

Subject-based knowledge organisation: An OER for supporting (digital) humanities research

Koraljka Golub^{a,*} and Olivia Pestana^b

^a*Institute, Department of Cultural Sciences, Faculty of Arts and Humanities, Linnaeus University, Växjö, Sweden*

^b*Department of Communication and Information Sciences, Faculty of Arts and Humanities, University of Porto, Porto, Portugal*

Humanities scholars can today engage in research inquiry using data from a range of varied collections which are often characterised by poor subject access, often resulting in systems that underperform and even effectively prevent access to data, information and knowledge. In spite of the availability of professional standards and guidelines to provide quality-controlled subject access through knowledge organisation systems (KOS), subject access in such collections is rarely based on KOS. At the same time, KOS themselves may come with problems such as being slow to update, being rigidly structured and not incorporating end-users' vocabulary. It may therefore be useful to consider methods for remediating these deficiencies in KOSs, such as collecting user-generated metadata via social tagging or complementing automated indexing techniques with manual ones. To help address the above problems, the paper discusses these challenges and points to possible solutions in different contexts. It does so by reflecting on an open educational resource (OER) devoted to this theme, titled Introduction to Knowledge Organisation Systems for Digital Humanities. It was developed as part of an EU project called DiMPAH (Digital Methods Platform for the Arts and Humanities), 2021–2023, creating seven OERs for inclusion in DARIAH Teach.

Keywords: Knowledge organisation systems, social tagging, automatic subject indexing, subject access, user search interface, open educational resource, OER, DARIAH Teach, DiMPAH

1. Introduction

Humanities scholars can today engage in a historical, cultural, or linguistic inquiry using data (or “capta”) from archives, open corpora, administrative records, publisher databases, websites and more (Borgman, 2021). However, the plethora and diversity of digitally available information resources poses immense challenges for the field of knowledge organisation (ibid.), leading to many domains to opt for automatic solutions (most plainly demonstrated in how web pages are indexed and searched based on automated techniques). Yet automated systems are not a silver bullet in all cases: they underperform and even effectively prevent access to data, information and knowledge.

*Corresponding author: Koraljka Golub, Institute, Department of Cultural Sciences, Faculty of Arts and Humanities, Linnaeus University, H325, Hus F, Växjö, Sweden. E-mail: koraljka.golub@lnu.se.

This has been particularly detrimental in the humanities, both for secondary information resources like publications (Golub et al., 2020) as well as for primary resources (e.g., museum objects, archival documents, research data sets) that are often crucial research resources for the humanities researcher (see, for example, Golub et al., 2022).

In spite of international professional standards and guidelines to provide quality-controlled subject access based on knowledge organisation systems (KOS), such as information retrieval thesauri or subject headings systems, subject access is rarely based on KOS in archives and museums. This is problematic because relying solely on automatic methods will often fail – especially for objects with little or no text (like multimedia, digital twins, 3D models of artifacts, or artistic performance recording), old texts that have been too poorly digitised for reliable optical character recognition (OCR), or texts such as fictional or philosophical works that are so complex or abstract that even a human reader would be troubled in identifying the subject matter. Even libraries or databases of journal articles, which have subject metadata, often fail their users because their search interfaces do not adequately integrate and leverage said metadata.

At the same time, KOS themselves may come with problems such as being slow update and rigidly structured, not to mention imposing labels different from those of the end-user. It may therefore be useful to consider end-user metadata via social tagging. Automatic solutions, despite the reservations listed above, could also be of use if implemented in a complementary fashion. Furthermore, all these approaches (KOS, automated, and social) face challenges when it comes to new types of resources such as digital performing arts, digitised historical newspapers, digital twins, and 3D cultural objects.

To help address the above problems, the paper discusses these challenges and points to possible solutions in different contexts. It does so by reflecting on an open educational resource (OER) devoted to this theme, titled *Introduction to Knowledge Organisation Systems for Digital Humanities*. It was developed as part of an EU project called DiMPAH (Digital Methods Platform for the Arts and Humanities), 2021–2023, which as responsible the creation of seven OERs for inclusion in DARIAH Teach (<https://teach.dariah.eu>), a European-based platform for OERs in the digital arts, humanities and heritage (Papadopoulos et al., 2022). DARIAH Teach includes OERs that are extensible, open source and open access; they further asynchronous and flexible learning and allow easy localisation and adaptation.

The remainder of the paper is structured as follows. In the Background section we show the need for quality subject access to support research in (digital) humanities as well as point to the opportunity that OERs like our own provide in making the discipline of knowledge organisation available beyond its traditional context of library and information sciences (LIS). We present three different ways to organise resources and discuss their pros and cons in the Selected Approaches section. Under Use Cases examples of implementing the different approaches in practice are shown, including combined approaches. How these approaches were presented in the OER is described

in the next section. Finally, guidelines for future research and practice and provided in the Concluding Remarks section.

2. Background

2.1. Challenges of subject searching in the humanities

Searching by subject has proven to be very common amongst end-users despite being the most challenging type of search (see below). For example, subject searching is frequent in online search systems such as library catalogues (Hider & Liu, 2013; Hunter, 1991; Villén-Rueda et al., 2007), online museums (Baca, 2004; Liew, 2004), bibliographic databases (Siegfried et al., 1993), repositories (Heery et al., 2006), discovery services (Meadow & Meadow, 2012) and related digital search services (Patel et al., 2005). Subject search access to both primary resources (e.g. museum objects, archival documents like diaries or letters, research data sets) and secondary ones (e.g. academic publications) in humanities and heritage are needed by the researcher and the general public (some of whom are self-taught experts, e.g. hobbyists interested in genealogy), students, teachers (e.g. for preparing school lessons or excursions) and pupils, not to mention information experts who work in institutions curating and managing such content.

Finding resources online is directly dependent on the quality of search systems. In comparison to known item searching (e.g., queries for objects whose title, author, etc. is known beforehand), searching by subject is much more challenging. This is the result of difficulties in formulating search queries due to user's insufficient knowledge of the subject matter and/or of the online collection(s) and their resulting inability to use right search terms; insufficient knowledge of searching (i.e. how to formulate a search query to reflect the information need); as well as challenges arising from semantic ambiguities inherent to natural language such as polysemy, homonymy and synonymy that may lead to false positives, an overwhelming number of results and results missing.

Also, texts do not always explicitly name concepts that they write about. In many humanities disciplines and works of literary fiction language is often purposefully metaphorical. Texts from different historical periods often use different terms for the same concept and concepts will have different connotations over time. The problem is exacerbated with non-textual media, such as those often found in museums, which first require the visual system to be determined and expressed in text (Svenonius, 1994). New digital collections present further challenges, such as those of intangible cultural heritage, digital twins, 3D/AR cultural objects and historical newspapers.

In all, related research has shown that full-text searching is not enough. Knapp et al. (1998) established that the most effective way of online searching databases in the humanities is to combine free-text searching with the use of KOS-based indexing. KOS are particularly needed in large databases covering many subjects (Markey,

2007; Tibbo, 1994) as well as in databases of primary sources (Bair & Carlson, 2008) such as museum objects, which cannot be queried using full-text searches alone. Tibbo (1994) makes the point that the exponentially increasing volume of information objects available online leads to information overload and entropy, rather than increasing benefit from access to information. Although full-text indexing works for some tasks, for others it creates information overload and prevents the searcher from gaining a comprehensive overview on a topic: if a query returns thousands of retrieved documents, few searchers will browse beyond the first dozen or two hits.

2.2. OERs as an opportunity for knowledge organisation

Education in knowledge organisation is most often conducted within academic programs in library and information sciences (LIS) (Hider, 2018; Hjørland, 2022). With OERs becoming widely available (Mishra et al., 2022; Butcher, 2015, p. 5), an opportunity arises to share education on this topic to learners with different backgrounds, including those in the arts, cultural heritage, and humanities. Papadopoulos et al. (2022, p. 2) consider OERs the third pillar of education, given that they create “an additional pedagogic space for our students alongside the classroom (the first pillar), be it online or virtual, and more traditional secondary sources (articles, monographs, websites, videos, etc.), the second pillar”. Developing an OER in the form of a course contributes, on the one hand, to a student’s self-training and, on the other hand, helps teachers to create complementary opportunities to traditional training, stimulating individual reflection and the practice, helping fulfil learning outcomes.

2.3. Selected approaches

2.3.1. Knowledge organisation systems (KOS)

A process known as assigned subject indexing could be used as one solution to the challenges discussed in the Background section; and this is the dominant approach in the knowledge organisation (KO) community. In this process subject terms are taken from established KOS such as subject headings systems, thesauri and classification systems. These are designed to help the user select a more specific concept to increase precision, to discover broader or related concepts to increase recall, to help the user disambiguate between homonyms, or to discover which term is best used to name a concept. In addition, hierarchical browsing of classification schemes and other systems with hierarchical structures could help the user improve their understanding of their own needs and to formulate their queries more accurately.

The international ISO indexing standard of 1985, which was confirmed in 2020 (International Organization for Standardization, 1985), prescribes general techniques for subject indexing and clearly states that these are to be applied “by any agency in which human indexers analyse the subjects of documents and express these subjects in indexing terms” (International Organization for Standardization, 1985, p. 1), defining documents to be “any item amenable to cataloguing or indexing, specifically including

also non-print media and three-dimensional objects or realia". The standard gives a document-oriented definition of manual subject indexing as a process involving three steps: (1) determining the subject content of a document; (2) a conceptual analysis to decide which aspects of the content should be represented; (3) translation of those concepts or aspects into a KOS.

In order to counter high recall and low precision (i.e., a results list that includes many relevant items from the collection but far more irrelevant ones), a common problem in large text-based automated search systems, specific subject indexing should be implemented, involving (1) indexing policies that promote a high level of specificity and (2) indexing languages that are deep and detailed for any given topic (Tibbo, 1994). The KOS needs to be extensive in order to account for the fact that any topic can appear in many different contexts, and topics may be addressed from a very wide range of different perspectives (ibid.). Furthermore, specific disciplines will require their own specific KOS, rather than a one-size-fits-all approach (ibid.).

2.3.2. *Social tagging*

Since the emergence of Web 2.0 technologies, a complementary approach to professional subject indexing was found in social tagging services. However, social tagging is not the dominant approach-of-choice because of a number of disadvantages. These include a lack of indexing rules, different users using different words for the same concept, homonyms not being disambiguated, hierarchical and other relationships between tags being often not being there, tags being written in different forms (singular/plural, spelling variations etc.), tags being unlimited in quantity or having relevance for personal use only (e.g., a "to read" tag) (see, e.g., Furner, 2010; Kipp et al., 2015). At the same time, social tags are characterised by the natural everyday language that the users are familiar with and can relate to. In line with this, in her review of tagging literature, Rafferty (2018) concludes that while tagging may underperform in comparison to established subject indexing systems, it will still "complement, enrich, and . . . enhance conventional retrieval systems" (p. 510). These findings were concurred by Rolla (2009), Kipp and Campbell (2010), Golub et al. (2014).

2.3.3. *Automatic assigned indexing*

In spite of all the potential benefits of KOS, professional organisation of information based on KOS is very resource intensive. Just creating, managing, maintaining and updating KOS requires resources. Conducting subject indexing and classification based on the KOS for every single document takes lots of time for libraries, museums, archives, digital humanities projects etc., especially as they are increasingly facing the need to do more with less. This leads us to the sad situation we have today: many resources remain undiscovered due to increasing reliance on purely automatic approaches, and even when good KOS are used by an institution or a project, they are rarely visible at the level of the search interface (see below).

With this, cultural heritage institutions are failing to meet established cataloguing

objectives, especially the one where the user should be able to find everything there is on a certain topic in their database and retrieve only relevant information resources. Thus, another question arises, can we use automatic means to support KOS-based subject indexing?

Challenges with automation involve a number of issues. It is assumed that concepts have names, which can be more common in, for example, natural sciences, but much less so in humanities and social sciences. Determining a subject automatically is logically positivistic: a subject is considered to be a string occurring above a certain frequency, is not a stop word and is in a given location, such as a title (Svenonius, 2000, pp. 46–49). As mentioned earlier, automation is hard in documents with little text, or where text does not specifically mention terms for concepts it addresses, which is common in humanistic writings. Also, in algorithms inferences are made such as: if document A is on subject X, then if document B is sufficiently similar to document A (e.g., they share similar words or references), then document B is on that subject. Again, this holds true only to a certain extent and in certain types of resources. Further, there are no theoretical justifications for vector manipulations, such as the cosine measure that is often used to obtain vector similarities and applying them for subject computation.

Automatic subject indexing is hard to compute because texts are a complex cognitive and social phenomenon, and cognitive understanding of text engages many knowledge sources, sustains multiple inferences, and involves personal interpretation (Moens, 2000, pp. 7–10). Automatic understanding of text involves linguistic coding (vocabulary, syntax, and semantics of the language and discourse properties), domain world knowledge, shared knowledge between the creator and user of the text, and the complete context of the understanding at a specific point in time including the ideology, norms, background of the user and the purposes of using the text.

Automatic subject indexing tools are not sufficiently robust to deal with all these complexities. So, to what degree can they be applied in practice? This leads us to another problem, that of evaluation: research of automatic tools is usually conducted in laboratory conditions, excluding the complexities of real-life systems and situations (Lancaster, 2003, p. 334) and some scholars have been calling for this situation to improve. For example, see, Golub et al. (2016) who propose a comprehensive evaluation framework involving three major steps: 1) evaluating indexing quality directly through assessment by an evaluator or through comparison with a gold standard; 2) evaluating the quality of computer-assisted indexing directly in the context of an indexing workflow, and; 3) evaluating indexing quality indirectly through analysing retrieval performance.

While evaluation approaches often assume that human indexing is best, and that the task of automatic indexing is to meet the standards of human indexers, more serious scholarship needs to be devoted to evaluation in order to further our understanding of the value of automatic subject assignment tools and to enable us to provide a fully informed input for their development and enhancement. Hjørland (2011) proposes that the ideal formula for the future of indexing is that the human indexer takes what

automatic indexing is good at (once this is understood), and invest their resources on the value-added indexing that requires human judgment and interpretation. This would be in line with machine-aided indexing (MAI) in operative systems like Medical Text Indexer or Data Harmony's MAI software which has been adopted by a number of organisations.

2.3.4. *The role of the search interface*

As written earlier, it is important to provide quality subject access to a vast range of heterogeneous information objects in digital collections in order to make them visible. This includes both primary (e.g. research objects) and secondary sources (e.g. publications). The general objective of subject indexing should be that it allows the user to find everything in the collection that is relevant to a certain topic. KOS need to be applied to help address challenges of the natural language in subject searching as well as scarce text of information objects (like museum objects). While international standards, policies and practices to support this are in place, the next question then is to what degree the benefits of KOS are implemented at the level of the search interface – it is only then that the KOS actually become useful to the end-user.

Literature has pointed to 18 functionalities common across cultural heritage institutions as well as 3 additional image-related ones that are important for collections with images and information objects other than publications (Golub et al., 2021). While more research involving users is needed to confirm the best ways to implement different functionalities, let us consider several of the suggested functionalities:

1. Searching by concepts from KOS, including individual facets or concepts that compose a complex term (e.g., a class). This includes the ability to search by not just final and complete built classes or pre-coordinated index terms, but also in-built facets of the classes or the index terms. Furthermore, this means that the user can search by a term rather than just a symbol of the class (the end-user should not be expected to use class symbols).
2. Browsing by concepts from KOS, which is especially useful for those new to the document collection. Most beneficial are hierarchically structured concept schemes, such as hierarchical classification systems or information retrieval thesauri. At the narrowest hierarchical levels there should be a manageable number of information resources – perhaps not more than several dozen or so. If there are many more, the structure should be further developed to include more narrower concepts.
3. Automatic translation of user search terms into KOS terms. If the user writes a synonym not used per se, the system automatically translates it into the preferred term denoting the same concept. The system resorts to the KOS to be able to do that.
4. Showing narrower terms and broader terms. When the user types a search term, also narrower and broader terms are shown for them to explore around and consider choosing a more specific or a more general term. It also helps the user with disambiguation.

5. Disambiguation – offering the user different concepts (e.g., are you looking for jaguar as an animal or jaguar as car?).
6. Linking in the metadata record all other information resources with the same index term. This allows the user to click on the term in the metadata records and directly retrieve all other metadata record with exactly the same term.

2.3.5. *Combining the best of the three worlds*

In the sections above we have discussed advantages of KOS for information search and retrieval. However, KOS may not be neutral but could instead mediate bias. When bias is integrated in the KOS, it could also prevent access to information. Commonly present biases in large Western KOS are dominance of Christianity, gender bias, and lack of topics related to minorities (for more information, see, for example, Olson, 2002). One example to address such problems is Europeana's WEAVE project that aims to help increase visibility of community and minority collections, such as the Roma heritage, through appropriate curation and metadata.

At the same time, KOS may be slow to update so that new terms are not possible to use in a search system. Unless the search terms known to the user are in the KOS or used by a document author, that document would not be retrieved even if relevant. Similarly, some older KOS in particular have been developed more to match the professional language or the language of the expert, rather than the end-user that could be using different terminology (consider Latin versus common language terms). Thus, the user entering a search term that is not in the KOS even if the concept is, would retrieve no documents unless a full-text retrieval is enabled and the author used the same term in creating the work as the user did when searching for it. Furthermore, it is also important to remember indexing consistency levels tend to be lower in larger KOS and in organisations with higher exhaustivity and specificity indexing policies (like large academic libraries).

Social or collaborative tagging has advantages of adding further user perspectives that may have been overseen by managers of KOS. This would include more recent terms that may not have made it yet into a KOS. However, the tags are often uncontrolled and even different singular/plural forms or spelling variants are counted as different subject terms; hierarchies that are so beneficial for subject browsing are very hard to derive based on social tagging only.

Automatic subject indexing does not require professional resources and may be thus affordable; however, services that want to ensure high quality subject access cannot rely solely on automatic solutions. While purely automatic solutions can show high consistency (a characteristic sometimes used in the literature to argue for automatic indexing), they can also be consistently wrong. Automatic methods are successful to the degree to which they are able to understand discourse, a challenge that is especially high in humanities resources.

Further, if using derived automatic indexing (keywords from text) rather than assigned automatic indexing (keywords from KOS), documents will match user query only if the author used the same term as user's search term. Documents also reflect

the author's bias that may well be reflected in automatically extracted terms. An advantage of derived indexing is that more recent terminology could be reflected in the document so users would be able to find a match when searching based on more recent terms or when looking for new concepts.

3. Use cases

The two use cases are given below to demonstrate the strengths of the three different approaches applied in different contexts. The first case shows how combining automatic, social and KOS-based approaches yields enhanced subject access points. The second one demonstrates the value of KOS-based approach for LGBTQI fiction where automatic methods do not work effectively due to the challenging characteristics of such documents, as discussed above.

3.1. *Enhancing social tagging with KOS and automatic suggestions*

Project named EnTag (Enhanced Tagging for Discovery) (Golub et al., 2014) aimed to explore the potential of applying an established KOS for enhancing social tagging, with the ultimate purpose of improving subject indexing and information retrieval. The KOS used was the Dewey Decimal Classification (DDC) with mappings to Library of Congress Subject Headings (LCSH). When coming to a document to be tagged, the user is first offered automatic suggestions from DDC/LCSH. The excerpt from the DDC hierarchy is also given for users to confirm the right context for the term.

The results of the study demonstrated the importance of KOS suggestions for indexing and retrieval in the following:

- To help produce ideas of which tags to use – the users appreciated the suggestions being offered rather than them needing to think of any on their own.
- To make it easier to find focus for the tagging – the users chose more specific terms than they would have thought of themselves. For example, while they would normally choose just the general term 'slavery', the more specific suggestions from the KOS allowed them to consider the more specific ones like 'slavery and emancipation' or 'slavery – political science'. This in turn helps the searcher distinguish between similar documents and is better aligned to indexing recommendations to always assign the most specific index term available.
- To ensure indexing consistency – a significant disadvantage of social tagging is thus addressed as users are more likely to choose same word forms and terms for the one concept, rather than their own variation when they do not have a KOS to choose from.
- To increase the number of access points in retrieval – combining automated, social and KOS-based indexing has been shown to increase the number of access points in comparison to just social tagging. This means that more relevant documents are likely to be retrieved.

For more about the study, please refer to Golub et al. (2014).

3.2. Discovering LGBTQI fiction

Fiction is characterised by language that is often symbolic, indirect or implicit, rather than specifically and explicitly manifesting its themes in the text. This is exacerbated in LGBTQI fiction as LGBTQI themes may be subtly present and expressed in contested historical language. Such textual characteristics make it challenging to apply automatic subject indexing.

Professional fiction indexing based on KOS is usually limited to genre, complemented with facets of time and place. This is in spite of the fact that users' complex needs often cover "a combination of different aspects, such as specific genres or plot elements, engagement or novelty" (Bogers & Petras, 2017, p. 15). This implies the need to address more subjective aspects of literary fiction known as "appeal characteristics" (Saricks & Wyatt, 2019) such as pacing, characterization, storyline, frame/setting, tone, and language/style. Many libraries therefore import social tag clouds for works of fiction from LibraryThing (Johansson & Golub, 2019).

Looking further into LGBTQI fiction in Sweden, the general KOS used by most public libraries, Swedish Subject Headings (SAO – Svenska ämnesord), covers only a few LGBTQI terms which are too general. In order to allow researchers and general public interested in LGBTQI fiction find what they are looking for in a better way, the Queerlit (QLIT) project was set up in which a specific thesaurus for LGBTQI themes was developed, using the international *Homosaurus* as a basis (for more information, see Golub et al., 2022).

The Queerlit project is also building a dedicated database with an advanced search interface to support the subject searching of Swedish LGBTQI fiction. A search interface is as important as underlying metadata to support good subject access for the end-user.

4. An OER for knowledge organisation in digital humanities

The OER created to introduce these topics – entitled *Introduction to Knowledge Organisation Systems for Digital Humanities* – is aimed at Master's students and working professionals in any area of arts and humanities or cultural institutions seeking to acquire introductory understanding of KOS and their applicability for search and retrieval of resources across a range of digital collections and retrieval systems. The information given is aimed at providing the learner with a foundation to help implement optimal solutions in cultural heritage collections, institutions and digital humanities projects, resulting in enhanced access to humanistic resources for researchers, the general public and other stakeholders. Learners who take the entire course would be able to understand main principles of knowledge organisation, structures of knowledge organisation systems, as well as premises for their application. In this way, the OER provides the learner with knowledge about knowledge organisation, to which they usually do not have access as part of their main education.

More specifically, by the end of the course, the learner should be able to:

- 1) Understand the key principles behind knowledge organisation systems in the context of cultural heritage and digital humanities research;
- 2) Understand the pros and cons of professional, social (i.e., non-professional) and automatic subject indexing in this context;
- 3) Understand the role of the search interface in subject access;
- 4) Develop basic knowledge organisation skill sets;
- 5) Evaluate the most appropriate knowledge organisation system and type of subject indexing, or any combination thereof, for a given challenge in the humanities/heritage context;
- 6) Identify which search interface functionalities need to be implemented for optimal subject access;
- 7) Apply appropriate knowledge organisation approaches to help tackle global challenges from a humanities/heritage perspective;
- 8) Develop a strategy for implementing a good combination of knowledge organisation approaches and systems, including subject access functionalities at the level of the interface.

However, as with other DARIAH Teach courses, only a part of the course can be taken by the learner or the teachers, so special consideration has been given to structuring the course materials to allow taking just one resource – a unit, a lesson, an interactive element, an assignment etc. – or a combination thereof.

The learning activities and instructional materials comprise a variety of learning modalities. The course is divided into four units, each containing several lessons. The lessons incorporate multimodal and interactive teaching materials, such as videos, quizzes, a timeline as well as extensive further reading to guide learners through the topics. Case studies and scenarios from cultural heritage and related fields are used to enable students to put into practice the theoretical concepts and best practices explored.

Course development over the three years of the DiMPAH project (2021–2023) took place under close guidance of DARIAH Teach team of experts. Furthermore, three focus group interviews with potential users – Master’s students, higher education teachers, cultural heritage professionals – were conducted (in Portugal, Sweden and France). All these informed the OER development.

5. Concluding remarks

We have discussed the key role that quality subject metadata play in making information resources findable (professional, social, automatic); we have also demonstrated the need for quality search and browse interfaces of information retrieval systems, both equally important for discovery. What is also important is to inform the design of both KOS and search interfaces by end-user requirements. The adoption of KOS for improved discovery cannot be divorced from approaches from related disciplines such as human computer interaction or information behaviour that would allow for

a clearer understanding of humanities scholars' information practices in specific contexts, as well as that of the general public. The user's needs must be thoroughly and continuously researched to inform the development of KOS, their implementation at the time of indexing and at the level of the search interface.

One illustrative example is that of the ResearchSpace project which demonstrated that DH research practices and the ways humanities scholars interact with their sources do not match with what is conventionally expressed in databases: data-based organisation does not always work for researchers like historians where the organising unit should be contextualised and form part of a linkable, variable narrative than an atomistic unit of datum. Understanding humanities researchers is key to understanding what kinds of knowledge organization systems, processes and standards we should create and provide. Future research should focus on gaining a deep understanding of the context of needs, search, interaction and use.

While the field of knowledge organisation has been finding applications in numerous areas outside its home field of LIS to help address the almost universal need for organising information, we have also witnessed that, in many domains of human endeavour, resources are being organised in an ad hoc manner, often resulting in systems that underperform or even effectively prevent access to data, information and knowledge. In order to help deliver the best solutions for organising resources in (digital) humanities, it is important to bring the two communities of research and practise together and explore their combined potential.

One way of doing is to make knowledge about possible solutions to the challenges of subject access in the arts, heritage and humanities available beyond the LIS educational programmes. Creating OERs in quality-controlled platforms such as DARIAH Teach provide a great potential for uptake of this knowledge in communities of arts, heritage and humanities scholars, whose subject information access needs have not been addressed by widely available solutions in existing information systems.

Acknowledgments

This work was co-funded by the Erasmus+ Programme of the European Union, project Digital Methods Platform for Arts and Humanities (2020-1-SE01-KA203-0778789). The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

- Baca, M. (2004). Fear of authority? Authority control and thesaurus building for art and material culture information. *Cataloging and Classification Quarterly*, 38(3-4), 143-151. doi: 10.1300/J104v38n03_13.

- Bair, S. & Carlson, S. (2008). Where keywords fail: Using metadata to facilitate digital humanities scholarship. *Journal of Library Metadata*, 8(3), 249-262. doi: 10.1080/19386380802398503.
- Bogers, T., & Petras, V. (2017). An In-Depth Analysis of Tags and Controlled Metadata for Book Search. In *Proceedings of iConference 2017 iSchools*. iConference Proceedings Vol. 2. doi: 10.9776/17002.
- Butcher, N. (2015). *A Basic Guide to Open Educational Resources (OER)*. Edited by Asha Kanwar and Stamenka Uvalić-Trumbić. UNESCO and Commonwealth of Learning. <http://hdl.handle.net/11599/36>.
- Furner, J. (2010). Folksonomies. In *Encyclopedia of Library and Information Sciences*, edited by Marcia J. Bates and Mary Niles Maack, 1858-66. CRC Press. <http://works.bepress.com/furner/5>.
- Golub, K., Bergenmar, J., & Humelsjö, S. (2022). Searching for Swedish LGBTQI fiction: challenges and solutions. *Journal of Documentation*, 78(7), 464-484. doi: 10.1108/JD-06-2022-0138.
- Golub, K., Lykke, M., & Tudhope, D. (2014). Enhancing social tagging with automated keywords from the Dewey Decimal Classification. *Journal of Documentation*, 70(5), 801-828. doi: 10.1108/JD-05-2013-0056.
- Golub, K., Tyrkkö, J., Hansson, J., & Ahlström, I. (2020). Subject indexing in humanities: A comparison between a local university repository and an international bibliographic service. *Journal of Documentation*, 76(6), 1193-1214. doi: 10.1108/JD-12-2019-0231.
- Golub, K., Ziolkowski, P.M., & Zlodi, G. (2022). Organizing subject access to cultural heritage in Swedish online museums. *Journal of Documentation*, 78(7), 211-247. doi: 10.1108/jd-05-2021-0094.
- Golub, K., Lykke, M., & Tudhope, D. (2014). Enhancing Social Tagging with Automated Keywords from the Dewey Decimal Classification. *Journal of Documentation*, 70(5), 801-28. doi: 10.1108/JD-05-2013-0056.
- Golub, K. (2016). Potential and Challenges of Subject Access in Libraries Today on the Example of Swedish Libraries. *International Information & Library Review*, 48, 204-10. doi: 10.1080/10572317.2016.1205406.
- Heery, R., Lyon, L., Tsinaraki, C., Brody, T., Koch, T. & Doerr, M. (2006). Report on Digital Repositories: An Evaluation Study on the Development and Implementation of Community Repositories to Support Research (And Learning and Teaching), DELOS2 Network of Excellence on Digital Libraries, Deliverable 5.1.1. <http://citeseerx.ist.psu.edu/viewdoc/download?jsessionid=19458833E5FA5CC45117776085BD0E87?doi=10.1.1.101.5976&rep=rep1&type=pdf> (accessed 30 November 2019).
- Hider, P. (2018). The terminological and disciplinary origins of information and knowledge organization. *Education for Information*, 34(2), 135-161. doi: 10.3233/EFI-180165.
- Hider, P., & Liu, Y.-H. (2013). The use of RDA elements in support of FRBR user tasks. *Cataloging and Classification Quarterly*, 51(8), 857-872. doi: 10.1080/01639374.2013.825827.
- Hjørland, B. (2022). Education in knowledge organization. ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli. <https://www.isko.org/cyclo/education>.
- Hunter, N.R. (1991). Successes and failures of patrons searching the online catalog at a large academic library: A transaction log analysis. *RQ*, 30(3), 395-402. <https://www.jstor.org/stable/25828813>.
- ISO (1985). Documentation – Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms (5963:1985). <https://www.iso.org/standard/12158.html>.
- Johansson, S., & Golub, K. (2019). LibraryThing for Libraries? How Tag Moderation and Size Limitations Affect Tag Clouds. *Knowledge Organization*, 46(4), 245-259. doi: 10.5771/0943-7444-2019-4-245.
- Kipp, M.E.I., Beak, J., & Graf, A.M. (2015). Tagging of banned and challenged books. *Knowledge Organization*, 42(5), 276-83. doi: 10.5771/0943-7444-2015-5-276.
- Kipp, M.E.I., & Campbell, D.G. (2010). Searching with tags: Do tags help users find things? *Knowledge Organization*, 37(4), 239-55.
- Knapp, S.D., Cohen, L.B., & Juedes, D.R. (1998). A natural language thesaurus for the humanities: The need for a database search aid. *Library Quarterly*, 68(4), 406-430. doi: 10.1086/603001.
- Lancaster, F.W. (2003). *Indexing and abstracting in theory and practice*. 3rd ed. University of Illinois.
- Liew, C.L. (2004). Online cultural heritage exhibitions: A survey of information retrieval features. *Program Electronic Library and Information Systems*, 39(1), 4-24. doi: 10.1108/00330330510578778.
- Markey, K. (2007). The online library catalogue: Paradise lost and paradise regained? *D-Lib Magazine*, 13(1/2). doi: 10.1045/january2007-markey.
- Meadow, K., & Meadow, J. (2012). Search query quality and web-scale discovery: A qualitative

- and quantitative analysis. *College and Undergraduate Libraries*, 19(2-4), 163-175. doi: 10.1080/10691316.2012.693434.
- Mishra, M., Dash, M.K., Sudarsan, D., Santos, C.A.G., Mishra, S.K., Kar, D., Bhat, I.A., Panda, B.K., Sethy, M., & da Silva, R.M. (2022). Assessment of trend and current pattern of open educational resources: A bibliometric analysis. *The Journal of Academic Librarianship*, 48(3), Article 102520. doi: 10.1016/j.acalib.2022.102520.
- Moens, M.F. (2000). *Automatic Indexing and Abstracting of Document Texts*. Kluwer.
- Olson, H.A. (2002). *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Kluwer.
- Papadopoulos, C., Rasterhoff, C., & Schreiberman, S. (2022). Open Educational Resources as the Third Pillar in Project-Based Learning During COVID-19: The Case of #dariahTeach. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, 6(1). doi: 10.18357/kula.205.
- Patel, M., Koch, T., Doerr, M., & Tsinaraki, C. (2005). Semantic Interoperability in Digital Library Systems, DELOS2 Network of Excellence on Digital Libraries, Deliverable 5.3.1, available at: <http://delos-wp5.ukoln.ac.uk/project-outcomes/SI-in-DLs/SI-in-DLs.pdf> (accessed 30 November 2019).
- Rafferty, P. (2018). Tagging. *Knowledge Organization*, 45(6), 500-16.
- Rolla, P.J. (2009). User tags versus subject headings: Can user-supplied data improve subject access to library collections. *Library Resources & Technical Services*, 53(3), 174-84.
- Saricks, J.G., & Wyatt, N. (2019). *The Readers' Advisory Guide to Genre Fiction*. Third Edition, ALA Editions.
- Siegfried, S., Bates, M.J., & Wilde, D.N. (1993). A profile of end-user searching behavior by humanities scholars: The Getty online project report no. 2. *Journal of the American Society for Information Science*, 44(5), 273-291. [https://doi.org/10.1002/\(SICI\)1097-4571\(199306\)44:5<273::AID-ASI3>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-4571(199306)44:5<273::AID-ASI3>3.0.CO;2-X).
- Svenonius, E. (1994). Access to nonbook materials: The limits of subject indexing for visual and aural languages. *Journal of the American Society for Information Science*, 45(8), 600-606.
- Svenonius, E. (2000). *The Intellectual Foundations of Information Organization*. MIT Press.
- Tibbo, H.R. (1994). Indexing for the humanities. *Journal of the American Society for Information Science*, 45(8), 607-619. [https://doi.org/10.1002/\(SICI\)1097-4571\(199409\)45:8<607::AID-ASI16>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-4571(199409)45:8<607::AID-ASI16>3.0.CO;2-X).
- Villén-Rueda, L., Senso, J.A., & De Moya-Anegón, F. (2007). The use of OPAC in a large academic library: A transactional log analysis study of subject searching. *The Journal of Academic Librarianship*, 33(3), 327-337.