
Innovation in Knowledge Organisation: A conference report

Matt Moore and Kelly Tall*

The inaugural Innovation in Knowledge Organisation (IKO) conference was held in Singapore 8-9 June 2015. IKO was created to showcase innovations in knowledge organisation, whether through the application of innovative methods and technologies, or novel combinations of methods and technologies. This article provides an overview of the key presentation and discussion topics for the conference with reflections on the broader themes that emerged across the two days and what these might mean for information professionals.

INTRODUCTION

“Taxonomy Bootcamp [in the US] is a very solid conference in knowledge organisation. It has been running for over 12 years but last year was exciting. There was a step change in the case studies that were presented. The world was moving on and I wanted to bring some of the excitement and innovation to Singapore”. – Patrick Lambe.

The Innovation in Knowledge Organisation (IKO) conference was hosted by Straits Knowledge (a Singapore-based knowledge management consultancy), PebbleRoad (a Singapore-based user experience design consultancy) and Synaptica (a global taxonomy software company). Additional sponsors were IHS (a global industry information company) and PoolParty (a semantic technology platform).

The focus was on knowledge organisation. This is a topic the authors have touched on many times in *Online Currents* so this conference was of immediate relevance to us. One of the authors was delivering a plenary on building competencies in the knowledge organisation and felt there was much of value to share with the *Online Currents* readership.

The nature of the conference was high energy – although 26 presentations were given, most of these were short form and much of the conference was given over to discussion and interaction. This was not a conference to sit back and fall asleep in.

In this article, the authors provide an overview of the material presented and the discussions they had with participants. It begins with plenary presentations and case studies, continues with an overview of discussions within the clinics and ends with some broader reflections on the conference topics and what they mean for information professionals in Australia.

GETTING STARTED

The morning of the first day began with a keynote by Patrick Lambe from Straits Knowledge. Patrick’s keynote had two focus areas. The first half was the domains that he wanted to integrate over the course of the conference – knowledge organisation methods, technologies and standards. The second half was somewhat unusual in that his focus was on sources of failure within knowledge management projects rather than the tales of success that generally make up case studies discussed at conferences. The six sources of failure were related to these (successful) project features: embed these projects in everyday work, focus on what matters, involve the right stakeholders, balance the various elements of these projects, demonstrate and deliver value, and build capability over time. On the one hand, this made for a dark opening but on the other, all the subsequent case studies presented would

* Matt Moore has spent 15 years working in knowledge and information management, learning and development and internal communications with organisations such as IBM, Oracle and the Australian Securities and Investment Commission. He lectures at the University of Technology Sydney and is a former chair of the New South Wales Knowledge Management Forum. Kelly Tall is an experienced market researcher with a passion for data visualisation. She has worked with many well-known global and Australian brands. All websites viewed August 2015.



appear highly successful in contrast. Lambe's work¹ has been as concerned with learning from failure as much as celebrating success for a long time and we share his desire for practitioners to share their experiences honestly with each other. This honesty can be hard to find on conference podiums so it was an important step in setting the tone for the conference.

The rest of the morning was devoted to a series of pecha kucha presentations on knowledge organisation topics. Pecha kucha is, as one speaker noted, a brutal discipline. Each speaker has to present 20 slides each on an automatic timer of 20 seconds per slide. The eight topics covered were:

- Taxonomies, ontologies and knowledge graphs.
- Autoclassification and text analytics.
- Linked data.
- Standards for taxonomies and Linked data.
- No SQL and graph databases.
- Open source search (eg Lucene, Solr, Elasticsearch).
- Search based applications.
- Internet of Things.

We've covered many of these topics in previous articles so we will move on to the application.

CASE STUDIES

The afternoon of the first day and the morning of the second saw a series of case studies. Each case study presenter had the opportunity to present a six minute "pitch" to the whole audience before then presenting it for 20 minutes, three times to different audiences. It is worth outlining the case studies at a high level to get a sense of the current state of the art in knowledge organisation. More information on these can be found on the IKO website.²

Agnes Molnar (Search Explained, Hungary). Molnar had a global software client who already had an enterprise search implementation but wanted a richer search experience. The client had migrated most of their content to Microsoft Office 365 so the project used Office Graph – a machine learning engine that analyses content, people and activities within Office 365 environments and via applications like Delve, and can proactively present information to users in particular scenarios. Apart from ensuring the content was stored properly, the main challenges were in explaining how the system worked to users and also the limited customisation in the current version of Office Graph.

Ahren Lehnert (FMC Technologies, USA). Lehnert works at an oilfield equipment and services company with 18,000 employees around the globe. There was no single search instance or singular reporting on search queries which meant both a fractured user experience and limited insight to improve search results and information management. There was also no common controlled vocabulary for metadata values. A firm-wide taxonomy was created that leveraged the analysis of search terms. The technical solution involved SharePoint 2013 search; Smartlogic Semaphore for taxonomy management, text mining, and auto-categorisation; and SAS JMP Text Miner for search term analysis.

Charlie Hull (Flax Consulting, UK). Reed Specialist Recruitment offers recruitment services across 350 offices worldwide. The existing recruitment search application used an Oracle database: searches took several minutes to complete and returned an un-ranked list of results, which staff would then have to work through by hand. Flax designed a search infrastructure based on the open source search engine Apache Lucene/Solr, a custom Java indexer and text extraction with Apache Tika. The new search engine provided sub-second search response times and ranking capability. Users required significant retraining to take advantage of these – they simply had forgotten to trust search.

¹ Lambe P, "Why Knowledge Managers should be Sued" (actKM Conference 2007) <http://www.actkm.org/userfiles/File/actkm2007conf/Day%201%20Keynote%202007%20-%20Why%20Knowledge%20Managers%20Should%20be%20Sued%20-%20Patrick%20Lambe.pdf>.

² Materials from the conference including presentations, case study materials and videos can be found at <http://www.ikoconference.org/>.



David Clarke (Synaptica, UK). Synaptica have built an image annotating tool – OASIS – that allows users to explore images, tag them and link them to external sources via Linked Open Data.

Elisa Erriquez and Jenni Wardell (BBC, UK). The BBC has been at the forefront of taking a semantic approach to their content. The publication of their 2010 World Cup coverage was based on a semantic model (entities including teams, players and matches) that allowed them to dynamically serve up content to users. The approach has been expanded to BBC Sport and BBC News. External sources such as Wikidata, Dbpedia, MusicBrainz, GeoNames, and Statistics.gov.uk are integrated into the platform using a linked data approach.

Maish Nichani (Pebble Road, Singapore). Pebble Road worked with a Singaporean government agency that runs regular household surveys. They carried out a field study with the surveyors to identify their issues – which were less around a survey app (as might be expected) and more about scheduling interviews. They created an app that sourced location, distance, directions (Google), climate (National Environment Agency), bus, taxi (Land Transport Authority), age of residents (Client organisation) and provided the integrated information for users.

Marjorie Hlava (Access Innovations, USA). Access Innovations used Linked Data and taxonomies to create a quick-start “smart” thesaurus with a large scientific publisher and a large scientific association with a publishing platform – both cases involved hundreds of thousands of articles. Thesaurus terms in articles were identified with the full text material and the thesaurus terms themselves were linked to external sources such as dbpedia and wikidata. In many cases, the specific thesaurus terms were not discussed in these sources so entries had to be created for them. The results of this effort were staff time savings and an improved end user experience.

Nor Azlinayati Abdul Manaf (MIMOS Berhad, Malaysia) presented two case studies. In the first, the Malaysian Agricultural Research and Development Institute (MARDI) had no way of connecting the knowledge of the whole value chain of paddy (wet rice cultivation) research. The PADIPEDIA application was built including a web-based portal with searching capabilities, basic analysis and reporting function using semantic technology. The development of paddy ontology is an attempt to utilise semantic web technology for organising knowledge. The second case study, the Malaysian Science and Technology Information Centre (MASTIC) required a system that could provide insight into Malaysia’s R&D capabilities and activities. KRSTE.my allows users to explore the relationships between research results, researchers and topic areas through structured natural language processing and semantic search.

Eric Tsui (Hong Kong Poly University, Hong Kong). HKPolyU used a mix of taxonomy, folksonomy tags and a clustering algorithm to enhance the navigation functions of a tourism web site.

Haliza Jailani and Kia Siang Hock (National Library Board, Singapore) each presented a case study from Singapore’s National Library. The merger of National Archives Singapore (NAS) and the National Library Board in November 2012 meant that the harmonisation of library and archives metadata, name headings and controlled vocabularies was needed so the public can discover resources without having to go through two portals. The schemas used by the NLB (Dublin Core) and the NAS (ISAD G) were mapped. A quarter of a million records and named entities were converted into RDF and published as linked data. Meanwhile, text mining based on Hadoop was deployed with over 10 million items identifying 1 billion associations. The NLB works in English, Chinese, Malay and Tamil and links between entities in these languages had to be created. Overall use of materials by the public has increased.

Mark Glikson (Gumbuya, Singapore). Gumbuya provides the software for O-Pulse – a portal for investment advisors to manage client relationships, compliance, and communications with portfolio managers. Starting with a Semantic O/S, O-Pulse was able to leverage schematised data structures, defining real-world entities, things and objects semantically without the need for database setup or manipulation. O-Pulse was able to deliver the initial version of their portal with 40% savings in both time and cost, when compared to alternative options using the traditional development process.

Mark Garlinghouse (Hinton Information Services, Singapore). A *Fortune 500* food and beverage manufacturer with a team of about 1,000 globally-distributed R&D professionals restructured their R&D capability and identified that the R&D repository and R&D content assets were woefully



underutilised. IHS created knowledge bases from the client's internal content and linked it to the open source and proprietary content sets that the client already used or selected to license for this project. The solution used IHS Goldfire to semantically index and link the internal and external content and IHS Knowledge Collections as a platform for proprietary content. The quality of literature reviews at the initiation of new research projects has increased and the access to patent literature has enabled researchers to identify IP white space to target in their research.

Patrick Lambe (Straits Knowledge) and Loh Yu Gene (Stronium Sdn Bhd, Malaysia). Straits Knowledge have been conducting knowledge mapping exercises with clients for over a decade using paper and transcriptions of content into Excel – which was highly labour-intensive. Straits Knowledge worked with Stronium, a software development company, to build a knowledge audit tool using a graph database foundation. The graph database provided a flexible foundation so that all requirements did not need to be tightly tied down at the beginning and usability and functionality issues could be solved along the way. Analysis of knowledge audit results needs to be highly flexible, and flexible querying is also a feature of graph databases.

As a participant, some overarching themes emerged from these case studies:

- There are few parts of life that are not being touched by these technologies – the case studies involved public media companies, scientists and engineers, library patrons, recruiters and job hunters, software developers and tourists.
- Sophisticated technologies are entering the mainstream – eg Microsoft Delve, Lucene Solr. These technologies are largely invisible to users in the front end – indeed their purpose is to make information exploration seamless. However they have significant impacts on infrastructure in the back end and those who manage it.
- Case studies came from all over the world. That included local (ie Singaporean/Malaysian) groups who are pushing ahead with new techniques to solve real problems and bring real value to users.
- For those that have been operating in this space for some time, certain practice areas have not expanded as might be expected. Only Eric Tsui's presentation explicitly mentioned folksonomies. Had this event been run in 2008 then user-based, free-text tagging would have been centre-stage. The world has moved on.
- Some things have changed less than expected. We would have been surprised had it been said in 2008 that SharePoint would remain the dominant enterprise content management system. It was under attack from Google Apps and many other innovative technology providers. However, SharePoint was still a major topic of discussion – not only in the case studies, but also in the clinics (see below).

INTERLUDE: THE FUTURE OF SEARCH

Charlie Hull returned for the plenary session of the second day to talk about the role of search in the world of big data – especially the velocity dimension of the volume/velocity/variety view of big data. To put it crudely, he discussed how real-time data feeds could be queried to both identify patterns and also formulate new datasets. The cases used include network monitoring, social media trends, Internet of Things tracking, and customer interaction pattern recognition.

Search is no longer simply about “search” – ie information retrieval triggered by a user entering a string of text. It is moving into the real-time, the predictive, and the visual. This is disturbing as many organisations struggle to implement basic search technologies – recent surveys indicate that only 11% of responding organisations have a cross-organisation enterprise search (AIIM 2014)³ and 45% of users in responding organisations are dissatisfied with information findability (Findwise 2014).⁴

The authors suggest that search and findability are entering a twin-track world where technology providers produce ever more sophisticated tools for specialised knowledge retrieval that are deployed

³ Association for Information and Image Management (AIIM), *Search and Discovery – Exploiting Knowledge, Minimizing Risk* (2014) <http://www.aiim.org/Research-and-Publications/Research/Industry-Watch/Search-and-Discovery-2014>.

⁴ Findwise, *Enterprise Search and Findability Survey* (2014) <http://www2.findwise.com/findabilitysurvey2014>.



by a small number of sophisticated organisations – and the larger number of search deployments continue to disappoint due to failures in implementation. Whether the mainstream can catch up with the vanguard remains to be seen.

GETTING DOWN TO BUSINESS: CLINIC TOPICS

The afternoons of the first and second days were devoted to group discussions. The themes for the group discussions were developed by the participants themselves.

The clinic topics were as follows:

- (1) Getting started – scoping the project for safe fail and learning fast.
- (2) Process for building taxonomies and ontologies.
- (3) Getting buy-in and measuring success – for the organisation and for users.
- (4) Finding the right mix between automation and human effort.
- (5) Maintaining data quality – tags, content, sources.
- (6) Integrating multiple sources (internal and external) and creating a coherent environment – includes federated search.
- (7) What standards should we adopt? How do we choose and implement them?
- (8) SharePoint, SharePoint, SharePoint!
- (9) Exploiting convergence and deploying a portfolio of technologies.
- (10) KOS maintenance and governance – includes implications for policy.

Some broader themes also came through in the list of clinic topics and the ensuing discussions. There was a focus on the practical. Much of the discussion focused on the “how” of undertaking these projects, rather than concerns with the “why”. This might be expected with an audience of practitioners. Ethical issues were not necessarily top of mind but the discussions that we heard often surfaced broader “so what” questions about KOS projects (and explicitly in the “buy-in” discussion).

There was also a combination of concerns. Some of the issues were technical (eg integration, convergence, automation), some were methodological (eg scoping, process for building, standards, governance), some were user- and business-led (buy-in). KOS projects are “messy” – they are not purely technical efforts but involve design, technology and business elements combined.

Finally there was a range of expertise. Some participants were highly experienced and technically capable whereas some were starting from scratch. This range in expertise made for conversations that were both fruitful (because of the variety of viewpoints) and yet also challenging (because of the differences in experience). In the short-term, this breadth of expertise is only likely to increase among information professionals unless we seek to build our collective competencies. Developing a framework for self-management of competencies related to knowledge organisation was the topic of a final plenary from Matt Moore.

REFLECTIONS AND CONCLUSIONS

Overall, there are three words that stuck in the authors’ minds after the conference:

“Change” – The speed of technological change is now a cliché and the world is full of “hockey-stick” graphs. The authors have been writing articles on knowledge organisation for *Online Currents* since 2009 and what is new is not the technologies under development but the scale of their implementation. As the case studies above indicate, people are doing real things with linked data. In terms of the hype cycle, we are hopefully moving into the plateau of productivity. As we move from “proof of concept” to “business as usual”, we will learn exactly what works and what does not.

“Open” – Dave Clarke from Synaptica observed that the importance of open data was a recurring theme. It is easier to access linked open datasets – with APIs and clear protocols for reuse – than closed ones. In Australia, we can see initiatives such as Govhack promoting such activities in the public sector but less so in the private domain – for now. Organisations have always had to balance what they make public versus what they keep private. While some public information disclosure is a function of regulation (eg annual financial statements), much is driven by the self-interest of the organisation (eg marketing materials and press releases). The calculus of openness will shift in many

industries as data dissemination becomes easier and adventurous market players take the opportunity to get their data embedded in the everyday lives of their customers and consumers.

“Messy” – KOS projects are messy in a number of senses. They involve work with both technology and human beings. The multiple standards that underpin them are in flux. And while the main driver for change is technology, there are multiple platforms and many legacy systems. Even with SharePoint (so good that the clinic sessions named it thrice), are we talking about SharePoint 2013 or Office 365 or SharePoint 2010/2007/2003 or some version of the above that your organisation has customised in its own special way? This mess has to be managed productively and dynamically. Organisations can rarely afford the time and money to stop their efforts while they rebuild from ground up. Success in a messy environment requires both a wide variety of skill sets and a certain mindset. The mindset is a comfort with ambiguity. This may be a challenge for information professionals because we tend to prefer precision and order.

This event prompted some reflections on the past seven years since we began writing these articles. It will be interesting (and perhaps embarrassing) to look back from IKO 8 in 2022 at this review of the inaugural event. Some of the observations that we make here will doubtless prove naive. Some of the technologies mentioned will not succeed and other technologies that are around today but unmentioned here will doubtless end up being hugely influential. Many of the human factor issues will remain unchanged (excluding a major shift in human evolution in the interim).

The conference was a great deal of fun to be involved in with a good balance of content delivery and audience participation. There was a good mix of attendees – practitioners, consultants, academics and vendors. The number of people (approximately 100) also allowed the opportunity to talk to a significant proportion of other participants. Most were from Singapore but some were from further afield – including an attendee from Scotland! In spirit, it felt very similar to Australian events such as ACT-KM and Disrupt Sydney, where a premium is placed on engagement with a smaller group of enthusiasts rather than the delivery of marketing information to many hundreds of people.

While there are events in Australia that take a similar form to IKO, there are none that currently cover the same content area (perhaps the closest being those arranged by the now seemingly defunct Institute for Metadata Management). The annual Taxonomy Bootcamp in the US and the bi-annual ISKO UK conference are the closest in content globally.

The question for Australian information professionals is how they will keep up with current developments in the knowledge organisation domain as it accelerates.

