

Richard P. Smiraglia—University of Wisconsin, Milwaukee

## **Modulation and Specialization in North American Knowledge Organization: Visualizing Pioneers**

### **Abstract**

Pioneers are those who, in some way, lead their peers to new destinations. In the evolution of a domain, the pioneers might very well be those who have followed a theoretical principle in some particularly ardent manner, thus leading the rest of the domain toward an evolving research front. The present paper is an attempt to use the tools of domain analysis to diachronically analyze the domain of knowledge organization as it is evolving in North America. That is we use bibliometric tools to identify the axes that define North American knowledge organization and its scientists, who are its pioneers. The evolution of a North American chapter of the International Society for Knowledge Organization (ISKO) marks a growth in coherence of a long active research area. An interesting research question is: what are the characteristics of North American scholarship in knowledge organization? Author co-citation analysis of North American authors whose work appeared in the journal *Knowledge Organization* is contrasted with author co-citation analysis of authors from outside North America. North American leaders are clearly identified, and some themes—such as knowledge organization online—that are emergent topics in North America are identified.

### **1. A Word About Pioneers**

Pioneers are those who, in some way, lead their peers to new destinations. Sometimes pioneers point the way or clear the path, sometimes they literally lead the way, and sometimes the term “pioneer” is used to designate role models. In the evolution of a domain, the pioneers might very well be those who have followed a theoretical principle in some particularly ardent manner, thus leading the rest of the domain toward an evolving research front. The call for papers for the 2009 North American Symposium on Knowledge Organization (NASKO 2009) set out a theme of “Pioneering North American Contributions,” and asked particularly for papers that follow Rayward (2004, 676) in presenting “evidence-based narratives that are not primarily celebratory, but that offer an opportunity for ‘detailed critical assessment of matters of importance.’” The present paper is an attempt to use the tools of domain analysis to analyze the domain of knowledge organization as it is evolving in North America. That is, we will use our own tools—particularly bibliometric tools—to identify the axes that define North American knowledge organization and its scientists. In this case, we are treating the members of the domain as its pioneers. That is, by casting a metaphorical net in two directions we hope to illuminate the pioneering activities of those whose work is defining and delimiting the domain as it evolves.

The twentieth century saw an evolution in library-and-information science (LIS) from a very pragmatic beginning, rooted in practice, through major developments in bibliographic classification, to a position of rationalism influencing the development of models of information and information behavior near the turn of the twenty-first century. A parallel evolution (or perhaps it was embedded) moved us from the equation of bibliographic classification with knowledge organization to the current perception of Knowledge Organization Systems (KOS) as domain-centered and work-based. These trends have been

described by Smiraglia (2002a, 2002b), Hjørland (1998), and Hjørland and Albrechtsen (1999), among others. Mai (1999) suggests this evolution is part of a larger trend toward post-modern thought, in which the search for universal truth has given way to the pragmatic attempt to document our immediate surroundings.

Dahlberg (2006) suggests the field of knowledge organization at the beginning of the twenty-first century is based on a propositional concept of science, operating as a subfield of science-of-science, embracing primarily concepts, methods of indexing, universal systems, subject-based systems, and taxonomies. Hjørland (2003) sketches the boundaries of a knowledge organization that is much broader than LIS, and he makes a distinction between social and intellectual forms of knowledge organization, and stipulates the basic unit of knowledge organization is the semantic relation between concepts. In a 2008 special issue of *Knowledge Organization* (vol. 35 nos. 2/3), authors offered several perspectives on the parameters of the domain, including:

- Hjørland (p. 86): the social division of mental labor;
- Tennis (pp. 109-10): epistemology, theory, methodology, design, study, critique, identity;
- López-Huertas (pp. 118 ff.): multi-lingual, cultural, social, and ethical issues.

This list is partial, but it suggests some of the extensions of concept theory that define the domain at large, especially as viewed within LIS.

## **2. Knowledge Organization in North America**

The evolution of a formalized North American chapter of the International Society for Knowledge Organization (ISKO) marks a growth in coherence of a long active research area. What is not so clear is whether there is a particularly regional approach to the field. ISKO has active regional chapters in China, France, Spain, Italy, Germany-Austria-Switzerland, Russia, and Poland, among others. The advent of a North American group raises an interesting question about the approaches taken by North American scholars in this discipline. An interesting basic research question is: what are the characteristics of North American scholarship in knowledge organization? In this essentially preliminary study, the tools of domain analysis will be applied to the literature of knowledge organization by North American contributors to visualize thematic axes that represent the “pioneers” of an evolving domain.

There are several techniques used to visualize activity in domains, not the least of which is the production of a review article. But bibliometric techniques, including citation analysis, word and co-word analysis, and author co-citation analysis are techniques that can reveal the contours of progress over time, and identify core theoretical paradigms. White and McCain (1997) describe these techniques for visualization, which is seen as especially helpful for identification of evolving or emerging domains. Similarly, Hjørland (2002) suggests eleven approaches to understanding the intellectual content of a domain, of which bibliometric studies and empirical analyses focused on terminology are key. Identifying the

axes of the domain will point to its operational areas. Tennis (2003) describes two such axes: 1) areas of modulation—definitions that modulate the domain by stating its extension; and, 2) degrees of specialization, including “focus” and “intersection”—ways to qualify a domain by increasing its intension.

There is very little prior research to draw upon. Content analysis of the 2006 ISKO International conference program (Smiraglia 2006) suggested four axes of modulation were: ontology and classification, data modeling, cultural integration, and knowledge management. The 2008 International Conference in Montréal had more scattered axes of modulation (Smiraglia 2008), corresponding roughly to information retrieval, epistemology, warrants (cultures, languages, environments), and models. The 2008 conference was more highly dependent than might have been expected on monographs (234 monograph citations to 381 journal citations and 100 proceedings citations) suggesting humanistic influences remain important, particularly in the classical approaches to ontology. A recent dissertation (Friedman 2007) included national origin figures for papers submitted both to American Society for Information Science and Technology Special Interest Group/Classification Research (ASIST SIG/CR) conferences and to biennial ISKO International Conferences; 61% of the SIG/CR papers originated in the U.S. as did 23% of ISKO conference papers, giving us a relative sense of the proportion of work in knowledge organization that is taking place in the US, if not in all of North America. All of these indications must be considered preliminary and limited as points of departure for the present project.

## 2.1 NASKO 2007

The only prior indicator of North American research comes from analysis of the proceedings of NASKO, the first North American symposium, which was held in 2007. This symposium was quite small (13 papers and 3 posters were accepted), but three noticeable axes emerged from co-word analysis of the abstracts (Smiraglia 2007): faceted classification, social classification (tagging), and applications. Most papers cited recent journal articles, and 2/3 of the citations were unique, indicating a minimal sense of cohesion among the authors. Semantics and domain analysis seem to form critical thematic axes among the works cited by NASKO authors, as did facets and controlled vocabulary, information retrieval, international standards, and Web 2.0. Broughton, Vickery, Pollitt, and Hjørland were cited among the top 19, but the rest were North Americans (See Table 1).

**Table 1.** North American authors most frequently cited in NASKO 2007

Atherton-Cochrane	6
Olson	6
Smiraglia	6
Beghtol	5
Abbas	4
Drabenstott-Markey	4
Bates	3
Binns	3
Campbell	3

Hearst	3
Howarth	3
Kipp	3
Mai	3
Mann	3

To reiterate, we have only very preliminary indicators here about whether there might be a particularly North American approach to knowledge organization. We know that approximately a quarter of papers at ISKO international conferences come from North America, and that although North American researchers are contributing substantively to the domain at large, European authors are among those most heavily cited by North Americans.

**Table 2.** Some Thematic Axes

ISKO 2006	ISKO 2008	NASKO 2007
ontology and classification	information retrieval	faceted classification
data modeling	epistemology	social classification (tagging)
cultural integration	warrants (cultures, languages, environments)	applications
knowledge management	models	

We can see in Table 2 the thematic similarity between NASKO 2007 and the international conferences from 2006 and 2008. If there is a difference in emphasis it lies in a focus on applications in this specific instance.

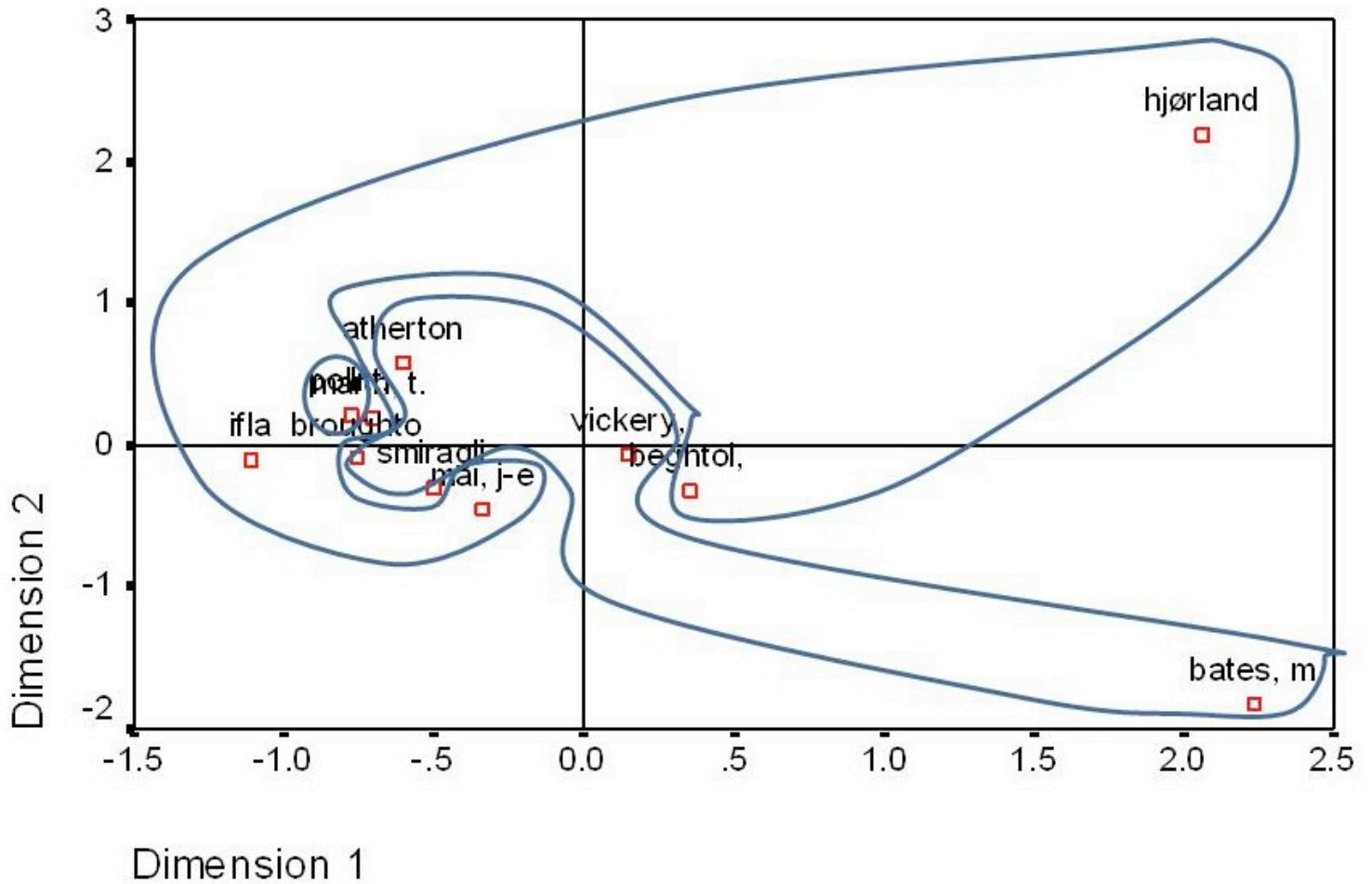
### 3.0 North American Contributions to the Journal *Knowledge Organization*

Several approaches to visualization of North American scholarship in knowledge organization are possible. Contributions to all ISKO conferences could be compiled, as could contributions to the ASIST SIG/CR, or thematic searching of the journal literature of information studies and allied disciplines could be used to yield a list of active scholars. For the present study, the North American contributions to the ISKO quarterly journal *Knowledge Organization* were used. This is considered just one possible approach to visualizing the domain, but it has value in rooting these preliminary indicators in the one formal domain publication venue that is indexed electronically.

Using ISI's *Web of Science*, all articles published in the journal *Knowledge Organization* (1993-2009) were located. In 1993 the journal changed its title from *International Classification*, and this change was considered to be epistemologically significant for the present study. Furthermore, the past three editors of the journal (including the present editor) have been North American scholars. In each article, author institutional affiliations were examined, and only those associated with institutions in North America were selected. A total of 336 authors were represented, of which 75 were from North America (24 from Canada, and 51 from the United States). A list of authors was compiled and then merged with the list of leading cited-authors from NASKO 2007. Author

co-citation figures were downloaded from the *Web of Science*, and authors with no co-citation were removed from the data matrix. *SPSS* then was used to create the multidimensional-scaled plot shown in Figure 1 (stress = .03540 and  $R^2 = .99680$  indicating the plot fits the data well).

**Figure 1.** Author Co-citation in North American Knowledge Organization



Bearing in mind that author co-citation illustrates the way in which a domain is being seen by those who cite these authors, this plot shows the perception that North American knowledge organization is anchored by two classic authors from information science—Bates and Hjørland, grounded by two classic North American authors—Atherton and Beghtol, integrated well with the British approach represented by Vickery, Broughton, and Pollitt, and moving along a more recent axis represented by Mai, Smiraglia, and IFLA's *FRBR* report. The Dendrogram that accompanies the *SPSS*-output (which is not shown here) suggests two clusters, as shown, with the upper cluster wrapped around the lower. In the diagram the names Pollitt overwrites the name Mann. Mann is not connected to the two clusters, but rather forms his own axis. In the summer of 2007, when the NASKO symposium was held, there was much attention to the Library of Congress' reports on the future of cataloging (Library of Congress 2007), to which Mann responded. Likely his presence in this map is an historical artifact (meaning the co-citation is likely an accident of timing) rather than an independent axis in North American knowledge organization. Reference to the individual works cited reveals that the other two clusters represent information retrieval (Bates, Vickery, Atherton) and classification (Hjørland, Beghtol, and Broughton). But we also see the perception of shared epistemologies—empiricism in the lower cluster and rationalism in the upper. By making reference to the analysis of NASKO 2007 reported above, we also can see that the primary axes are semantics and information retrieval—two aspects of information science that are essentially interconnected in classification and domain analysis.

To understand more fully what might constitute the foundations of the domain the citations in papers by our co-cited authors were analyzed. These authors cited 225 sources, among which the most cited were two works by Ranganathan (*Prolegomena* and *Elements of Library Classification*), and two papers by Albrechtsen and Jacob (1989. Dynamics of classification systems as boundaries) and Hjørland (2002. Domain analysis in information science). The subjects are traditional classification, and domain analysis, which accords with our earlier observation that these are the axes of modulation in North American KO.

This analysis is clearly preliminary, though it is sufficient to answer the research question about the characteristics of North American research in knowledge organization with broad strokes. We see an active research community with significant author productivity making steady progress in a variety of traditional and emerging areas. If we agree that pioneers lead by pointing the way and providing role models, we can see leadership in these results. The papers clearly fit within the parameters of the emergent domain of post-modern knowledge organization as described in the opening paragraph above: conceptual, largely broader than LIS, and domain-centered.

#### **4.0 Knowledge Organization outside North America**

To comprehend better the role of North American pioneers, we sought to create a contrasting portrait of knowledge organization by analyzing contributions from authors outside North America. An essential question is whether there is any difference between the

North American pursuit of knowledge organization and the approach to the field internationally. Again there is little prior research. López-Huertas and Contreras analyzed Spanish research in KO in 2004. They found increasing productivity but low internal coherence, with activity directed toward documentary languages, and thesauri, especially multi-cultural or multi-lingual thesauri. The two analyses of international conferences cited above indicated the presence of axes such as ontology, epistemology, and warrants (cultures, languages, environments) in addition to those revealed in the author co-citation analysis above.

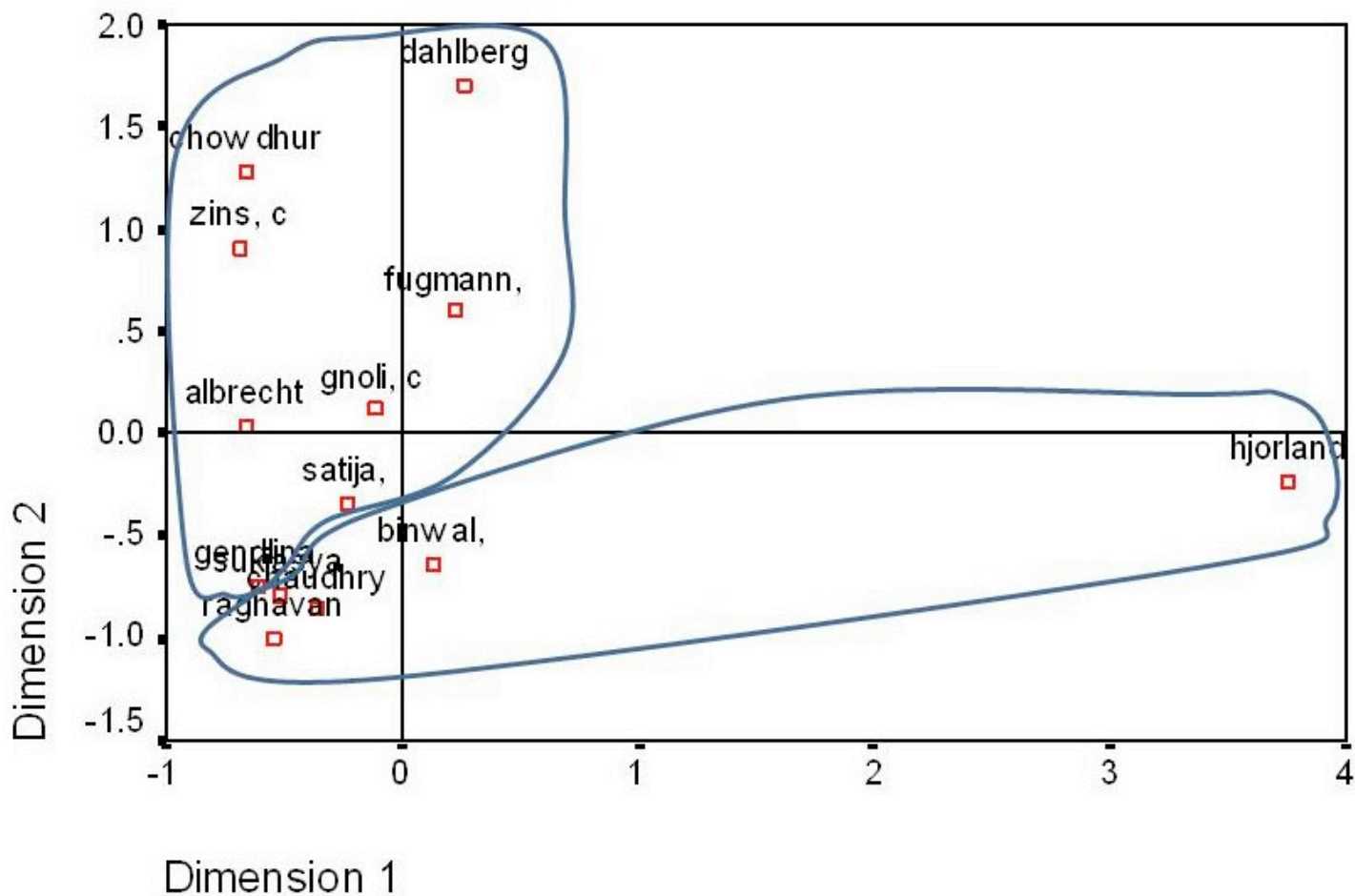
Once again *Web of Science* was consulted and all papers from authors outside North America were recorded together with their cited references. Papers from 31 countries ranging from Belgium to Uruguay were included, although most papers came from 11 countries, shown in Table 3.

**Table 3.** National Origin of Papers from Outside North America

France	34
Russia	34
Germany	26
England	19
India	19
Denmark	18
Italy	12
Singapore	11
Israel	9
Netherlands	8

A list of the 13 most prolific authors emerged, which included Albrechtsen, Binwal, Chaudhry, Chowdhury, Dahlberg, Fugmann, Gendlina, Gnoli, Hjørland, Raghavan, Satija, Sukiasyan, and Zins. ISI was used to compile an author co-citation plot for these authors, which appears in Figure 2 below (stress = .03250 and  $R^2 = .99707$  indicating again a good fit). As before, the Dendrogram provided with the *SPSS* output (but not reproduced here) reveals the shape and contents of two clusters. Once again Hjørland anchors the plot at some distance from the rest of the research front. Dahlberg also occupies an essential position as anchor to the more tightly clustered group at the left. Bearing in mind that this represents the perception of these authors by those who cite them, we can see two clusters anchored by Hjørland and Dahlberg.

Figure 2. Author Co-Citation in Knowledge Organization from Outside North America



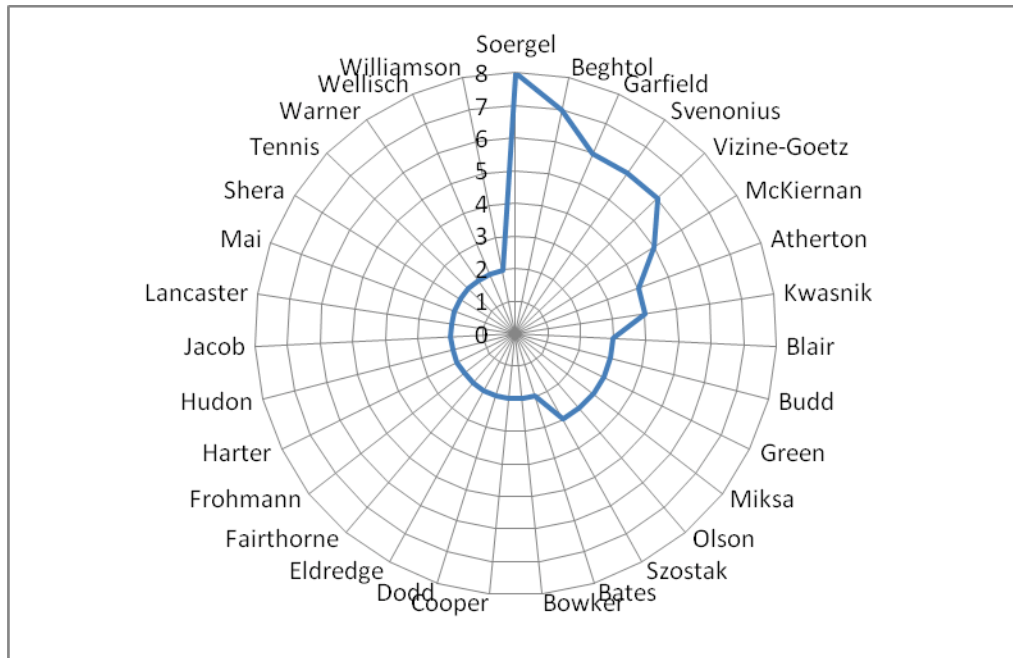
Although there are several anomalies in this map (for instance, that Albrechtsen and Hjørland, who often write together, are perceived to be in different clusters), the papers in the lower cluster seem to revolve around an information retrieval axis, while those in the cluster above revolve around a concept-based axis including classification, indexing, and domain analysis. Although there is a subtle terminological difference, in essence the essential axes are the same as those we observed in the North American analysis above.

To answer the question of how influential North American authors have been on their counterparts elsewhere in the world, cited references were analyzed. Often, documents in



classification research come from small local conferences, workshops or symposia and therefore, are of limited availability for citation. Interestingly, Russian and Indian authors tended to cite almost exclusively authors with their own nationality. But among the other authors the influence of North American scholarship was clear. Thirty-one North American authors are cited; those most frequently cited are named in Figure 3.

**Figure 3.** North American Authors Cited by Non-North American Peers



Soergel and Beghtol are the most cited, followed closely by Garfield, Svenonius, Vizine-Goetz and so forth, moving clockwise around the circle. Books frequently cited include Soergel's *Organizing Information* and Svenonius' *The Intellectual Foundation of Information Organization*, but also Atherton's *Classification Research* proceedings from 1965. Frequently cited papers include Kwasnik's (1999) "The role of classification in knowledge representation and discovery," Soergel's (1999) "The rise of ontologies," Beghtol's (1995) "Facets as interdisciplinary undiscovered knowledge," (1998) "Knowledge domains," and (2002) "A proposed ethical warrant for global knowledge representation and organization systems." Also popular are conference papers by Vizine-Goetz (e.g. 2001) concerning the use of *DDC* for organizing internet resources, and McKiernan's papers (e.g. 1998) about bookmarks—the beginning of research on what would become known as social classification.

## 5.0 Visualizing Pioneers

The analyses presented here are preliminary, of course, but also they are indicative. More research incorporating the conference papers of ISKO biennial conferences as well as those from the ASIST/SIG-CR workshops would yield a fuller picture of the domain (although, as yet those data are not in electronic form). Still, what we can see here is indicative of an evolving domain in knowledge organization, and of a strong regional component in North America. Tennis (2003) suggested a domain could be identified by the description of its total scope, which he calls its area of modulation, and by its degrees of specialization, which qualify the former. With only occasional and subtle differences in terminology we see that knowledge organization as the area of modulation is qualified primarily by axes that can be identified with information retrieval and semantics, or concept-theory. The latter axis is characterized by classification and domain analysis. A subtle epistemological distinction seems also to characterize the two axes, with empiricism associated with the information retrieval axis and rationalism with the semantic axis. Knowledge organization online, including social classification, is an emergent topic in the North American domain; multi-cultural and multi-lingual aspects are more apparent outside North America. Again, this analysis is preliminary and indicative, and is not definitive.

That said, North American pioneers do emerge from this analysis. At the very least, scholars like Atherton, Beghtol, Garfield, Svenonius, Soergel have led the way for the entire domain. Citation of their seminal works as well as their most recent contributions demonstrates the importance of their contributions. These and other scholars of knowledge organization in North America routinely develop theoretical principles that inform the domain globally, by serving as international role models, and bringing new ideas forward at a regular pace. Perhaps for those who populate the domain it is too soon to posit much more than this, which brings us back to the point with which we began—the members of this domain are, and continue to be, its pioneers.

## References

- Atherton, Pauline, ed. 1965. *Classification research: proceedings of the second International Study Conference held at Hotel Prins Hamlet, Elsinore, Denmark, 14th to 18th September, 1964*. Copenhagen: Munksgaard.
- Beghtol, Clare. 1995. 'Facets' as interdisciplinary undiscovered public knowledge. S.R. Ranganathan in India and L. Guttman in Israel. *Journal of documentation* 51: 194-224.
- Beghtol, Clare. 1998. Knowledge domains: multidisciplinary and bibliographic classification systems. *Knowledge organization* 25: 1-12.
- Beghtol, Clare. 2002. A proposed ethical warrant for global knowledge representation and organization systems. *Journal of documentation* 58: 507-32.
- Dahlberg, Ingetraut. 2006. Knowledge organization: a new science? *Knowledge organization* 33: 11-19.
- Friedman, Alon. 2007. Concept map as "sign;" concept mapping in knowledge organization through a semiotics lens. Ph.D. dissertation, Long Island University.

- Hjørland, Birger. 2008. What is knowledge organization (KO)? *Knowledge organization* 35: 86-101.
- Hjørland, Birger. 1998. Theory and meta-theory of information science: a new interpretation. *Journal of documentation* 54: 606-21.
- Hjørland, Birger. 2003. Fundamentals of knowledge organization. *Knowledge organization* 30: 87-111.
- Hjørland, Birger and Hanne Albrechtsen. (1999). An analysis of some trends in classification research. *Knowledge organization* 26: 131-9.
- Kwasnik, Barbara. 1999. The role of classification in knowledge representation and discovery. *Library trends* 48n1: 22-47.
- Library of Congress, Working Group on the Future of Bibliographic Control. 2007. Draft report. <http://www.loc.gov/today/pr/2007/07-244.html> (accessed 8 June 2009).
- López-Huertas, María J. 2008. Some current research questions in the field of knowledge organization. *Knowledge organization* 35: 113-36.
- López-Huertas, María J. and Evaristo Jiménez Contreras. 2004. Spanish research in knowledge organization 1999-2001. *Knowledge organization* 31: 136-50.
- Mai, Jens-Erik. 1999. A postmodern theory of knowledge organization. *Proceedings of the 62nd annual meeting of the American Society for Information Science*, ed. Larry Woods. Medford, NJ: Information Today, 1999, pp. 547-56.
- McKiernan, Gerry. 1998. Beyond bookmarks: a review of frameworks, features, and functionalities of schemes for organizing the Web. *Internet reference services quarterly* 3n1: 69-82.
- North American Symposium on Knowledge Organization. 2009. Call for papers. <https://netfiles.uiuc.edu/klabarre/shared/iskona/callforpaper.html> (accessed 8 June 2009).
- Rayward, W. Boyd. 2004. When and why is a pioneer: history and heritage in library and information science. *Library trends* 52n4: 671-82.
- Smiraglia, Richard P. 2002a. The progress of theory in knowledge organization. *Library trends* 50: 300-49.
- Smiraglia, Richard P. 2002b. Further progress in theory in knowledge organization. *Canadian journal of information and library science*. 26 n2/3: 30-49.
- Smiraglia, Richard P. 2006. Whither knowledge organization: an editorial. *Knowledge organization* 33: 8-10.
- Smiraglia, Richard P. 2007. A glimpse at knowledge organization in North America: an editorial. *Knowledge organization* 34: 69-71.
- Smiraglia, Richard P. 2008. ISKO 10's Bookshelf: an editorial. *Knowledge organization* 35: 187-91.
- Soergel, Dagobert. 1985. *Organizing information: principles of data base and retrieval systems*. Academic Press Library And Information Science Series. San Diego: Academic Press.
- Soergel, Dagobert. 1999. The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science* 50: 1119-20.

- Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Digital libraries and electronic publishing. Cambridge, Mass.: MIT Press.
- Tennis, Joseph T. 2003. Two axes of domains for domain analysis. *Knowledge organization* 30: 191-5.
- Tennis, Joseph T. 2008. Epistemology, theory, and methodology in knowledge organization: toward a classification, metatheory, and research framework. *Knowledge organization* 35: 102-12.
- Vizine-Goetz, Diane. Dewey in CORC: classification in metadata and pathfinders. *Journal of internet cataloging* 4n1/2: 67-80.
- White, Howard D. and Katherine W. McCain. 1997. Visualization of literatures. In *Annual review of information science and technology* 32, Medford N.J.: Information Today, pp. 99-168.