Replication and Accumulation in Knowledge Organization— An Editorial

Richard P. Smiraglia

Smiraglia, Richard P. 2017. "Replication and Accumulation in Knowledge Organization—An Editorial." *Knowledge Organization* 44(5): 315-317. 15 references.

1.0 Knowledge organization is science

The Economist is a good place for the well-informed layperson to get news about scientific discoveries. In the issue from October 8, 2016 there is a story about a phenomenon called "autophagy," which is a system of recycling cellular components. The story is about a 2016 Nobel prize winner for physiology or medicine, Yoshimori Ohsumi of the Tokyo Institute of Technology, whose work nailed down details of the process. The story tells us the research began in 1988 with a type of yeast, and progressed through several iterations in each of which a different gene was disabled until it became clear how the process worked, at which point it could be scaled up to human research. This is the typical progress of research in almost any science. Experiments are conducted in a stepwise fashion, sometimes changing one variable in each new experiment while controlling the others, at other times sheer replication is used either to confirm a finding, or to provide sufficient justification for scaling up or moving external parameters. This is how science leads to theory, and it is how theory leads to laws, and it is laws that dictate the core of any coherent domain.

It is important for research in knowledge organization to embrace replication and step-wise experimentation on a much larger scale than has been undertaken to date

Knowledge organization is the science of the order of knowledge, based on the central unit of the concept. It is understood that concepts derive their meaning from domain-dependency. But if this is the case, then it is important for research in knowledge organization to embrace replication and step-wise experimentation on a much larger scale than has been undertaken to date.

Nagel (1979) was writing at the pinnacle of empirical methodology in modern science (needless to add, perhaps, at the University of Chicago, the institute at the pinnacle of empirical methodology for most of the twentieth century) and even he felt it necessary to defend "science" over against "common sense." He wanted readers to understand that (3) "sciences are organized bodies of knowledge and … in all of them a classification of their materials into significant types or kinds … is an indispensable task;" and (4) "it is the desire for explanations which are at once systematic and controllable by factual evidence that generates science; and it is the organization and classification of knowledge on the basis of explanatory principles that is the distinctive goal of the sciences."

It is the desire for explanations which are at once systematic and controllable by factual evidence that generates science

It is the organization and classification of knowledge on the basis of explanatory principles that is the distinctive goal of the sciences

He did so by explaining that science is the essential task of isolating variables so as to generate reliable explanations. But, more to the point of this essay, he wrote (32): "A deductive scientific explanation, whose explicandum is the occurrence of some event or the possession of some property by a given object, must satisfy two logical conditions. The premises must contain at least one universal law, whose inclusion in the premises is essential for the deduction of the explicandum. And the premises must also contain a suitable number of initial conditions." He means, in colloquial terms, a deductive expla-

nation of an observable must contain two parts. There must be an appeal to general occurrences of the sort under observation. And, there must be a sufficient number of determinative variables. If you hope to explain why water is falling on your head in your front yard, you must know both about weather and jokers with garden hoses, and you must know that the water is coming either from a cloud or from a garden hose. Science means reducing all of the knowables and unknowables to classified positions. This is the role of knowledge organization in the advancement of humanity. It is not enough to distinguish pragmatically between rain from above and rain from the sky; rather, it is critical to explain the difference between the sky and mere aboveness and to classify different instantiations of the same event (water on your head) appropriately.

2.0 Science is cumulative and that requires replication

There has been some empirical validation provided by replication in knowledge organization, but not much. In two papers in 2002 (a, b), now fifteen years ago, I outlined some key areas of empirical research that seemed to have provided demonstrable evidence of what Nagel called "a universal" (if not quite yet laws). These had to do with the recurrence of power law distributions in author productivity, instantiation as a phenomenon of information objects, and external validity for library collections. Authors writing about folksonomy in this journal's pages have confirmed those distributions (Munk and Mørk 2007a and b). In 2007 Marija Petek reported her deliberate and accurate replication of my research into derivative bibliographic relationships, this time in a Sloveninan online catalog. Her results confirmed most of the results of my earlier work, and also pointed to a few points of departure; it was a brilliant example of both replication and accumulation. Her research replicated and therefore demonstrated the theoretical reliability of the earlier work. But her work also pushed the envelope to make the science of instantiation cumulative. More recently a number of authors have taken up the cudgel of replication in domain analysis to expand the domain (Smiraglia 2015; López-Huertas and Smiraglia 2016). Most authors still cite papers from the mid-1990s when they write about domain analysis, ignoring much of two decades of research that has, as Nagel suggested, contributed to the statement of universal explanations. If two decades of explanations have been published in our domain, then authors writing today should demonstrate the accumulative nature of our science by citing the latest work, and not work from decades ago that has been supplanted. For example, in Smiraglia (2015, 32) we learn that in two decades only eight domains have been analyzed twice and only five have been analyzed three times; and none of these have yet been submitted to meta-analysis, which would bring forward theoretical positions concerning the conceptual content of those domains.

Another rich area ripe for empirical analysis is facet analytical theory (CRG [1955] 2007; Broughton 2006), which is having an anniversary of sorts in our field even as it has been taken over by outsiders (who misuse it) to mean smoother website design. Why does this happen? Because scholars in KO fail both to meta-analyze empirical research, and to promote their empirical findings outside the KO domain. Such is the state of research that our most cited, most prolific and most influential mentor has written about the "Paradox of Atheoretical Classification," by which he means classifications that defy Nagel by avoiding at all costs empirical evidence (Hjørland 2016). Even contributors to our journal, such as Zhao and Wei in this issue have written (328): "Besides, the knowledge of knowledge organization in China is rapidly accumulative."

3.0 Knowledge organization must be more cumulative, and there must be more replication

Knowledge organization, like any science, must rely on empirical evidence and that means experimentation must rely on replication. Our founder's words speak eloquently to the point. For example, in a 2008 interview Dahlberg said (85): "The widespread acceptance of KO … implies that it be moved out of its present place under 'classification and indexing' in LIS establishments, where it is even sometimes neglected …." A year later she wrote in our journal (169 ff.): "The noematic concept of knowledge is best fitted for our programme of concept work in knowledge organization, i.e., the known as result of an act of cognition" and "each true statement about a certain item of reference delivers a knowledge element about this together with a characteristic of its concept. The sum of necessary statements about such an item of reference forms the whole of characteristics of its concept." This implies, of course, that the "sum of necessary statements" about any concept must be known, classified, and sorted, scientifically. This is possible only with sufficient replication. In 2014 Dahlberg wrote (86): "Generally accepted knowledge" carries the seal of science, resulting from verifiable dicta or else from intersubjective agreement in form of generally accepted definitions as opposed to subjective knowledge acquired by experience or learning. In the latter meaning, knowledge serves as a kind of spiritual warrant, which means that reminiscence depends on remembered data, which fact explains why people differ in opinion on identical phenomena, for

each relies on different angles of vision and items of recollection." This is exactly Nagel's point—that science relies not on spiritual warrant, not even on common human sense, but rather on replicated and replicable and therefore empirically verifiable controlled observation, the results of which are classified.

I have served on many program committees for knowledge organization conferences where a common response from reviewers is a denigrating: "oh dear, another X treated in Y plan in KO." But this, of course, is exactly the course of science. KO as a domain should not be denigrating replication, but rather, should be encouraging it. Instead of "oh dear," referees in this case should do their job, search the literature, return a list of prior work and require authors to relate proposed research to the published, and peer-acknowledged, science. This is the only way a science can grow.

References

Broughton, Vanda. 2006. "The Need for a Faceted Classification as the Basis of all Methods of Information Retrieval." *Aslib Proceedings: New Information Perspectives* 58 nos. 1/2": 49-72.

Classification Research Group. (1955) 1997. "The Need for a Faceted Classification as the Basis of all Methods of Information Retrieval: Memorandum of the Classification Research Group: Received, for Information, by the L. A. Library Research Committee, May 1955." In *From Classification to "Knowledge Organization: Dorking revisited or Past is Prelude,"* ed. A. Gilchrist. The Hague: FID, 1-9.

Dahlberg, Ingetraut. 2008. "Interview with Ingetraut Dahlberg December 2007." *Knowledge Organization* 35: 82-85.

Dahlberg, Ingetraut. 2009. "Brief Communication: Concepts and Terms—ISKO's Major Challenge." *Knowledge Organization* 36: 169-77.

Dahlberg, Ingetraut. 2014. "Brief Communication: What is Knowledge Organization?" *Knowledge Organization* 41: 85-91.

Hjørland, Birger. 2016. "The Paradox of Atheoretical Classification." *Knowledge Organization* 43: 313-23.

"Understanding Autophagy." 2016. *Economist* 8 October, 70.

L**ó**pez-Huertas, Maria J. 2015. *Special Issue: Domain Analysis Revisited. Knowledge Organization* 42.

Munk, Timme Bisgard and Kristian Mørk. 2007a. "Folksonomy, The Power Law & The Significance of The Least Effort." *Knowledge Organization* 34: 16-33.

Munk, Timme Bisgard and Kristian Mørk. 2007b. "Folksonomies, Tagging Communities and Tagging Strategies: An Empirical Study." *Knowledge Organization* 34: 115-27.

Nagel, Ernest. 1979. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Indianapolis: Hackett.

Petek, Marija. 2007. "Derivative Bibliographic Relationships in the Slovenian Online Catalogue COBIB." *Journal of Documentation* 63: 398-423.

Smiraglia, Richard P. 2002a. "Progress Toward Theory in Knowledge Organization." *Library Trends* 50 no. 3: 300-49.

Smiraglia, Richard P. 2002b. "Further Progress Toward Theory in Knowledge Organization". *Canadian Journal of Information and Library Science* 26 no. 2/3: 30–49.

Smiraglia, Richard P. 2015. *Domain Analysis for Knowledge Organization: Tools for Ontology Extraction*.