Knowledge Federation
from a Knowledge Organization Perspective:
A Position Paper

Alexander Sigel
Cologne, DE
sigel@asigel.de

Abstract. The new field of knowledge federation and the existing field of knowledge organization share quite a number of objects of experience and understanding. In particular, they have the common objective of facilitating insight by supporting gradual subject-centric organization of knowledge (of inter-related concepts and of information resources relevant to such concepts) in an environment which is conceptually heterogeneous and where the federation of scattered knowledge fragments is impeded by obstacles on several layers. The Topic Maps approach is particularly apt for tackling such obstacles, both on the conceptual and technical level, i.e. likewise for humans and the Semantic Web.

This paper reviews and discusses crucial issues of joint interest for these fields from a modern knowledge organization perspective. It contributes to the advancement of knowledge federation research by reinterpreting basic ideas of conceptual knowledge organization in the light of new knowledge federation challenges pertaining to discourse communities with open, emergent knowledge systems.

The paper highlights how modern (distributed) knowledge organization (and distributed knowledge management) approaches, combined with Topic Maps, may cross-fertilize with and facilitate knowledge federation.

Determining the relevance of an information resource to a concept, and thus determining its subject, is a hermeneutic act somehow relative to the interpreter. Different communities and usage contexts entail validly differing viewpoints, leading to differing conceptualizations. However, traditional knowledge organization and traditional knowledge management ultimately had to force different viewpoints into one, centralized homogeneous conceptual system (e.g. a taxonomy). The internet increased conceptual heterogeneity and brought us a more democratic, user-contributed subject organization, lately e.g. in the form of folksonomies and tagsonomies (cf. the tagsocratic project).
1 Introduction

The term knowledge federation has been proposed by Dino Karabeg [1] for the concept comprising the “shared interest in (...) the creation, organization and presentation of knowledge (and of knowledge artifacts) where both the individual/partial and the collective/general views are maintained and reconciled, and which affords the intrinsic advantages of both”. The Call for Abstract for this Knowledge Federation Conference [2] states as a purpose of knowledge federation “to join together, organize, reconcile and present the fragmented pieces or sources of information that pertain to a given subject.” Knowledge federation strives at “bringing together heterogeneous information resources under possibly many different criteria, such as same context, or same subject”. “Ontology crafting” is mentioned as one example of a social mapping process in knowledge federation. Knowledge federation is said to have potential for “furthering the aims of the Semantic Web and knowledge organization”.

This all suggests that the existing field of knowledge organization and the new field of knowledge federation, though having different foci, share quite a number of objects of experience and understanding. Hence, it is worthwhile looking at selected commonalities in order to foster interchange and transfer.

Both knowledge federation and knowledge organization also share a particular relationship to the Topic Maps approach, which links them even closer. Topic mapping is seen as one means to realize improved federation, and several knowledge federation prototypes are built with topic maps. TMRAP, a topic maps-driven protocol for knowledge federation has been proposed in [3]. The relationship between knowledge organization and Topic Maps has been characterized elsewhere [4-6]. Modern knowledge organization theory allows for the accommodation of multiple viewpoints, and these can be modeled with Topic Maps in a natural way:


Dino Karabeg, citing this passage, states that this “view clearly points towards knowledge federation” [1].

This paper reviews and discusses crucial issues of joint interest for knowledge federation and knowledge organization from a modern knowledge organization perspective. It contributes to the advancement of knowledge federation research by
reinterpreting basic ideas of conceptual knowledge organization in the light of new knowledge federation challenges pertaining to discourse communities with open, emergent knowledge systems. It also highlights, how modern (distributed) knowledge organization (and distributed knowledge management) approaches, combined with Topic Maps, may cross-fertilize with and facilitate knowledge federation.

To this end, this paper is organized around the following questions:

- What is knowledge organization about and good for?
- Which are important issues in knowledge organization?
- How do knowledge organization, integration and federation combine and differ?
- Can knowledge organization support multiple conflicting viewpoints?

This will allow us to enter into the discussion of what knowledge federation challenges exist to which knowledge organization may contribute.

1 What is Knowledge Organization about and good for?

This section introduces knowledge organization (KO) and major approaches in order to be able to discuss commonalities between both fields thereafter.

Because KO, though mainly rooted in library, information, documentation and classification science, is a pluridisciplinary field centered around the process and result of the organization of knowledge, there exist diverse conceptualizations, definitions and methods. Which are major approaches in KO? Hjørland discerns between the following seven approaches to KO, proposing “to mediate between different views and develop the system in accordance with the goals and values of the users for which the system is intended” [7, 80]:

1. The traditional approach expressed by classification systems in libraries and databases, e.g. universal systems (c. 1876);
2. The facet-analytical approach, Ranganathan (1933);
3. The information retrieval tradition (1950s);
4. User-oriented and cognitive views (1970s);
5. Bibliometric approaches (1963);
6. The domain analytic approach (1994);
7. Other approaches. [7, pp. 87ff].

Where knowledge federation highlights individual views and their reconciliation into community views, I regard the user-oriented view, the domain analytic approach, and other approaches like discourse analysis, as most apt.

A wealth of material on KO can be found via ISKO, the International Society for Knowledge Organization (www.isko.org), in particular in the proceedings of the international and national conferences, and in the international journal “Knowledge Organization”.

Position paper (draft 2008-08-04)
http://knowledgefederation.org/Knowledge_Federation_2008
A classical definition by Dahlberg, the founding mother of KO, can be found in [5, p. 392f.]. It is structured around her “systematifier”, which according to McInerney (1997) “looks at the facets of any knowledge field to represent concepts embedded within it and can be applied to multidisciplinary, transdisciplinary and other forms of interdisciplinary endeavors”. Therefore I suggest that a good starter for structuring a definition of the knowledge federation field may be Dahlberg's systematifier.

To follow the current discussion on what KO is about, I highly recommend the recent special issue “What is Knowledge Organization?” [7] of the journal “Knowledge Organization”.

There, Hjørland [7, 86f., online at http://dlist.sir.arizona.edu/2402/] discerns between KO in the narrow meaning in the library and information science field, and KO in the broader meaning.

“In the narrow meaning Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, bibliographic databases, archives and other kinds of 'memory institutions' by librarians, archivists, information specialists, subject specialists, as well as by computer algorithms and laymen. KO as a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations, works and concepts.”

According to him, the broad meaning of KO relates both to “how knowledge is socially organized and how reality is organized.” He claims that “KO in the narrow sense cannot develop a fruitful body of knowledge without considering KO in the broader perspective”. Because knowledge federation clearly also relates to social KO processes, both KO in the narrow and broader sense are relevant to knowledge federation.

Tennis [7, 103] defines KO as “the field of scholarship concerned with the design, study and critique of the processes of organizing and representing documents that societies see as worthy of preserving.”

Both Tennis and Hjørland see KO also rooted in the theory of knowledge. For a thorough treatment of the philosophical and epistemological background of subject representation, I recommend Hjørland's works.

Earlier I have tried to describe what KO is in my view:

- “KO is the interdisciplinary field that theoretically reflects the practical activity of organizing knowledge for specific purposes and discourse communities” [5, p. 385]
- “KO is interested in optimizing the organization (the conceptual access structure) of knowledge repositories in order to support easier retrieval, creation and sharing of knowledge for user communities.” [3]
- “KO is about organizing objects of thought (and associated carriers of information) so that humans can work with them more easily. The central aim is improved access, more sophisticated finding aids, and a clearer overview. To this end, KO deals with structured metadata (for example,
indexing). KO expresses and orders statements about matter, which are comprised of concepts and relations.” [5, p. 384]

- “KO (...) is about how to arrive at and represent appropriate conceptual access structures to aid working with knowledge (knowledge networks or knowledge spaces).” [5, 386]

- “KO is about aiding people in their knowledge work by organizing knowledge through (virtual or overlaid) conceptual access structures. Such structures allow us, in addition to providing easier access (...), to gain overview and understanding.” [5, 397].

In this sense, KO is – in a nutshell - the field concerned with ordering concepts, which in turn allows ordering of all types of assets relevant to these concepts. KO provides a methodology for intellectually organizing knowledge in the form of assertions. Assertions are statements about subject matter, comprised of concepts and relations, which can be ordered according to various criteria. Concepts can be interconnected by typed relations, resulting in virtual conceptual overlay structures in the form of lightweight semi-formal ontologies. By organizing the concepts, the assets relevant to such concepts can be flexibly organized in multiple ways.

KO is the science of structuring and systematically arranging concepts (objects of thought, knowledge units/knowledge items), according to their inherent knowledge elements (characteristics), and the application of concepts and classes ordered this way for the assignment of worth knowing contents of referents (objects/subjects) of all kinds. In other words, if you want to organize information assets, KO provides a methodology which achieves order by organizing concepts as Knowledge Organization Systems (KOS) according to what defines these concepts, and by assigning statements to information assets about the aboutness of these assets (in the form of structured metadata composed of KOS elements). The statements represent what is implied by the asset’s content, i.e. why it is relevant to user communities.

A concept is defined by the synthesis of all necessary characteristics which can be stated about any referent (the essential statements or predicaments), represented by significations. According to Dahlberg's concept triangle the formal elements of a concept (knowledge unit) is defined by A. the referent (the subject referred to), B. the characteristics (the knowledge elements, or attributes), and C. the verbal form (the denotation, or name). The inner structure of concepts may be represented by interconnected categories, and complex concepts may be built by interconnecting concepts. Characteristics are also concepts and can themselves be systematically ordered according to formal categories. Categories are concepts gained from making the most general statement about a referent, e.g. “is a process”, “is a property”. Because categories are on a higher level, they seem to be more stable over time than ordinary concepts. Therefore, it is recommended to build KOS and indexing languages on top of categories, using conceptual analysis. KOS and indexing languages must contain all categories which shall be reliably found, anyway. This can guard against the arbitrary proliferation of concepts.
Concept relationships arise from existence of equal, similar, or functional dependent characteristics as components of concepts. The formation of concept systems is based on such concept relationships. Concept relationships can be systematically ordered, because characteristics can be systematically ordered. This can guard against the arbitrary proliferation of relations. According to Green, “relationships are at the very heart of knowledge organization” [7, p. 150ff.], but “despite the centrality of relationships, their expression in knowledge organization schemes seldom rises to full and systematic expression” [7, p. 158]. She suggests that studies showing that making use of the full and systematic expression of relationships does not improve retrieval effectiveness “may miss something”. “Either we are not implementing relationships properly, or we don’t understand the fundamental properties of the context in which we are working – or perhaps both!” [7, p. 158]. This leaves hope for a better understanding of relationships in knowledge federation and a better implementation with the Topic Map approach.

According to Svenonius (2000), “the essential and defining objective of a system for organizing information (...) is to bring essentially like information together, and to differentiate what is not exactly alike”, such that all relevant, however also only relevant information assets are found (collocation objective). The major question here is: Who shall define what is “like”? (Can this be valid for everyone? All the time and in all contexts? Etc.)

I claim that knowledge federation and knowledge organization have the common objective of facilitating insight by supporting gradual subject-centric organization of knowledge (of inter-related concepts and of information resources relevant to such concepts) in an environment which is conceptually heterogeneous and where the federation of scattered knowledge fragments is impeded by obstacles on several layers. Apart from logistic issues of scattered knowledge, and syntactic heterogeneity, the most interesting layer is semantic interoperability of the interlinked knowledge network. The remedy seems to be provision for an infrastructure supporting multiple viewpoints and their value-creating flexible mapping.

I also claim that the Topic Maps approach is particularly apt for “bringing together” scattered knowledge fragments, both on the conceptual and technical level, i.e. likewise for humans and the Semantic Web.

2 What are important issues in Knowledge Organization?

Gnoli presents “ten long-term research questions in KO” [7, 137], most of which I regard as relevant to knowledge federation research, so I will shortly comment on them:

Can KO principles be extended to a broader scope? Yes, I strongly believe that KO principles can be extended beyond the traditional LIS core. I would like to see more Topic Map systems implement KO principles, and I would like to find knowledge federation profiting from refurbished KO principles.
Can any ontological foundation be identified? Although general KOS should not be based on certain philosophical views and use certain philosophical ideas only as general structuring principles, in a federated environment it should be possible to have KOS being based on arbitrary philosophical views relevant for structuring. With respect to entities seen as dynamic over time rather than eternally stable, I support an evolutionary approach, but have not really seen practical solutions from KO. In a federated world, concepts changing their meaning over time may need more thought.

Gnoli writes: “If we really want to enable interoperability between different schemes and interdisciplinary research, we will always need some general scheme, at least as a switching device between systems based on different epistemologies”. Although a switching system minimizes the need for partial mappings, I doubt if in a federated world there will be only one switching system, given that ontology research has already come up with lots of upper ontologies. I rather believe that we will have conceptual overlays of trusted originators which we can flexibly switch on or off in our topic maps.

Should disciplines continue to be the structural base of KO? I agree with Gnoli that it should be possible to make statements about phenomena and retrieve them both independently of the disciplinary context and dependent on the disciplinary context. Because traditional KO has focused on universal KOS and on special KOS around disciplinary viewpoints, in knowledge federation land I see potential for reconciliation between these viewpoints.

How can viewpoint warrant be respected? Because KO, according to the epistemological claim, can be different to different communities, Beghtol had introduced the “viewpoint warrant”. Gnoli suggests a neutral switching system such that users could easily switch between different specific viewpoint systems. In addition, he would allow users to choose a KOS with a specific viewpoint as their preferred KOS. Of particular relevance to knowledge federation is the requirement articulated by Gnoli that “[u]sers of a system should be allowed to switch between different viewpoints, both to choose their preferred one, and to explore how related knowledge is expressed from different perspectives.” This could be implemented with the Topic Maps means of scoping and merging.

How can KO deal with changes in knowledge? An improved understanding may lead to changed characteristics, concepts, relations, and assertions. Therefore, in knowledge federation it must be possible to express versioning (e.g. predecessor, successor, and split relationships) and temporal validity, and this not only for one issuer, but for multiple viewpoints.

How can KOS represent all these dimensions? I agree with Gnoli that a modern analytico-synthetic would be able to express arbitrary content, including viewpoints. He points as potential problems out: cognitive overload and the right balance between analysis and synthesis. In sum, I am convinced that most existing KOS still bear the burden of decades of printed systems and do not take full advantage of electronic possibilities of digital semantic knowledge systems. It is possible to represent
arbitrary conceptualizations, but it will become more complicated, and it will be more
difficult to map between them.

How can software and formats be improved to better serve these needs? Again,
using the Topic Maps approach, or similar Semantic Web efforts, see e.g. [8], cited by
Gnoli.

Who should do KO? While traditionally this was the function of the KO
specialists/information professionals acting as gatekeepers, or of authors, now KO can
be done by readers, resulting in folksonomies, with the advantage of more
“democracy”, “as anyone can use his/her preferred terms (...) without being forced
into the rigidity of a pre-produced scheme”, and the disadvantage of “the obvious lack
of vocabulary control”. In my view, all three groups have their stake, because,
according to epistemological theory, they bring different viewpoints into the game.
Therefore, all contributors should be supported in knowledge federation systems. In
order to reduce the disadvantages of a pure bottom-up approach, Gnoli sees as current
trends the mediation by expert groups, and top-down tools, e.g. based on facet
analysis. For me, fuzzy.com and KO in crowdsourcing applications are very good
examples of systems where such an approach is natural and rising.

López-Huertas [7, p. 128] identified two broad research challenges: How can we
achieve quality in KO and KOS? How can we manage emergent knowledge in KOS?
Both are clearly relevant to knowledge federation, too.

3 How do Knowledge Organization, Integration and Federation combine and differ?

The common problem and object is that knowledge is fragmented. How do the
approaches to overcome this differ? In the following, I assume that “knowledge” is
something similar to “statements about subject matter”, whether relating to
information assets or not, and that the task is to bring statements together which are
regarded as like in a certain context.

Tanaka discerns integration from federation. Whereas in general “[f]ederation
assumes an open networked environment of heterogeneous, autonomous and
distributed resources”, he characterizes integration as “local and centralized
management and interoperation of resources in a closed environment” [9]. He defines
knowledge federation as “a way to select and combine multiple knowledge resources
from an open environment (...) into a single virtual entity representing a new
resource.” Tanaka further differentiates between knowledge federation and
knowledge organization/knowledge reorganization. In his view, knowledge
organization stands for the “structural organization (...) of knowledge resources”.

Karabeg [1] sees knowledge integration metaphorically like “knowledge
dictatorship”, and requires for knowledge federation to balance individual and
collective views, allowing members of the discourse community to “make their own qualified judgment”.

Indeed, traditional information and documentation systems and centers were oriented towards a central view which is defined by assuming what will serve the majority of the user community best. This view materialized as the thesaurus or classification system of this center, which had a coherent and hopefully consistent, but rather rigid structure. To this end, an artificial, neutral system was constructed which integrates views on concept definitions and their interrelations. Structural aspects are still prevailing, however relational and functional aspects are nowadays increasingly used. The environment was closed in the sense that authors or users could not change the document representation, and there was no room for alternative views. However, it was not completely closed, because centers moved from completely centralized to cooperative model between several institutions. The common view was still upheld by prescribing usage of the joint KOS, and by procedures of quality assurance. With the advent of the internet, documents of much more varying quality and background with a higher semantic heterogeneity had to be processed.

This led e.g. to Krause's (1996) work on a “layered model of information provision in which no longer a central agency exerts its authority in subject indexing and vocabulary control upon agencies located lower in the hierarchy, but in which a group of partners co-operate. Such a strategy does not result in uniform metadata, but leads to layers of heterogeneous metadata with different quality control procedures. Intellectually controlled high-quality subject schemata lie at the heart of those layers. Intelligent transfer components are sought which can improve on subject data on outer layers by using the structure of inner layers.” [4]

“Instead of normative authority we will have to live and copy with different layers of indexing quality and depth (among other attributes of heterogeneity). In addition to heterogeneity due to a multitude of independent actors, different domains and contexts require admission of different conceptualizations, which find their ways into KOS and subject metadata.” [5, p 416]

López-Huertas [7, p. 122 and 128] summarizes several years of KO research with the phrase “KOS design demands for integration of knowledge” and relates the quality of KO in the internet environment to the quality of knowledge integration. She discerns between two meanings of knowledge integration:

1. “[T]he capability for KOS to represent and organize knowledge avoiding a global, standardized view (...), and
2. “[T]he systems’ capabilities to improve design, to develop structural devices to represent and harmonize in classificatory structures heterogeneous information sources and different structures coexisting in the Internet.”

Semantic interoperability (comparability and compatibility of KOS) is a major issue in KO. However, semantic integration was rather focused on the integration of
whole KOS via switching systems and concept mapping. Semantic integration at the fine-granular assertion level has only begun with the Semantic Web.

In earlier publications, I have used the term “knowledge integration” to denote the process of bringing together scattered statements about like subject matter, which is represented as topic map fragments. However, I always assumed this integration only to be true for the integrator himself (or herself), as I had a P2P scenario in mind. Others can optionally import this view as additional layer, if they trust, but are not forced to. I also used the term “knowledge aggregation” to underline the aspect of bringing together like jigsaw puzzle parts, but possibly recombining the parts to new blocks. This presupposed that the aggregation uses well-defined relations between fragments, and does not throw every item on piles.

But how can one aggregate statements (assertions) about the same subject from different resources? More complex statements can be created by aggregating existing statements, resulting in a semantic network which is an index adhering to a sophisticated index language with a grammar. Aggregation is defined as at least two statements sharing one argument. A first sketch of how this might work has been given in [8].

I appreciate the term “knowledge federation”, because it does better express the autonomous multi-actor setting, the “democratic” aspects and the focus on social processes, and now prefer it over the term knowledge integration. However, I would not say that knowledge integration always implies “knowledge dictatorship”, because integrating something into my view does not presuppose that I will necessarily prescribe everyone my result of the integration of knowledge fragments. It is just an offer, or even a valuable intangible good.

In my view, KO can well comprise processes of knowledge federation, and KO can also be democratic by supporting multiple viewpoints.

4 Can Knowledge Organization support multiple conflicting viewpoints?

Determining the relevance of an information resource to a concept, and thus determining its subject, is a hermeneutic act somehow relative to the interpreter. Different communities and usage contexts entail validly differing viewpoints, leading to differing conceptualizations. However, traditional knowledge organization and traditional knowledge management ultimately had to force different viewpoints into one, centralized homogeneous conceptual system (e.g. a taxonomy). The internet increased conceptual heterogeneity and brought us a more democratic, user-contributed subject organization, lately e.g. in the form of folksonomies and tagsonomies (cf. the tagsocratic project).

It is not the function of KO to exert “dictatorship” by imposing a central, dominating view on everyone on how knowledge is to be organized. Its function is to provide
means for easier conceptual access by communities. Up to recently, KO had to find a compromise because it was not possible to have more than one view. Traditional KO was rather oriented towards an ideal of finding the “right place” for each concept, if not some kind of “ultimate truth”:

“Therefore, institutes of knowledge organization ought to be established where work on the definition and systematization (establishing the correct system positions of each concept) in a given field of knowledge can take place.” [7, p. 85]

This contrasts with more open epistemological positions. “Democratic indexing”, a reconciliation process, where users collaboratively chose the terms used in indexing, has been proposed by Hidderley & Rafferty in 1997 [10, 11]. It is well compatible with user- and domain-oriented approaches of KO: “Allowing conflicting multiple conceptualizations introduces indexing views, or view-based indexing. This leads to the idea, adapted from user-oriented indexing, that the different viewpoints of user groups must be modeled in an ontology-based relevance model, and the indexing must consist of qualitative argumentations (relevance reasons) why a certain document is relevant from this special viewpoint.” [5, p. 398]. I suggested “any KO approach to provide ways to express and handle multiple, typically conflicting conceptualizations. (...) If we take seriously the idea that different discourse communities conceptualize differently, the goal is no longer to find a universal classification as a compromise, but to maximize the utility for both by separating both contexts and maximizing locally to the needs of each community. We can have two different views about the aboutness of a document, and we thus can conceptually adapt the subject representation of a document to the user communities.” [5, p. 408]

This KO theory has still to find its way into practical systems, but knowledge federation would benefit from it.

5 Summary and Conclusion

I have argued that KO shares with knowledge federation the common objective of facilitating insight by supporting gradual subject-centric organization of knowledge in an environment which is conceptually heterogeneous and where the federation of scattered knowledge fragments is impeded by obstacles on several layers. I have shown that KO can offer several approaches, some of them e.g. user-oriented, and have introduced KOS consisting of concepts and relations. I have presented comments on important issues in KO (arguably the most important being semantic interoperability and multiple viewpoints), and have sketched how KO combines with and differs from knowledge integration and knowledge federation. I have shown that modern KO can cope with multiple viewpoints, leading to view-based, “democratic” indexing.

Because with biased KO eyes I see the main challenges in knowledge federation in coping with semantic heterogeneity and supporting multiple viewpoints and their
emergent reconciliation over time, I am convinced that KO has much to offer – but not the ready-made patented solution. More dialogue and research is necessary.

References