.4%) participants worked with data repositories that supported multiple controlled vocabularies and only 35(24.6%) participants could annotate a subject term selected from a controlled vocabulary with their data repository. Moreover, only 37 (26.1%) participants were able to use software to automatically generate suggested subject terms selected from a controlled vocabulary.

| Actions Performed | Yes | No | Don't Know |
|--|-------|-------|------------|
| Entering free text | 89.8% | 6.8% | 3.4% |
| Selecting from a single controlled vocabulary | 81.4% | 13.6% | 5.1% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 49.2% | 40.7% | 10.2% |
| Annotating a subject term selected from a controlled vocabulary | 28.8% | 57.6% | 13.6% |
| Using software to generate suggested subject terms selected from a controlled vocabulary | 30.5% | 62.7% | 6.8% |

Table 2. Actions performed by data contributors

Currently when providing subject terms, 53(89.8%) data contributors entered free text and 48(81.4%) selected terms from single controlled vocabulary. 29(49.2%) selected terms from multiple controlled vocabularies. Only 17(28.8%) data contributors annotated subject terms selected from a controlled vocabulary, and 18(30.5%) had used software to generate suggested subject terms selected from a controlled vocabulary.

| Functions currently supported for data curators | Yes | No | Don't Know |
|--|-------|-------|------------|
| Entering free text | 84.4% | 13.0% | 2.6% |
| Selecting from a single controlled vocabulary | 70.1% | 24.7% | 5.2% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 41.6% | 48.1% | 10.4% |
| Annotating a subject term selected from a controlled vocabulary | 23.4% | 59.7% | 16.9% |
| Using software to generate suggested subject terms selected from a controlled vocabulary | 27.3% | 66.2% | 6.5% |

Table 3. Functions currently supported for data curators

For functions related to provide subject terms, 65(84.4%) data curators had their data repositories supporting entering free text, 54(70.1%) had ones supporting selecting terms from single controlled vocabulary, 32(41.6%) had their

repositories enabling multiple controlled vocabularies, only 18(23.4%) had function on repositories to annotate subject terms and 21(27.3%) had ones with features to generate subject terms automatically.

| Functions current support for repository developers | Yes | No | Don't Know |
|--|-------|-------|------------|
| Entering free text | 72.7% | 22.2% | 5.1% |
| Selecting from a single controlled vocabulary | 70.7% | 23.2% | 6.1% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 48.5% | 42.4% | 9.1% |
| Annotating a subject term selected from a controlled vocabulary | 24.2% | 59.6% | 16.2% |
| Using software to generate suggested subject terms selected from a controlled vocabulary | 28.3% | 65.7% | 6.1% |

Table 4. Functions currently supported for repository developers

72(72.7%) of data repository developers had their data repositories supporting function to enter free text, and 70(70.7%) had their repositories supporting function to select from a single controlled vocabulary. 48(48.5%) of developers had functions on data repositories to support selecting subject terms from multiple controlled vocabularies, and only 24(24.2%) had functions on data repositories to support annotating a subject term selected from a controlled vocabulary and 28(28.3%) had tools to use software to generate suggested subject terms selected from a controlled vocabulary.

What users want in the future

| Functions wanted | Yes | No | Don't Know |
|--|--------|--------|------------|
| Entering free text | 74.60% | 9.90% | 15.50% |
| Selecting from a single controlled vocabulary | 76.10% | 10.60% | 13.40% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 69.70% | 11.30% | 19.00% |
| Annotating a subject term selected from a controlled vocabulary | 57.70% | 12.00% | 30.30% |

| | | | |
|--|--------|-------|--------|
| Using software to generate suggested subject terms selected from a controlled vocabulary | 66.90% | 8.50% | 24.60% |
|--|--------|-------|--------|

Table 5. Functions that users want

Participants who answered the current use of controlled vocabulary also answered questions about what functions they would like to use in the future. For the total 142 participants, 106(74.6%) of them want to continue using function to enter free text and 108(76.1%) want to select terms from single controlled vocabulary. 99(69.7%) participants want to select terms from multiple controlled vocabularies. Total number of participants who want to annotate subject terms from controlled vocabulary increase to 82(57.7%). 95(66.9%) want to use software to generate suggested subject terms in the future.

| Functions data contributors would like to use in the future | Yes | No | Don't Know |
|--|--------|--------|------------|
| Entering free text | 83.10% | 11.90% | 5.10% |
| Selecting from a single controlled vocabulary | 88.10% | 8.50% | 3.40% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 83.10% | 8.50% | 8.50% |
| Annotating a subject term selected from a controlled vocabulary | 67.80% | 13.60% | 18.60% |
| Using software to generate suggested subject terms selected from a controlled vocabulary | 79.70% | 8.50% | 11.90% |

Table 6. Functions data contributors would like to use in the future

In the future, when data contributors would like to provide subject terms, 49(83.10%) of them want to enter free text, 52(88.10%) of them want to select terms from a single controlled vocabulary, and 49(83.10%) want to select terms from multiple controlled vocabularies. 40(67.80%) contributors want their repositories to have functions to annotate terms, and 47(79.70%) want their repositories to support software to generate terms. Compared to the percentage of users with each functions, 20(33.9%) more data contributors would like to select terms from multiple controlled vocabularies, 23(39%) more want to annotate subject terms from controlled vocabulary and 29(49.2%) more want to use software to generate suggested subject terms.

| Functions data curators would like to support | Yes | No | Don't Know |
|--|-------|-------|------------|
| Entering free text | 79.2% | 9.1% | 11.7% |
| Selecting from a single controlled vocabulary | 81.8% | 9.1% | 9.1% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 67.5% | 14.3% | 18.2% |
| Annotating a subject term selected from a controlled vocabulary | 54.5% | 13.0% | 32.5% |
| Using software to generate suggested subject terms selected from a controlled vocabulary | 66.2% | 11.7% | 22.1% |

Table 7. Functions data curators would like to support

In the future for functions related to provide subject terms, 61(79.2%) data curators want their data repositories to support entering free text, 63(81.8%) want the repositories to have functions to select terms from single controlled vocabulary and 52(67.5%) want the repositories to support selecting terms from multiple controlled vocabularies. Meanwhile, 42(54.5%) data curators would like their repositories to have functions to annotate terms and 51(66.2%) want their repositories to enable to generate terms automatically. There is 20(25.9%), 24(31.1%), and 30(38.9%) increase from the number of current use of functions to desired use of functions of selecting terms from multiple controlled vocabularies, annotating subject terms and generating terms automatically, respectively.

| Functions repository developers would like to support | Yes | No | Don't Know |
|--|-------|-------|------------|
| Entering free text | 76.8% | 7.1% | 16.2% |
| Selecting from a single controlled vocabulary | 76.8% | 11.1% | 12.1% |
| Selecting from multiple controlled vocabularies when describing a single dataset | 69.7% | 11.1% | 19.2% |
| Annotating a subject term selected from a controlled vocabulary | 58.6% | 11.1% | 30.3% |

| | | | |
|--|-------|------|-------|
| Using software to generate suggested subject terms selected from a controlled vocabulary | 70.7% | 6.1% | 23.2% |
|--|-------|------|-------|

Table 8. Functions repository developers would like to support

In the future, if possible, 76(76.8%) data repository developers want their repositories to support entering free text and selecting from a single controlled vocabulary. 69(69.7%) would like their repositories to support selecting from multiple controlled vocabularies. 58(58.6%) want the data repository to have function to annotate a subject term selected from a controlled vocabulary. 70(70.7%) want the data repositories to support using software to generate suggested subject terms selected from a controlled vocabulary. There is 21(21.2%), 34(34.4%), and 42(42.4%) increase in developer's desire to have repositories support functions to select terms from multiple controlled vocabularies, annotate subject terms and generate terms automatically, respectively.

DISCUSSION

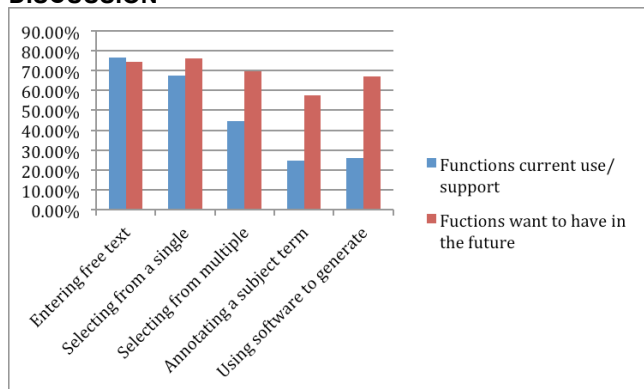


Figure 1. Functions that are currently used and desired in the future by all participants

Figure 1 shows the functions that are currently used by participants compared to functions that are desired in the future by all groups of users. The majority of users had functions to enter free text function and select terms from a single controlled vocabulary in their data repositories. About 40% users had annotation of subject terms function and less than 30% users had multiple controlled vocabularies supporting or software to generate terms in repositories. For future desired functions, though about 70% users want to keep the functions with entering free text and selecting from a single controlled vocabulary functions, many more users who don't have advanced tools, express the desire to have those tools to select terms across different disciplines, annotate terms, and generate terms automatically. The number of users who would like the data repository to support selecting subject terms from multiple

vocabularies is 1.56 times the number of current users who had this function on their repository. Similarly, the percentages of users who want the data repository to enable annotating subject terms and using software to generate terms are 2.35 and 2.56 times of the percentages of users who had these relevant functions. The dramatic increases imply a lack of advanced tools in current data repository and a huge demand for these controlled vocabulary services.

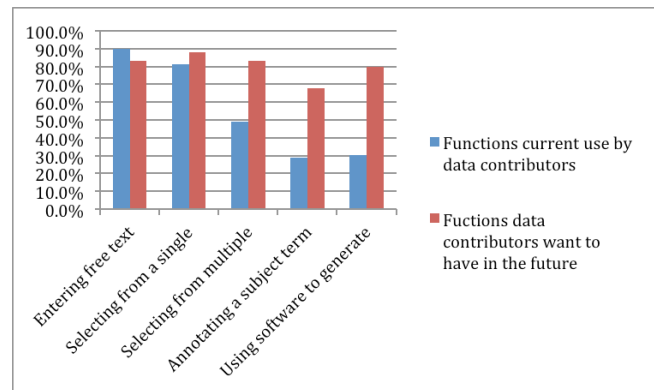


Figure 2. Data contributors

Figure 2 shows the percentage of data contributors who used or desired to use each function on data repository. The result is consistent with general users' result in Figure 1. Most data contributors made use of the first two basic functions. Less than half of data contributors used multiple vocabularies and less than 30% used annotation or software to generate terms. From their future desire we find most data contributors want to continue using functions to enter free text and select terms from a single controlled vocabulary. Meanwhile, the percentages of users who want to select terms from multiple vocabularies, annotate a subject term and use software to generate terms are 1.69, 2.35, and 2.61 times of the percentages of current users, which increase dramatically. It implies that many data contributors don't have those advanced tools available on their repository but they have strong interest in using them. They are a large potential user group of multi-disciplinary vocabularies, annotation function, and automatic indexing software.

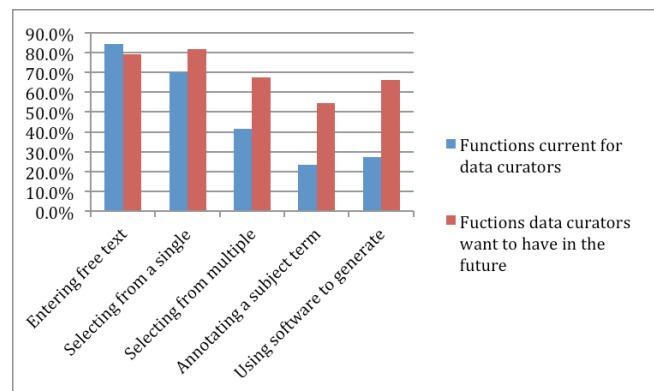


Figure 3. Functions used by data curators

Figure 3 shows the percentage of data curators who used or would like to use each function on data repository. From the results, we find most data curators had their repositories supporting functions to enter free text and select subject terms from a single controlled vocabulary. However, less than half of repositories support multiple controlled vocabularies and less than 30% of repositories support annotation or automatic indexing. It is surprising to find the controlled vocabulary tools of current data repositories so outdated and unintelligent.

Like data contributors, data curators want to keep the two elemental functions that are most used and also show a large interest in having the repository supporting multiple controlled vocabularies, annotation, and automatic indexing functions. The percentages of users who want to have the three functions are 1.62, 2.33 and 2.42 times of the percentages of users who had the relevant functions. This result shows that large groups of data curators also would like to use advanced tools. Their demand for functions with controlled vocabularies is huge and unsatisfied.

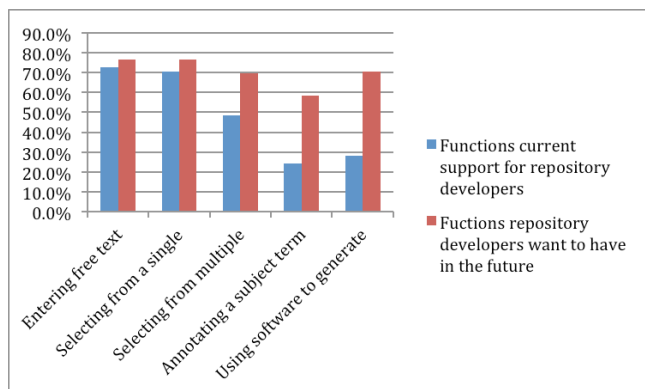


Figure 4. Functions supported for repository developers

Figure 4 shows the percentage of repository developers who use or would like to use each function. Similarly, the majority of repository developers had functions with entering free text and selecting terms from a single controlled vocabulary. Less than 50% had functions support multiple vocabularies and less than 30% had functions to annotate a subject term or use software to generate subject term. Like data contributors and data curators, most repository developers want to keep the two primary tools but also have desires to have more advanced tools. The percentages of developers who want to have tools to select terms from multiple vocabularies, annotate a subject term, and generate subject terms automatically are 1.44, 2.42, and 2.50 times of the percentages of current developers who had these tools. Therefore, we note that most data repository developers not only want the data repositories to support basic function, but also want to have the advanced tools. This shows their needs for advanced tools and also implies that developers have the intention and interest to develop such tools or install them on their data repository.

Most of the three user groups want to continue using functions to enter free text and to select terms from a single controlled vocabulary. Although many users had not used advanced tools or did not support them on their data repository, all groups of users indicate a strong interest in having such tools to select terms from multiple controlled vocabularies, annotate subject term and generate suggested subject terms automatically. The results indicate that most users, regardless of groups, want those tools available on their repository and would like to make use of them.

CONCLUSION

In this paper, we performed research on the current status of controlled vocabulary functionality and the potential for future desirable tools. We explored participants' use of both basic and advanced controlled vocabulary technologies, identified and labeled user groups, and investigated different groups of users' interests.

Our key findings:

- Most participants and data repositories already support basic functions of controlled vocabulary
- Many users do not have advanced controlled vocabulary tools on their repositories
- All groups of users want to keep basic tools such as entering free text and selecting subject terms from a single controlled vocabulary
- Most users express a strong interest in having and using advanced tools to select subject terms from multiple controlled vocabularies, annotate subject terms from controlled vocabulary, and generate subject terms automatically

The results indicate that there is a lack of those advanced tools in data repositories. These findings provide an impetus for technologies like HIVE to be developed and installed on repositories to satisfy the majority of users' needs.

Next Steps

As a potential next step in this research, we could design a new survey to ask users who already have access to advanced tools and find out whether or not those tools can improve their work quality or efficiency. Similarly, researchers could conduct an experiment to let users work with basic tools and advanced tools, and compare their generated subject terms. Through these experiments, researchers could investigate whether and to what extent the advanced tools can assist users to obtain significantly better results compared to basic tools.

ACKNOWLEDGMENTS

We would like to acknowledge National Science Foundation under Grant Number #OCI-0830944 and #OCI-0940841 for their support of this research.

REFERENCES

- Baker, T., Vandenbussche, P.-Y., & Vatant, B. (2013). Requirements for vocabulary preservation and governance. *Library Hi Tech*, 31(4), 657-668. Doi: 10.1108/LHT-03-2013-0027
- BioSharing. (2015). Retrieved from <https://www.biosharing.org/>
- Conway, M. C., Greenberg, J., Moore, R., Whitton, M., & Zhang, L. (2013). Advancing the DFC Semantic Technology Platform via HIVE Innovation. Proceedings of the 7th Metadata and Semantics Research Conference. Thessaloniki, Greece, November 19-22, 2013, pp. 14-21. Doi: 10.1007/978-3-319-03437-9_3
- DataNet Federation Consortium (DFC). (2015). Retrieved from <http://datafed.org/>
- Decker, S., Erdmann, M., Fensel, D., & Studer, R. (1999). *Ontobroker: Ontology based access to distributed and semi-structured information* (pp. 351-369). Springer US.
- Greenberg, J., Losee, R., Pérez Agüera, J. R., Scherle, R., White, H., & Willis, C. (2011). HIVE: Helping Interdisciplinary Vocabulary Engineering. *Bulletin of the American Society for Information Science and Technology*, 37(4): 23-26. Retrieved from http://www.asis.org/Bulletin/Apr-11/AprMay11_Greenberg_etAl.html
- Greenberg, J., Rowell, C., Rajavi, K., Conway, M., & Lander, H. (2013). HIVEing across U.S. DataNets. Research Data Management Implementations Workshop, NSF/Coalition for Academic Scientific Computation (CASC), Arlington, VA, March 13-15, 2013.
- Harper, C. A., & Tillett, B. B. (2009). Library of Congress controlled vocabularies and their application to the Semantic Web. *Cataloging & Classification Quarterly*, 43(3-4), 47-68. Doi: 10.1300/J104v43n03_03
- Isaac, A., & Baker, T. (2015). Linked data practice at different levels of semantic precision: The perspective of libraries, archives and museums. *Bulletin of the Association for Information Science and Technology*, 41(4), 34-39.
- Lanzenberger, M., Sampson, J., & Rester, M. (2010). Ontology visualization: Tools and techniques for visual representation of semi-structured meta-data. *Journal of Universal Computer Science* 169(7), 1036-1054.
- Long Term Ecological Research Network (LTER). (2012). LTER Controlled Vocabulary Tools & Resources. Retrieved from http://im.lternet.edu/vocab_resources.
- Ma, J. (2007). SPEC Kit 298: Metadata. Washington, DC: Association of Research Libraries: 13, 28. <http://publications.arl.org/Metadata-SPEC-Kit-298/>
- Nagypál, G. (2005). Improving information retrieval effectiveness by using domain knowledge stored in ontologies. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops* (pp. 780-789). Springer Berlin Heidelberg.
- National Center for Biological Ontologies (NCBO). (2015). BioPortal. Retrieved from bioportal.bioontology.org/
- National Information Standards Organization (NISO). (2005). Guidelines for the Construction, Format, and Management of Monolingual Thesauri. Retrieved from http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf.
- National Science Foundation (NSF). (2011). Dissemination and Sharing of Research Results. Retrieved from <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- Office of Science and Technology Policy (OSTP). (2013). Expanding Public Access to the Results of Federally Funded Research. Retrieved from <https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>
- Park, J.-R., Tosaka, Y. (2010). Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability. *Information Technology and Libraries*, 29(3), 104-116.
- Rowell, C. (2013). Controlled vocabulary use by data repositories: Determining status and potential for promoting interoperability. Retrieved from <https://cdr.lib.unc.edu/record/uuid:3d55802c-54c1-44ee-b056-54be1c142a52>.
- Rowell, C., Greenberg, J., Zhang, Y., Ogletree, A. (2015). Controlled Vocabularies for Scientific Data: Users and Desired Functionalities. figshare. Retrieved from <http://dx.doi.org/10.6084/m9.figshare.1483478>
- Shah, U., Finin, T., Joshi, A., Cost, R. S., & Matfield, J. (2002). Information retrieval on the semantic web. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 461-468). ACM.
- SKOS Simple Knowledge Organization System. <http://www.w3.org/2004/02/skos/>
- Svenonius, E. (1986). Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, 37(5), 331-340.
- Wache, H., Voegelé, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., & Hübner, S. (2001). Ontology-based integration of information—a survey of existing approaches. In *IJCAI-01 workshop: ontologies and information sharing* (Vol. 2001, pp. 108-117).