Further progress toward theory in knowledge organization

Theory is a system of testable explanatory statements derived from research. Theory explains the domains in which we operate, the phenomena found in those domains, and the ways in which they might be affected by manipulation or change. Theory is derived from the deliberate, self-conscious, and controlled observation of phenomena, whether this has taken place in the positivist empirical paradigm or in the qualitative paradigm. In knowledge organization, the generation of theory has moved from an epistemic stance of pragmatism, based on observation of the construction of retrieval tools, to empiricism, based on the results of empirical research. In three areas, research has accumulated to a degree sufficient to approach the positing of preliminary theoretical statements. These are: 1) author productivity and the distribution of name headings; 2) the work phenomenon; and 3) external validity. Research in these areas is examined, leading to the positing of three preliminary theoretical statements. These are: 1) most names in bibliographic databases will occur only once, and a very small number, which can be predicted by Lotka’s Law, will occur many times; 2) the work phenomenon reflects an association with Lotka’s law, such that many works exist in only one instantiation, but a large proportion evolves steadily over time; and 3) there is a beginning of evidence that there are grounds for external validity in the examination of knowledge entities. Further domain-specific analysis in each of these areas is necessary to lead to strengthened predictive capability of theory that might emerge. Other theoretical statements might soon be possible. Such explanations could give us real predictive power for the development of sophisticated systems for the retrieval of knowledge entities.