City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/15029/

Link to published version:

"A point along a line": moving knowledge organization to the next level

David Bawden
Centre for Information Science, City University London

> *Our time is just a point along a line*
> *That runs forever with no end*
> *I never thought that we would come to find*
> *Ourselves upon these rocks again*
> (Al Stewart, *Lord Grenville*)

When any technique or technology has been in use for some while, it tends to reach a plateau stage, after rapid growth, and questions are asked about where it goes next, and what comes after it. To my mind, the formal vocabularies (FVs) of knowledge organization (KO) - taxonomies and thesauri - are at this stage, but I fear that the questions being asked may not be the right ones. [I should immediately say that this is not a criticism of the papers in this issue, which I have not seen at the time of writing.]

A first set of questions, in the early days, were about what these FVs should be like; how should they be constructed and used. These questions were largely answered many years ago, and incorporated into textbooks and standards, although there is always room for new tweaks.

A second set of questions were about how FVs related to other methods, such as categorization and free-term indexing. These questions were also answered satisfactorily decades ago, although oddly they seem to resurface regularly.

A third set of questions relates to how FVs can be used in new digital environments, bearing in mind that their basic forms were devised in a age of print on paper. These questions also have been answered in a general sense, although there is still work to do in the adaption of FVs to specific new formats, as they emerge.

It seems to me, again emphasizing that I am not criticizing the papers of this issue, that those of us interested in FVs, and with a belief in their continuing value, have a tendency to continue to ask the same questions as we have in the past, and - not surprisingly - to find the same answers. This may help ensure that we are using FVs to the best effect in the ways we are used to, but it does nothing to move us on from the plateau stage to a new phase of rapid growth in use in new ways and new environments. There will be a continuing need to ask such questions in particular contexts, as consultancy or short-term applied research, but there is also need to move on.

I see three ways in which we can, and should, be asking deeper questions.

First, although there is already a substantial body of theory underlying concepts and classification in particular - much of it, one suspects, little known to many of those who work with newer forms of KO - there is still a need for studies of the underlying theory. In particular, when we ask whether, and how, FVs can be used in newer environments, we need a better understanding of how much they embody deep principles, and how much they simply a pragmatic response to the contexts of their time.

Second, we need many more domain studies, of the kind consistently advocated by Birger Hjørland. These would encompass both the nature of information and knowledge in the domain, and the ways in which users of that domain make use of, and find, information. This moves the research focus away from the traditional KO concerns, and into epistemology on the one hand, and information behaviour on the other.

Third, there is a need for convincing studies of the value and impact of FVs, and of knowledge organization generally. For example, a question asked from the earliest days was whether FVs or free text were more effective for retrieval; indeed, studies of this question still emerge from time to time. The answer, known for a long while, is that the best solution is to have both available; but that this is more expensive, and potentially complex for users. The complexity issue may be addressed by the domain studies of point two. The economic issue requires studies using the best available methods for assessing the value and impact of information services generally; for example, contingent valuation, vignettes/personas and critical incidents. We may well find that, as in many other areas, standard products will be largely automated, while 'luxury goods' justify expert human intervention; but research is needed as to exactly what this means.

That such a programme of research is urgently needed is evident, from the growing importance of organized knowledge, in fields as disparate as molecular biology (Mayor and Robinson 2014) and communities of fans of films, television programmes, comics and the like (Price 2015). This latter kind of 'public' application is likely to spread much more widely, as personal information management and 'lifelogging' become established. Convincing such communities of the value of the principles underlying taxonomies and thesauri is likely to be a significant for the future as convincing more conventional information providers of their continuing pragmatics and economic benefits. Both are essential, and both should figure in a research programme to move to KO to the next level.

References

Mayor, C. and Robinson, L. (2014). Ontological realism and classification: structures and concepts in the Gene Ontology, *Journal of the Association for Information Science and Technology*, 65(4), 686-697

Price, L. (2015) Fans as gatekeepers: the role of cult media fans in collecting, preserving and sharing fanworks. Paper presented at the European Fan Cultures conference, Rotterdam, November 2015.