IKO
INNOVATIONS
IN KNOWLEDGE
ORGANISATION

Conference Guide

Innovations in Knowledge Organisation Conference

8 - 9 June 2015
e2i Campus, 80 Jurong East, St 21, Singapore 609607
www.ikoconference.org

ORGANISERS

This conference is curated by three international thought leaders and practitioners in knowledge organisation, supported by a distinguished Advisory Board:

synaptica®
knowledge organization systems

Dave Clarke

Dave is co-founder and CEO of the Synaptica® group of companies (www.synaptica.com), providers of enterprise software solutions for taxonomy and ontology management. David leads R&D at Synaptica, and is currently developing a range of Linked Data software solutions for ontology management, semantic indexing and content annotation.

Straits Knowledge
Redefining corporate knowledge

Patrick Lambe

Patrick is Principal Consultant of Straits Knowledge (www.straitsknowledge.com), and the author of *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness (Oxford 2007)*, one of the leading books in knowledge organisation. He consults, teaches and trains on taxonomy development and knowledge organisation around the world.

PEBBLEROAD

Maish Nichani

Maish is the founder of PebbleRoad (www.pebbleroad.com), an innovative user experience design consultancy that helps organisations turn complex, challenging problems into elegant, innovative designs. He has curated learning events on designing the search experience, building search based applications, using pattern libraries, service design, and usability.

OUR SPONSORS

poolparty

PoolParty is a world-class semantic technology suite that offers sharply focused solutions to your knowledge organization and content business. The software is a semantic middleware that enriches your information with metadata, links your assets automatically and contextualizes your business! End-users as taxonomists and data analysts can easily build and enhance an information network without coding skills. Boehringer, Credit Suisse, Roche and The World Bank among many other customers profit from transforming data into knowledge with PoolParty.

Experience PoolParty: www.poolparty.biz

IHS

IHS combines core competencies managing "must have" content with work flow tools and advanced analytics to deliver client solutions focused on critical information and insight. IHS helps the world's leading companies to improve engineering, R&D and innovation outcomes by providing authoritative, updated content alongside internal knowledge bases. Documentation management of authoritative engineering information sources from the IHS Standards Expert and IHS Knowledge Collections with Goldfire deliver world-class tools optimised for the information that your organisation needs the way that you need it.

Hinton Information Services (http://www.hintoninfo.com/EN/) is the authorised reseller for IHS in Singapore. Contact us today to discuss how your use of authoritative engineering information can be optimised across your organization.

Email: sgsupport@hintoninfo.com

SUPPORTING ORGANISATIONS

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION

IKI-SEA
THE INSTITUTE FOR KNOWLEDGE AND INNOVATION SOUTH-EAST ASIA BANGKOK UNIVERSITY

Hong Kong Knowledge Management Society

The Hong Kong Polytechnic University 香港理工大學

KMIRC 知識管理及創新研究中心
Knowledge Management and Innovation Research Centre

actKM FORUM

CONFERENCE PROGRAMME

Day 1 (8 June 2015)

LEARN

9:00am	**Welcome Address**	*Welcome to the conference from the Conference Curators.*
9:15am	**Opening Keynote** - Patrick Lambe "The New Frontier in Knowledge Organisation: integrating technologies, standards and methods to deliver extraordinary results"	
10:00am	Tea Break & Networking Session	
10:30am	**Knowledge Organisation Tutorials** in Pecha Kucha format - each topic, 20 slides x 20 seconds each! • Taxonomies, Ontologies and Knowledge Graphs – *Patrick Lambe* • Autoclassification and Text Analytics – *Ahren Lehnert* • Linked Data – *Dave Clarke* • Standards for Taxonomies and Linked Data – *Marjorie Hlava* • No SQL and Graph Databases – *Loh Yu Gene* • Open Source Search – *Charlie Hull* • Search Based Applications – *Agnes Molnar* • Internet of Things – *Maish Nichani*	*Get a rapid grounding in the elements of Knowledge Organisation from the experts.*
11:30am	**Panel on Knowledge Organisation** - Q&A	*Explore the topics raised in the Tutorials with our panel experts.*
12:00pm	Lunch	

EXPLORE

1:00pm	**Case Study Café (1)** - Case Pitches 1) "Next generation SharePoint: Office Graph, the technologies it powers, and the opportunities and challenges that follow" – *Agnes Molnar* 2) "Using text mining and analytics to improve the search experience and enterprise taxonomy for an oil and gas products and services company" – *Ahren Lehnert* 3) "Implementing Open Source Search for a major specialist recruitment firm" – *Charlie Hull* 4) "OASIS: constructing knowledge bases around high resolution images using ontologies and Linked Data" – *Dave Clarke* 5) "Leveraging Linked Data to connect content and products across the BBC " – *Elisa Erriquez & Jenni Wardell* 6) "Design Methodology for a context aware mobile search based application, using diverse data sources" – *Maish Nichani* 7) "Using Linked Data and taxonomies to create a quick-start 'smart' thesaurus" – *Marjorie Hlava* 8) "Padipedia: building a knowledgebase on padi production using semantic web technology for MARDI (Malaysia Ministry of Agriculture)" – *Nor Azlinayati Abdul Manaf*	*Case presenters for our first 8 case studies will give a short pitch introducing their case study to the plenary audience.*
2:00pm	Break out case study table discussions (a) Each case presenter will host in-depth discussion on their case study	*Decide which case study in-depth discussions you want to explore.*
2:20pm	(b) Move to another case study table discussion to explore a second case study of interest to you	
2:40pm	(c) Move to a third case study table discussion to explore a third case of interest to you	

3:00pm	Tea Break & Networking Session	
REFLECT		
3:30pm	**Knowledge Café** – How could I apply what we have covered, back in my organisation?	*Table discussions followed by an expert panel Q&A and discussion.*
4.45pm	Capturing key questions for Day 2	*Major questions captured on cards for consideration on Day 2.*
5:00pm	**Day 1 Close**	

Day 2 (9 June 2015)

EXPLORE		
9:00am	**Review of Day 1**	*Key themes and questions from Day 1.*
9:20am	**Day 2 Keynote** - Charlie Hull "The Future of Search"	
10:00am	Tea Break & Networking Session	
10:30am	**Case Study Café (2)** - Case pitches 9) "Lessons learnt from deploying a hybrid folksonomy-taxonomy approach to enhance navigation on a tourism information site in Hong Kong" – *Eric Tsui* 10) "Building knowledgebases using named entity recognition, query expansion and Linked Data, to support enhanced discovery" – *Haliza Jailani* 11) "Enhancing search and discovery across libraries, archives and museums using text, data and audio analytics, using a single OneSearch service" – *Kia Siang Hock* 12) "Deploying a Semantic Operating System in a Small Financial Services Company" – *Mark Glikson* 13) "KRSTE.my: using Linked Data and semantic search to build a one stop knowledgebase on Science and Technology R&D in Malaysia" – *Nor Azlinayati Abdul Manaf* 14) "Providing access to knowledge to accelerate innovation in an R&D function for a food and beverage manufacturer" – *Mark Garlinghouse* 15) "Using a graph database to build a highly scalable and flexible knowledge mapping tool with taxonomy management system connectivity" – *Patrick Lambe & Loh Yu Gene*	*Case presenters for our second 7 case studies will give a short pitch introducing their case study to the plenary audience.*
11:30am	Break out into case study table discussions (a) Each case presenter will host an in-depth table discussion on their case study	*Decide which case study in-depth discussions you want to explore.*
11:50am	(b) Move to another case study table discussion to explore a second case of interest to you	
12:10pm	(c) Move to a third case study table discussion to explore a third case of interest to you	
12.30pm	LUNCH	
APPLY		
1.30pm	**Knowledge Organisation Clinic (a)** - Framing the issues and opportunities	*Participants will organise themselves into groups around the topic areas and frame their needs.*

1:55pm	**Knowledge Organisation Clinic (b)** - Working with the experts on how to apply Knowledge Organisation methods in your organisation	*Participants will work with expert panelists at tables to explore their own situations and how they could apply knowledge organisation approaches and technologies.*
2:55pm	Tea Break & Networking Session	
REFLECT		
3.25pm	**Plenary** - Matt Moore "Building Competencies for Knowledge Organisation"	*Plenary*
4.00pm	**Panel Discussion** - The Future of Knowledge Organisation	*Exploring where the technologies, methods and standards are likely to go next*
4.45pm	**Conference Summary and Review**	
5.00pm	**Conference Close**	

CONNECT ON SOCIAL MEDIA

#ikosg2015

www.ikoconference.org/blog

CONTACT US

David Clarke
dave.clarke@synaptica.com

Patrick Lambe
plambe@straitsknowledge.com

Maish Nichani
maish@pebbleroad.com

Address
Straits Knowledge
77B Neil Road
Singapore 088903

Tel/Fax
+65 6221 0383

OUR SPEAKERS

Nor Azlinayati Abdul Manaf
Lead Knowledge Engineer
MIMOS Berhad, Malaysia

Nor Azlinayati is the Lead of the Knowledge Engineering Team for Artificial Intelligence Lab, MIMOS Berhad and has been involved in various Knowledge Modelling and Engineering projects such as KRSTE.my, PADIPEDIA, Intelligent Business Matching, Medical Data Coding and Harmonisation. She also authored and co-authored a few patents and publication papers mainly in Semantic Technology as well as Knowledge Modelling and Engineering domain. Her research focus is on Knowledge Representation and Reasoning. Prior to joining MIMOS Berhad, she has had over 8 years of teaching experience in various IT and Computer Science areas such as artificial intelligence, database system and programming languages. She was one of the instructors for the Protégé OWL Tutorial conducted at the University of Manchester for 3 years. She obtained her Bachelor Degree (Hons) in Computer Engineering and Masters Degree in Information Technology (Artificial Intelligence) from the Multimedia University.

David Clarke
CEO and Head R&D
Synaptica®, UK

Dave is co-founder and CEO of the Synaptica® group of companies, providers of enterprise software solutions for taxonomy and ontology management. His previous roles include CTO of Synapse Corporation and Global Taxonomy Director at Dow Jones. He served on the NISO committee responsible for authoring the 2005 version of the US national standard for controlled vocabularies, ANSI/NISO Z39.19, and is a Fellow of the Royal Society of Arts. David leads R&D at Synaptica®, and is currently developing a range of Linked Data software solutions for ontology management, semantic indexing and content annotation.

Elisa Erriquez
Linked Data Team
BBC, UK

Elisa joined the BBC from the Traffic Support Division of the multinational company Mouchel. Before that she worked for digital agencies and mobile applications companies. She now works in Platform API within the BBC on a team called Linked Data Platform, which focuses on semantic, linked data and linking all aspects of the data within the BBC. Originally from Italy, Elisa joined the Software Engineering Research Lab (SerLab) as a research assistant, after her degree in Computer Science, working on Federated Database Management research. During her time there, she enrolled for a Master degree, which she completed at the University of Liverpool (UK). She was later offered a PhD at the University of Liverpool where she worked on modelling trust and distrust in multi-agent systems, using argumentation systems.

Mark Garlinghouse
Hinton Information Services
Singapore

Mark has over 21 years of experience in the information industry, spending most of that time developing markets in Asia Pacific. Most recently, he held a number of executive management positions at Thomson Reuters, responsible for supporting the needs of customers in the Life Sciences, Intellectual Property Solutions, Scientific and Scholarly Research and Healthcare business segments. He worked closely with customers engaged in research and innovation intensive industries including pharma, biotech and national accounts. From a base in Singapore, Mark currently provides consulting services and advisory support to a number of companies in the information and SaaS sectors looking to expand their markets.

Mark Glikson
Founder, CEO & Chief Architect
Gumbuya, Singapore

Mark is Founder and CEO of Gumbuya and Chief Architect of its core platform. Gumbuya was founded in 2012 to realise Mark's vision of how Semantics and UX should be central to the way enterprises transform into digital businesses. Gumbuya has since grown from idea to full stack platform, the first Native Semantic O/S offering enterprises a revolutionary way to build modern digital experiences. Mark has a deep 20-year history in platform software, including close to a decade at Microsoft in Enterprise Architecture, Strategy and Leadership roles, most recently as the GM of Microsoft's Platform and Developer Group for Asia Pacific.

Haliza Jailani
Assistant Director (Metadata Services)
National Library Board, Singapore

Haliza started as a librarian in a public library and moved to cataloguing, implementing metadata and knowledge organisation systems for digital collections in 2003. She was programme manager of digital infrastructure which delivered a metadata management system, a meta-search engine and a digital preservation system in 2010. With a specialist degree in

knowledge organisation and management, her current portfolio includes integrating physical and digital cataloguing, up-skilling cataloguers to support the semantic web, implementation of NLB Linked Data and aggregation and harmonisation of metadata for specific services.

Marjorie Hlava

Author of the Taxobook & President of Access Innovations, Albuquerque, New Mexico

Marjorie is well known internationally for her work in the implementation of information science principles and the technology that supports them. She is president of Access Innovations, Inc., which she founded in 1978. She served as president of NFAIS, which awarded her the Miles Conrad lectureship in 2014. She has also served as president of ASIS&T, which awarded her the prestigious Watson Davis Award in 2014. She served two terms on the Board of Directors of the Special Libraries Association (SLA), which is presenting the John Cotton Dana Award to her in June 2015. She is now developing ontological structures to serve linked data, which she feels is the future of semantics and information science. She is the author of five books, including the recent Taxobook 3-volume series.

Charlie Hull

Flax Consulting, UK

Charlie is the co-founder and Managing Director of Flax, specialists in all aspects of search technology based in Cambridge, U.K. Flax has been developing powerful search applications based on open source software for clients including The Financial Times, UK Government, Australian Associated Press and Reed Specialist Recruitment since 2001. Charlie also runs the Cambridge Search Meetup and London Lucene/Solr User Group and regularly talks and writes on search-related subjects.

Kia Siang Hock

Deputy Director, Technology Architecture and Innovation National Library Board, Singapore

In Siang Hock's current role, he and his teams are heavily involved in the conceptualisation, Proof-of-Concepts (PoCs), design and development of various innovative services at NLB. He has over 20 years of diverse IT experience in the areas of IT management, architecture, applications, systems operations, and applied research and development. Prior to NLB, Siang Hock held various IT positions at the Singapore Institute of Management, Pacific Internet, Jurong Town Corporation and Defense Science Organisation.

Patrick Lambe

Author of Organising Knowledge & Principal Consultant Straits Knowledge, Singapore

Patrick is the author of *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness (Oxford 2007)*, one of the leading books in knowledge organisation. He is co-founder and Principal Consultant of Straits Knowledge, an Adjunct Professor in KM at the Hong Kong Polytechnic University, Visiting Professor in the KIM PhD programme at Bangkok University, 2-term past President of the Information and Knowledge Management Society and a member of the editorial advisory board of the Journal of Knowledge Management, Knowledge Management For Development Journal, and Journal of Entrepreneurship, Management and Innovation. He consults, teaches and trains on taxonomy development and knowledge organisation around the world.

Ahren E. Lehnert

Taxonomy & Search Analyst FMC Technologies, Houston, Texas

Ahren is an information management professional with over ten years of experience in taxonomy, search, and content and records management. He has developed enterprise and eCommerce taxonomies in a consulting capacity for clients in a broad range of industries. Ahren is now the Taxonomy & Search Analyst at FMC Technologies, an oil and gas products and services provider based in Houston, Texas, where he is responsible for enterprise taxonomy development, auto-categorisation, and search. In addition, he is actively involved in text analytics, ECM, and other projects aimed at improving content management and retrieval throughout the organisation.

Loh Yu Gene

Executive Director Stronium Sdn Bhd, Malaysia

Gene is experienced in software development—having led teams in Malaysia and abroad—and implementing solutions for organisations of various sizes. He has keen interest in open systems and the adoption of rapidly advancing technologies in building applications for knowledge management, including the use of Linked Data and graph databases. Gene has a degree from the University of Western Australia with majors in Computer Science and Information Technology.

Agnes Molnar

CEO
Search Explained, Hungary

Agnes is CEO and Managing Consultant of Search Explained, specialising in Enterprise Search and Information Architecture. She has worked for various companies throughout the world, developing dozens of Enterprise Search implementations for both commercial and government organisations. She is especially renowned for her familiarity and extensive knowledge in SharePoint and other Content Management Systems. Since 2008, Agnes has been a consistent recipient of the prestigious Microsoft Most Valuable Professional (MVP) Award, for actively sharing her technical knowledge. She is a regular speaker at technical conferences and symposiums around the globe. She has also co-authored several books and white papers. She also maintains her passion and dedication on the subject through her blog, where she shares troubleshooting tips, best practices, and other useful resources in Content Management with a light and wholesome approach.

Matt Moore

Manager, Market Operations
PwC, Sydney, Australia

Matt has spent 15 years working in knowledge and information management, learning, marketing, sales operations and communications with organisations such as PwC, IBM, Oracle and the Australian Securities and Investment Commission. He lectures at the University of Technology Sydney, writes for Online Currents and is a former chair of the New South Wales Knowledge Management Forum.

Maish Nichani

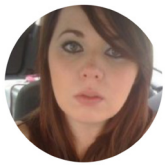
Author of Organising Digital Information for Others and founder of PebbleRoad, Singapore

Maish is a highly respected thought leader and practitioner in the field of user experience design with over 10 years of experience in designing complex information environments. His projects cover user needs analysis, technology selection, usability design, information architecture and intranet design. He is the founder and Principal of PebbleRoad, an innovative user experience design consultancy that helps organisations turn complex, challenging problems into elegant, innovative designs. He has been an active member of the design community since 2000, having spoken at international conferences, held workshops, written articles and two books on information design. Maish has curated and led learning events on designing the search experience, the art of persuasive design, building search based applications, using pattern libraries and content modelling, service design, and usability.

Eric Tsui

Professor
Hong Kong Polytechnic University

Appointed as the Professor of Knowledge Management at the Hong Kong Polytechnic University under the President's Distinguished Professionals Scheme in September 2002, Eric joined the University as a full time staff in March 2005. In the past 10 years, he has designed and delivered numerous public and custom-designed knowledge management and technology workshops. He has also consulted for many government departments and private organisations in Australia, Hong Kong, Thailand, Singapore, Malaysia, and Brunei. Eric is Vice President of the Hong Kong Knowledge Management Society, an honorary advisor on KM to the Hong Kong Police College, and a member of the IT Technical Advisory Sub-Committee in the Hong Kong Hospital Authority. Eric is also the Chairman of the Advisory Committee on eLearning, Chairman of the HKPolyU's MOOC Steering Group and Facilitator of the Community of Practice on use of Learning Management System to enhance Teaching and Learning at his university.



Jenni Wardell

Linked Data Team
BBC, UK

Jenni joined the BBC in December 2014 as Product Owner for the Linked Data Platform. Before that she worked for digital agencies and mobile applications companies, as well as a consultant for several exciting start ups. She now works in Platform API within the BBC on a team called Linked Data Platform, which focuses on semantic, linked data and linking all aspects of the data within the BBC. Originally from Canada, Jenni started out working for Lenovo doing technical roadmapping for the Government, Education and Healthcare sector clients for their 5 year refresh cycles and planning on their large research initiatives. Eventually preferring to be more hands on, Jenni moved into the world of consulting for mobile and application design and development, with a focus on UI/UX. In 2014 Jenni also became an advisor on the EU CLIPS project, helping advise from a B2B perspective of the software innovations for smart cities. Now with the application of strong user experience in the BBC and providing all of the data the BBC has to offer, it is an exciting initiative.

CASE STUDY CAFE

DAY 1

You will be able to attend in-depth table discussions on three of these cases on Day 1. The case outlines are provided to help you decide which case discussions you would like to attend.

1) AGNES MOLNAR

Next Generation SharePoint: Office Graph, the technologies it powers, and the opportunities and challenges that follow

1. About the Case Organization
The organization of this case study is a global software development company, headquartered in the US. Their main assets are their software and the special knowledge behind.

2. About the Challenge
The company has a very special domain of knowledge which is very hard to track. Also, their teams are dispersed in several offices around the world. Sharing knowledge as well as finding or even more discovering knowledge in the company has been always one of the biggest challenges.

Although this company has been having a very strong Enterprise Search implementation, it was still not enough for all their knowledge discovery scenarios.

3. What We Did
Since the customer had already migrated most of their content to Office 365, it was obvious to leverage the benefits of Office Graph and Delve. Although these new, revolutionary tools are brand new releases from Microsoft, the POC was very successful. The users started to be able to discover important knowledge management content – which made their existing search-driven experience fuller and richer.

4. Challenges and Lessons Learned
The main challenge was (as always) the content readiness: we had to make sure everything is stored properly. As a side-effect, the improvements on the content made their Search-Driven Application better too, although this was not the main object of the project.

The second challenge was to teach the users how to use this new approach and toolset, and how to combine with and differentiate from the common search experience. Although the user experience in Delve is very ergonomic and easy-to-use, the algorithms behind it and the structure of Office Graph is very challenging to explain – therefore hard to set the expectation bar.

The third challenge still remains, although we'll be able to solve this very soon. Namely, the Office Graph cannot be "customized" today. It consumes signals from Office 365 sources only, for a specific set of content only. With the upcoming updates, we'll be able to send signals from custom sources, and this will enhance its value even more, very soon.

5. Impact and Benefits
In general, thanks to the very good and easy user adoption, the project is already successful, even in this early phase of the supporting tools. Working in any industry, the value of Office Graph and Delve is unquestionable – I have seen very similar success stories in global manufacturing and pharma industries, too.

Usually, the adoption cycle of new tools like Office Graph and Delve is much longer. In my opinion, the main reason of these early success stories is the gap these tools fill – they give the enterprises not a new tool, but a new approach as well, which helps the users get their job done much more effectively.

6. Next Steps

As a next step, we'll definitely start using the upcoming enhancements, like sending new signals to Office Graph. Being able to add on-premises content to the Graph will make a significant increase in its value. Also, we're investigating in what kind of custom application layer would fit the customer's special needs, besides Delve's out-of-the-box experience.

2) AHREN LEHNERT

Using text mining and analytics to improve the search experience and enterprise taxonomy for an oil and gas products and services company

1. About the Case Organization

FMC Technologies is an oilfield equipment and services company based in Houston, Texas specializing in Subsea Technologies, Surface Technologies, and Energy Infrastructure. FMC Technologies has more than 18,000 employees in 24 facilities around the globe.

2. About the Challenge

In our move from six separate search instances in four separate SharePoint farms to a single instance of SharePoint 2013 search, we needed to consolidate search terms and phrases from across the globe to seed typeahead search suggestions, gather terminology for taxonomy development for automatic categorization, and develop singular "best bet" locations for common activities.

Prior to our approach there was no single search instance or singular reporting on search queries providing insight into the types of information being sought by employees across the globe. The search experience was confusing and fragmented and search results were poor. Additionally, there was no common controlled vocabulary for metadata values applied to all searchable content, resulting in a fragmented information retrieval experience. The new search implementation and use of the FMCTI Taxonomy for the automatic application of metadata affects all SharePoint users at FMC Technologies since SharePoint is the platform for our Intranet.

3. What We Did

We collected all search terms and phrases from the various search instances, regularized the output by consolidating duplicates, correcting corruptions in the information output, and eliminating search term noise. We then analyzed the final, consolidated list of search terms and phrases, looking for patterns in the information to further develop the FMCTI Taxonomy and determine best bet locations for promotion in search results.

Our Intranet platform consists of four, separate instances of SharePoint 2010 and our new single search tool is SharePoint 2013 search. We use Smartlogic Semaphore to manage the FMCTI Taxonomy, for text mining, and auto-categorization. We used SAS JMP Text Miner for search term analysis. There were no standards applied to the analysis, but we use various terminology standards for the creation of terms for the FMCTI Taxonomy.

The innovation in this approach was that lack of consolidation and reporting prompted us to use other analytical methods to understand better the search needs of our employees globally. Rather than reacting to what we believed or assumed people were searching for or extending our learnings from earlier user behavior testing, we used real data to arrive at an understanding of the actual topics users were seeking in content.

4. Challenges and Lessons Learned

Our main challenge was extracting and consolidating data from disparate sources due to duplicate searches with separate forms and counts, searches in multiple languages, and corrupted information in the output. Once we are on a single search, analysis of search terms will be easier as the search counts will be consistent. Multiple language searches will continue to be a challenge, but we will map common concepts to a single English term form to facilitate information retrieval in multiple languages.

Lessons learnt:

There is such a thing as too much analysis. Sometimes having a broader view of the information is better than having detailed, nuanced results. We will spend less time on initial analysis and more time on regular analysis in smaller chunks over greater periods of time.

Advice:
The time spent on pre-launch analysis is worth the effort to avoid starting from scratch and further alienating already frustrated users by implementing a search which appears to have no connection to what existed before. The analysis provided a baseline from which to measure search success and provide real data from which to make decisions.

5. Impact and Benefits
Business benefits are as yet to be determined. The prospective business impact is that employees will spend less time searching for information and understand the value of categorizing content in order to group content and understand relationships between concepts to make new information discoveries. Extracting more value from existing information is critical, particularly when an industry faces a downturn.

It's hard to say if the benefits could have come from other sources. However, the lack of trust and use of the search tools as they were showed that any improvement would be a success. We will know that our effort made a difference as we monitor the rate of search use and successful and unsuccessful searches. We hope to extend this analysis to determine connections between information retrieved and actual business decisions and their outcomes.

Reasons for success:
The need for an improvement in search bordered on desperation. Many employees have been asking for improved search. The delivery of an improved search tool and user experience coupled with the automatic application of standardized metadata will not only provide a better search experience but will also bolster the design of search-based applications to improve execution in specific use cases.

6. Next Steps
Our next steps are to continue to enhance search functionality with improvements in the search user experience and additional search features. In addition, we will actively pursue the development of search-based applications within SharePoint and connections to external information in other systems and file shares. Expanded search consolidation will also broaden the use of classification of content with standard metadata from the FMCTI Taxonomy. Finally, continued analysis of search terms and response to user requests for content classification will aid in the further development of the FMCTI Taxonomy and lead to improved content classification.

3) CHARLIE HULL

Implementing open source search for a major specialist recruiting firm

1. About the Case Organization
Founded in 1960, Reed Specialist Recruitment is a specialist provider of permanent, contract, temporary and outsourced recruitment solutions, and IT and HR consulting. Reed operates in Europe, the Middle East and Asia Pacific and has more than 3,000 permanent employees working out of 350 offices worldwide across 30 specialisms.

2. About the Challenge
The main objective was to replace the existing recruitment search function which 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. The search function was used by all Reed's recruitment consultants to find candidates, jobs and companies and its slow speed and varying quality led to an increased workload.

3. What We Did
Flax designed a flexible and powerful search infrastructure based on the open source search engine Apache Lucene/Solr. A custom indexer written in Java navigated the highly complex legacy database structure used by Reed. Using Apache Tika, plain text was extracted from CVs (Resumes) in Microsoft Office & PDF formats and in many languages including English, Polish and Chinese. A test framework was built to gauge and tune performance. A single search server provides the entire search capability over 12 million records, with sub-second search response times.

4. Challenges and Lessons Learned
The complexity of the existing data model led to a highly flexible pipeline being required, to allow us to easily change how data was mapped into the search index. The new search engine was also such a radical change in terms of performance and features that users required significant retraining to take advantage of these – they simply had forgotten to trust search. Organizations planning this kind of migration should focus on the user experience in particular. Moving to an open source platform may also

require a change in development culture and thus training.

5. Impact and Benefits
The project was delivered on time and on budget – unusual for a project of this scale – and the project was hailed as a great success by the client, part of a large-scale reorganisation of their IT capability. Reed have continued to innovate in search and have developed many further systems using a similar platform. The project was a success due to careful planning, tight project management and the clear benefits of moving to an open source solution – flexibility, scalability, no vendor lock-in and no license fees.

6. Next Steps
Flax has continued to work with Reed to develop powerful search-based solutions for other areas of their business. Flax has also since worked with other recruitment firms to develop similar capabilities.

4) DAVE CLARKE

OASIS: Constructing knowledgebases around high resolution images using ontologies and Linked Data

1. About the Case Organization
Synaptica produces software solutions for: building and managing taxonomies and crosswalks; designing and deploying knowledge organization systems; indexing and enriching content; and optimizing search, navigation and discovery.

2. About the Challenge
Visual images provide a valuable complement to textual information, but a vast amount of the information inside photographs, paintings, diagrams and drawings can be seen but not searched - it has been inaccessible to traditional query methods.

Many business applications could benefit from the ability to search inside images including: medical and scientific imagery, reconnaissance and intelligence, engineering and design, forensics and security, education and cultural heritage.

3. What We Did
Synaptica built a software system called OASIS that allows points and regions inside images to be highlighted and annotated. These visual features are then tagged using taxonomies and Knowledge Organization Systems. The software makes visual content searchable with pin point accuracy. It also promotes knowledge discovery as the application dynamically identifies features and related concepts as the user freely pans and zooms around an image.

The key technologies behind the solution are Linked Data and RDF graph databases. These allow users to connect to vast amounts of high-quality structured information in the Linked Open Data cloud, including authoritative Knowledge Organization Systems and ontologies. The extensive use of ISO and W3C standards and specifications ensures data portability and systems interoperability.

The fusion of several core technologies (high definition imagery, Linked Data, Knowledge Organisation Systems and semantic annotation) represents an innovative solution that opens up new opportunities for enriching visual content.

4. Challenges and Lessons Learned
Working with external data sources from the Linked Open Data cloud presents a number of challenges: (i) external data can be accessed by live queries to remote third-party servers, but these remote systems may not be able to provide adequate uptime availability or responsiveness; (ii) copies of external data can be ingested into local systems, but some datasets, such as DBPedia, may be too large to be accommodated on the available data storage; (iii) while graph databases out-perform relational databases at pattern-based queries, relational databases typically out-perform graph-databases at indexed or full text searches.

Synaptica responded to these challenges by building a flexible system that can simultaneously query data from any number of internal or external data stores. Low volatility data of a manageable size can be ingested while high-volatility or very large datasets can be accessed on remote servers in real-time.

5. Impact and Benefits
The result of the effort is a robust and scalable general purpose toolset that can be used to build taxonomies, access external Linked Data, and annotate image content. By leveraging Linked Open Data, much of which is available without license fees, the time and cost to deploy knowledge organization systems can be greatly reduced.

5) ELISA ERRIQUEZ & JENNI WARDELL

Leveraging Linked Data to connect content and products across the BBC

1. About the Case Organization

The British Broadcasting Corporation (BBC) is the UK's public-service broadcaster. It is the world's oldest national broadcasting organisation and the largest broadcaster in the world by number of employees, with over 20,000 staff in total, of which 16,672 are in public sector broadcasting.

Jenni and I are part of the Nexus A&C team (previously known as Linked Data Platform), which is part of the BBC Platform, BBC Digital business unit; responsible for all digital media services. It was created to help the BBC solve some of the immense challenges as the pace of digital change increased.

2. About the Challenge

In 2010, just before the Football World Cup, the sport team realized that Sport journalists spent a lot of time managing content within the CMS, creating indexes, deciding whether to remove a certain article, deciding where in the page the article should appear. From a journalist's point of view, it was more important to create content than deciding the layout of the page. Therefore it was decided that a new approach was needed to free the journalists, and a tagging or semantic annotation approach was the way forward.

3. What We Did

Initially, the approach was trialed for a microsite, hosting the content about the 2010 Football World Cup, an approach now evolved and used by many of the major web sites at the BBC (BBC News, BBC Sport, BBC News App, K&L, etc).

The microsite used methodologies such as Domain Driven Design and technologies such as RDF and OWL. A special team called DSP (Dynamic Semantic Publishing) was created to achieve this result.

The strategy adopted allowed the evolution from a relational content model and static publishing framework towards a fully dynamic semantic publishing (DSP) architecture. The DSP architectural approach underpinned the re-launched BBC Sports site and the BBC's Olympics 2012 online content.

DSP used linked data technology to automate the aggregation, publishing and re-purposing of interrelated content objects according to an ontological domain-modelled information architecture, providing a greatly improved user experience and high levels of user engagement.

The DSP architecture curates and publishes HTML and RDF aggregations based on embedded Linked Data identifiers, ontologies and associated inferences.

4. Challenges and Lessons Learned

BBC News, BBC Sport and a large number of other web sites across the BBC are authored and published using an in-house bespoke content management/production system ("CPS") with an associated static publishing delivery chain.

The BBC World Cup 2010 site featured a 700-plus team, group and player pages, which were powered by the high-performance DSP architecture.

Previously, BBC Sport would never have considered creating this number of indexes in the CPS, as each index would need an editor to keep it up to date with the latest stories, even where automation rules had been set up. To put this scale of task into perspective, the World Cup site had more index pages than the rest of the BBC Sport site in its entirety.

Obviously we encountered numerous challenges during the development of the Linked Data Approach. The main ones were discovering the power of inference (too powerful?) and the use of external data. A balanced understanding of inference through the use of weak and strong semantics has been the key to resolving these issues. However, some problems still remain and we are still looking into innovative solutions. The key to the success of the project has been the initial use of a small domain, as a prototype to prove the validity and value of the approach.

5. Impact and Benefits

This approach enabled the BBC to support greater breadth and scale, which was previously impossible using a static CMS and associated static publishing chain. The DSP system has now evolved into a set of RESTful APIs and a Triple Store storing more than 37 million triples, managed by our Nexus A&C team. The Linked Data is managed through these APIs and various tools to allow the user to create new "Tags" and to link content.

The most noticeable thing about this work is the scale of the approach of Linked Data at the BBC. Main stakeholders include BBC News, BBC Sport and BBC Worldwide.

The adoption of Linked Data has freed time for journalists, has made opening data possible, enabled quick implementation of large sites, such as the London Olympics.

6. Next Steps

The next step the BBC is currently taking is developing the project called myBBC.myBBC is an innovation programme that will deliver the tools, services and capabilities the BBC needs to make the best use of data to deliver more personalised experiences for its audiences. The two key sections of the project are called Profile and Personalisation.

Profile will provide a destination that joins up all of an audience member's BBC experiences and data and will enable a truly cross-platform, cross-product experience. It will put the audience in control of their data and personalisation, allow them to discover new products and let them share their experiences with other audience members.

The Personalisation proposition allows the BBC to gain a deeper understanding of its audience by offering product teams a set of services that enable audiences to tailor their BBC experience.

The myBBC team and the Linked Data team (Nexus A&C) are going to be working closely with each other to use linked data technologies to their full potential to deliver a product that will allow the audiences to find and consume content in the way that they specify across all BBC products and services, "placing the BBC back in the audiences' hands."

6) MAISH NICHANI

Design methodology for a context aware mobile search based application, using diverse data sources

1. About the Case Organization

The organisation is a government agency that carries out regular household surveys.

2. About the Challenge

Our client carries out household surveys. They recruit freelancers to do the job. They wanted to help them better optimise the time taken and success rate of the surveys.

A key challenge was to plan the sequence of the houses to visit. It turns out that the default list given to the surveyors was not optimised. The sequence of houses to visit depends on many contextual factors such as distance to a bus stop, rain and age of the residents, which governs when they are usually at home.

The challenge: how might we build a mobile app to help the surveyors better optimise their household visits route for a day?

3. What We Did

We first went on a field study with the surveyors. We followed them on their daily visits and noted down their strategies for the challenges they encountered. Eg some of them used color codes to highlight household preferences to do the survey online, by phone or by a visit.

We made a list of the factors that affected the visit route. There were many. Location, distance, climate, bus and taxi routes and even the age of the residents. We got data from diffrerent sources:

- Location – Google
- Distance – Google
- Directions - Google

- Climate – National Enviornment Agency
- Bus, Taxi – Land Transport Authoritiy
- Age of residents – Client organisation

We then created a mobile app prototype that used the findings along with data provided by the phone to create an experience that felt more like a guide-on-the-side.

The route listing changes depending on several factors:

- Starting location - from home or at a bus stop
- Ending point - at home or at a food court
- Public transport - shortest distance by bus or train
- Climate - depending if its going to rain or not

Age of residents - elderly prefer mornings, while those working prefer evenings.

4. Challenges and Lessons Learned

It is amazing what is available on the Internet, usually free to use. The challenge is to know what is required and how to use it correctly.

The field study helped us discover that climate data is important to surveyors - the route changes when it rains. Now that we know we can offer a better experience with climate data, the challenge was to know where to get it from. Luckily, National Environment Agency provides such data. But now we had to learn how to incorporate it using their API.

APIs can be a problem. Each provider will have their own API format. This makes integration harder. We hope that this will change in the near future when linked data becomes more prevalent.

We then discovered that using all this data increases the mobile data usage. Will surveyors pay for the extra data use? Or will the client pay a fixed amount for extra mobile data use? Our clients are thinking of offering some kind of reimbursement.

Finally we discovered that we have to use a service mindset. It is not just about the app but the entire experience - from the time they set out in the morning to the time they return in the evening.

During the field study we found that the surveyors were fumbling with the papers they were carrying. As part of the solution we decided to design a bag that could fit the phone and all the papers and stationery they had to carry.

5. Impact and Benefits

We tested the prototype with the surveyors. They felt that the app could help speed up their work. This was more for those new to the survey process. The app helped take away the planning work and gave more time for the surveyors to focus on the survey.

For the client, they realised that there is data available that can enrich their own data to offer a compelling experience to the surveyors.

The client also got instant updates on what is happening on the field. Without the app they would only know about the day in the evening, when the surveyors would log their day using an online app.

Lastly, the app analytics could be used to shed light on how the surveyors do their job: from the type of transport they choose to how they plan their routes. This info can be then used to improve the entire process.

6. Next Steps

The first version of the app is getting built. The client wants to take a phased approach to build the full app. There were concerns on the interation, data use, stability and security of the app. The first version is to evaluate these issues.

7) MARJORIE HLAVA

Using Linked Data and taxonomies to create a quick-start "smart" thesaurus

1. About the Case Organization

The two current applications of this approach are a large scientific publisher and a large association with a robust publishing platform.

In the first instance, the publisher has a growing, broadly based collection of over 100,000 articles from over 250,000 authors. The second is an association with a defined topical area; its 220,000 articles are from 900,000 authors and 29,000 institutions worldwide.

2. About the Challenge

Thesaurus applications have been in use for many years. With the increasingly complex and interconnected world available through digitization and the internet, the expected connectivity of information is growing at a logarithmic rate—everyone finds themselves buried in huge amounts of information. This has led to an ubiquitous need to track, find, and programmatically tag the information in meaningful ways for enhanced retrieval.

Increasingly, researchers are going directly to Google rather than to publisher websites to fulfill their research needs. This has a direct impact on the already struggling scholarly publishing industry.

Two instances of this challenge have led us to design the Smart Thesaurus. We have already proven the ROI of enriching content with subject metadata; further, we know that a thesaurus—with all of its relationships—provides a sound conceptual platform for information organization and retrieval. The next issue was how to enlarge that thinking to an actual working application for broad conceptual areas. We considered ontologies and other options for searching. We decided to build a proof-of-concept based on enriching and interlinking the data itself—combining Linked Data with a concept thesaurus.

3. What We Did

We started by enriching the data using standard thesaurus (taxonomy term) tagging with our Data Harmony suite of tools (Thesaurus Master, a vocabulary management tool, and M.A.I, our Machine Aided Indexer) to automatically apply the terms to the articles. Next, we used the M.A.I. to tag the data full text inline. That is: wherever a thesaurus concept was mentioned in the full text, triggered by the M.A.I. Concept Extractor Engine, we inserted the taxonomy term as a full XML string, effectively enriching the content directly in the full text of each article.

In parallel, we added a field for Linked Data—to hold URI to a persistent source, such as dbpedia or Wikidata, for each thesaurus term—to the Thesaurus Master application. Therefore, each preferred term can have a link to one or more external resources, from which we can query off definitions, links to other data sources, and other information to create dynamic web interfaces.

We adhered to the ANSI/NISO z39.19 Thesaurus Standard, which is comparable with the ISO 25964-1 Controlled Vocabulary Standard, for all thesaurus/taxonomy term record creation. We used DBpedia as a link to external resources and definitions where applicable.

4. Challenges and Lessons Learned

One problem we encountered is that most Linked Data sources are not granular enough to supply a link for every term in a highly specialized scientific thesaurus; the more specific the term, the less likely that a relevant Linked Data source is available.

The optimal solution to this problem is to create the missing Linked Data pages (on dbpedia, Wikidata, or whatever Linked Data source is in use) to enrich the source data in the publisher's specialized area; further, backlinks to the publisher's topic pages (or other relevant areas of their websites) should be added as external resources. In this way, the organization is contributing to and enriching the growing network of linked data sources available on the web as well as promoting themselves as thought leaders in the industry.

5. Impact and Benefits

The results are dramatic increase in uptake of the articles and interlinking of the core content as well as building the community for each topical area.

The benefits are clear.

- Staff time is saved by better information retrieval, freeing them for other activities
- Better search results for the end user (in this case, researchers)

- Portals with embedded Linked Data can stream dynamically generated content from external data sources (other websites, social media, news, images) alongside the publishers' own content, establishing these portals as "one-stop shops" for researchers—this helps to make the sites "sticky" by keeping users from leaving to search other sites
- Topic pages on publishers' websites enhanced with definitions queried from Linked Data sources without staff curation of content

Without the core thesaurus and the automatic linking this set of tasks would have taken several man years of labor to produce the same result. This solution, with the ever increasing burden of so much content, is scalable.

6. Next Steps

After combining content stored in a highly structured XML markup language (such as JATS), a well-formed topical thesaurus, and Linked Data, the next logical step is to implement a robust RDF triples database underneath the content and other data structures. This will help to make Smart Content—websites that are not only enhanced for human users, but completely machine-readable and able to draw inferences between concepts, objects, and entities elsewhere on the web.

8) NOR AZLINAYATI ABDUL MANAF

Padipedia: building a knowledgebase on paddy production using semantic web technology for MARDI (Malaysian Ministry of Agriculture)

1. About the Case Organization

The Malaysian Agricultural Research and Development Institute (MARDI) is a leading agricultural research centre in Malaysia. MARDI was established in 1969 with the main objectives of generating and promoting new, appropriate and efficient technologies towards the advancement of the agriculture, food and agro-based industries. It is located in Serdang, Selangor and has 32 branches nationwide. MARDI is mandated to conduct research in the fields of science, technical, economy, and social with regards to production, utilization and processing of all crops (except rubber, oil palm and cocoa), livestock and food.

2. About the Challenge

Main objective of the effort

To preserve knowledge about research and development on paddy and make it accessible for knowledge discovery.

Issue or problem

The ageing workforce represents an issue with knowledge loss as retirement occurs. Knowledge loss will threaten the paddy and rice research process and activities in order to improve/sustain the rice industry in Malaysia. Electronic and written memos, journals and books provide expert knowledge but they do not connect the knowledge of the whole value chain of paddy research. Thus the paddy ontology is built to be the repository for storing and connecting the knowledge.

Prior to PADIPEDIA implementation, various divisions that are distributed nationwide were "disconnected" from involvement in the paddy research and development activities, making it more difficult to have an integrated knowledge repository.

3. What We Did

The PADIPEDIA application is a web-based portal with searching capabilities, basic analysis and reporting function using Semantic Technology. The development of paddy ontology is an attempt to utilize semantic web technology for organizing knowledge. The paddy ontology will be the repository for storing and connecting the knowledge of the whole value chain of paddy production which includes breeding, agronomy, production system, pest and disease management, post-harvest and product development which is in line with the research scope of MARDI. Padipedia culminated from the need to preserve paddy knowledge and make it accessible for knowledge discovery. Users can obtain paddy-related information not only from within MARDI but from external sites as well.

Innovations about this effort are:

- Ability to perform Semantic Search based on concepts rather than keywords.
- Ability to perform question answering using natural language.
- Provides user access to published knowledge base in the form of open linked data which goes beyond PADIPEDIA knowledge base by linking users to other publicly available knowledge repositories in the world within the Open Linked Data Cloud.

Ability to perform Decision Support based on knowledge stored in PADIPEDIA Knowledge Base.

4. Challenges and Lessons Learned

- Gathering the background knowledge from the subject matter expert.
- Capture shared understanding of domain of interest from various discipline of paddy research (breeders, agronomist, plant pathologist, food scientist)
- People perceive this as another IT project

Overcome the hurdles by conducting several workshops with the domain experts. Need to explain thoroughly to them.

Advice to others

Identify the real champions who are responsible for providing all the materials required to implement the system.

5. Impact and Benefits

Impacts:-

a. Availability of strategic and reliable paddy info that provide knowledge to researchers and policy makers;
b. enable networking, collaboration among knowledge workers;
c. Repackaging of information for specific needs that can reduce search time for information and learning material;
d. Online streaming of information which offers real time information from source
e. Context based search which reduce search time; more accurate search to meet with specific work needs;

Benefits:-

a. Policymakers and Planners (Public & Private Sectors) - better decision making and projecting future venture.
b. Business community (Venture Capitals, Entrepreneurs, SMEs) - product and services innovations.
c. Research community – research extensions and product and services innovations.
d. Education community (Private & Public Institutions) – research extensions.
e. Individuals interested in Paddy and R&D activities related to paddy

6. Next Steps

MARDI is currently working on several ways to promote PADIPEDIA especially among the paddy community in Malaysia. Some of the main MARDI initiatives are:

- Continuous engagement with relevant organisations and other potential information provider through various meetings, discussions and forums;
- Organise road shows involving relevant organisations and researchers to get information on PADIPEDIA as well to capture updated information on R&D activities; and
- Taking part in many of its Ministry events organised by cluster agencies as a platform to promote PADIPEDIA to the public community.

MARDI also plans to extend the application to include knowledge from other commodities.

CASE STUDY CAFE

DAY 2

You will be able to attend in-depth table discussions on three of these cases on Day 2. The case outlines are provided to help you decide which case discussions you would like to attend.

9) ERIC TSUI

Lessons learnt from deploying a hybrid folksonomy-taxonomy approach to enhance navigation in a tourism industry KM system in Hong Kong

1. About the Case Organization

Showcasing a prototype of the TaxoFolk (Taxonomy-Folksonomy) system for enhanced knowledge navigation to the KM team of a large tourism company that already has strong KM governance and rooted in a culture of top down deployment of enterprise KM systems.

2. About the Challenge

- The company has already established taxonomies for classifying assets; however the taxonomy was set by a small group of staff and has not be updated regularly
- Taxonomy is getting out of date and users are increasingly frustrated as they waste time in locating knowledge assets
- Try to enlighten the team that taxonomy and search engines are not the only means to navigate knowledge assets; that user-labelled information can also be incorporated into the enterprise KM system
- Try to explain to the KM team that analyzing user labels/tags also help to better understand why certain words are used by users in qualifying information and this knowledge is valuable at taxonomy revamp stage as well as for search engine configuration. Additional vocabularies are often discovered from analyzing user tags/labels
- The KM team has approximately 12-15 members drawn from Corporate, HR, IT and other departments

3. What We Did

- Original research on combining taxonomy and folksonomy to enhance knowledge navigation in terms of a hybrid/extended site map
- Technologies involve user coding of clustering algorithms, (the TaxoFolk algorithm) codified IF-THEN rules to identify word stem and variants, checking of WORD-Net online dictionary and Wikipedia to verify the authenticity of words/terms and show relationships between them
- I explained the TaxoFolk algorithm, the rationale behind it, I showed the prototype and gave demonstrations

4. Challenges and Lessons Learned

- "Taxonomy" is very new to most Hong Kongers let alone the term Folksonomy. My biggest problem in promoting TaxoFolk is to survive the complexity of explaining several seemingly new and difficult-to-understand terms before the audience can ascertain what is the power of TaxoFolk
- Pick and approach companies/users that/who are familiar with information classification; companies that welcome bottom up contributions and especially those that are already using some kind of Web 2.0 tools
- Just deploy the prototype and let users try it for a pilot period. Let them experience the superiority in knowledge navigation and only then further explain how the hybrid navigation structure is put together
- Additional vocabularies revealed from the collected user tags/labels serve as an excellent source to further ascertain the gaps between use of naming and sometimes even the perceptions of corporate and perceptions of individual users, also useful for updating and maintaining the corporate taxonomy
- The system was not fully implemented, because the organization did not have the skills to re-develop the algorithms for TaxoFolk. However, when they finally upgraded their portal (SharePoint), they chose to adopt user tagging in the platform as a high priority.

5. Impact and Benefits

- Though the system was not implemented, the KM team finally realises the power and impact (offering an additional dimension to locate asset) of a tool like TaxoFolk
- Common tags used by users turn out to be a valuable source of information to consider during taxonomy revamp especially on the setting of terms and phrases; further discussion with users on how certain tags/labels are used reveals the mental models of many users
- At the time when I showed TaxoFolk to the organization, they were not piloting other projects hence I would say much of the change and understanding are attributed to TaxoFolk. The KM team appreciates that during taxonomy revamp, more consideration needs to be paid to a user's perspective on what they perceive and their use of common terms (which may well be different from terms that the organization has adopted)
- Delivering explanations at various levels of abstraction, patience, and provide a prototype or demo system for users to try

6. Next Steps

Tag clouds can also be generated from user contributed tags and are provided to users as an additional mean to navigate over knowledge assets. These tags have improved the findability of explicit assets in the organization. From the research perspective, we intend to apply for funds to support personalization of the TaxoFolk hybrid navigation system for individual users; at present it is one system for all users based on the contributed tags.

10) HALIZA JAILANI

Using knowledgebases using named entity recognition, query expansion and Linked Data, to support enhanced discovery

1. About the Case Organization

The National Library Board Singapore manages the National Library, 26 Public Libraries and the National Archives, with a mission to make knowledge come alive, spark imagination and create possibilities. NLB supports knowledge seeking activities by providing a trusted, accessible and globally-connected library and information service.

2. About the Challenge

With the merger of National Archives Singapore (NAS) and the National Library Board in November 2012, it became critical to harmonise library and archives metadata, name headings & controlled vocabularies so the public can discover resources from both agencies without having to go through two portals. Metadata and terminologies need to be aligned. Knowledge organisation processes need to be centralised for efficiency and cost savings, and common tools shared. Not many libraries and archives in the world have merged their collections, difficult to do as archives and libraries organise materials differently. The multi-level description of archives relates objects in a hierarchy and links the parts to a larger ensemble from the collection level perspective. Libraries organise at the item-level.

Names are largely not standardised in NAS, where 8 different databases are managed individually by different teams. As a result a single person may have more than one form of name. Search results for this person are not unified and resources are retrieved according to the name a user enters. Merging of NAS collection as it is with the NLB collection will cause search results to be more fragmented. A search for an entity whether a person, an organisation or a place will not pull content resources about the entity into a single list. Like other NLs, NLB uses names authorised by Library of Congress Name Authority Cooperative Program (NACO) which observe strict rules for capturing every part of a name. Where names cannot be established in NACO, NLB uses a separate list from a local file.

In addition, the library world is undergoing a shift in thinking on how libraries catalogue resources and enable them to be discovered. The long-time library cataloguing standard AACR2 (Anglo-American Cataloguing Rules 2nd Edition) was replaced with a less rigid standard that is more attuned to the needs of a digital world, called RDA or Resource, Description and Access. Increasing efforts are also being made to break the coded information out of the catalogue record so that they can be used as Linked Data and library materials can become searchable on the Internet.

3. What We Did

NLB used a three-pronged approach to organise the merged collections and improve resource discovery: (i) Data harmonisation & query expansion (ii) Linked Data implementation, (iii) Knowledge-base building.

NLB & NAS data harmonisation

Mapping was done between the archival schema, ISAD (G) and Dublin Core a generic schema for interoperability, where NLB had previously crosswalked 1.5m MARC records to DC. We managed the hierarchical ISAD (G) with item-level DC relationships and collection level description. About 6,000 people and organization names were extracted from the archives indexing system, de-duplicated and mapped to NACO and NLB local files. The finalised set was centralised in NLB's Knowledge Organisation System (KOS) and maintained in the Thesaurus & Taxonomy Editor (TTE), a vocabulary management tool. System enhancements and data integration were done for the archives, a new website launched and OneSearch, an integrated search for NLB (4.5m titles) and NAS (0.8m titles) was delivered in Aug last year. Due to the different forms of names, requests from owners, aliases and use in different languages, we are now working to provide a query expansion service that will enable the non-preferred names and associated forms of name to be searchable as desired.

NLB Linked Data implementation

We implemented a Linked Data management system and created an extensible data model and an extended service vocabulary to support a web feature for NLB & NAS websites. More than 250,000 records and named entities were transformed into RDF. We incorporated a selection of the National Heritage Board's galleries and museums metadata so as to create an RDF triple store of library, archive and museum data. Going beyond publishing the RDF data, NLB takes this a step further by conceptualizing a widget for NLB and NAS websites that will allow users to perform contextualised searches based on entities found on the website. Term extraction was done for unstructured data to enable nuggets of information to be extracted and added to NLB's triple store. Based on the raw data extracted, named entities were identified and added to the knowledge-base. The data was also analysed so that explicit relationships can be built into the extended service vocabulary.

Knowledge-base Building

The above two approaches require a strong knowledge-base to meet discovery needs effectively. NLB named entities are described by a set of properties such as latitude and longitude and dates of birth and death. These set of properties need to be enriched if we are to deliver a useful service. We have started enriching this knowledge-base and are working to enhance it by tapping on existing data on the Internet using Linked Data. By leveraging on other knowledge-bases, we hope to be part of a global network of contextualised data.

4. Challenges and Lessons Learned

Working beyond the library makes us realise that data exist at different levels and are organised differently because of this. Example, orientation or arrangement of work is important for museums and archives to understand the meaning and context of the artefacts and archival materials. NLB do not manage its content this way. Do we capture and control vocabularies for such purposes? Names may be authorised in different forms for Getty's Union List of Artist Names vs NACO. Which should be used or should a decision even be forced? Materials & technique and object/work type classifications for museums are at a very high granularity compared to the type categories found in libraries and archives. Compromises need to be made and interoperability can only occur at a high level. Insufficient metadata exist and vocabularies are not controlled or are unlikely to be controlled due to various reasons. NLB is looking at cleaning and reconciling data for the three agencies but this effort has its limitations.

5. Impact and Benefits

NAS content access and use has jumped substantially and we are now working to deliver OneSearch for the museum's artefacts on the SGCool website (80,000 items) by mid this year. Linked Data exposes poor quality data. A good outcome from this is a commitment to improve the metadata descriptions and the use of applicable standards that would aid in collection organisation and stakeholders' buy-in.

6. Next Steps

NLB is now in a position to share building blocks and learning points from its semantics-based resource discovery journey. NLB is also intending to share local terms, vocabularies and name headings to organisations with local cultural heritage collections and welcome their participation and use of the knowledge-base for their own resource description and discovery. NLB has set targets and resources for the coming years to increase the knowledge-base and publish them on international registries. Internally, we will also be transforming a further 700,000 authority and metadata records into RDF to enrich the web feature. We will continue to publish our RDF data on international registries and begin work with other national libraries to link the collection into a network of Linked Data.

11) KIA SIANG HOCK

Enhancing search and discovery across libraries, archives and museums using text, data and audio analytics, using a single OneSearch service

1. About the Case Organization

The National Library Board of Singapore (NLB) manages the National Library, the 26 Public Libraries and the National Archives of Singapore (NAS). NLB promotes reading, learning and information literacy by providing a trusted, accessible and globally-connected library and information service through the National Library and a comprehensive network of Public Libraries. NAS oversees the collection, preservation and management of public and private archival records, including government files, private memoirs, maps, photographs, oral history interviews and audio-visual materials.

2. About the Challenge

Libraries and archives collect and provide access to digital resources of significant national, cultural and heritage values. Their collections are painstakingly curated, described and made accessible.

A seismic shift in the information seeking behaviour has taken shape. With the ever-expanding reach of the Internet, users are accustomed to quick and easy access to content, and are expecting an online experience that is rich and instant. How can libraries and archives continue to reach out to their users in this era of shorten attention span?

Every year, NLB users collectively contribute to tens of millions of e-retrievals. We see every single one of these interactions as a golden opportunity to 'push' relevant content to the user. The success of the 'customers who bought this item also bought' recommendation feature at amazon.com is a clear testament of the power of pushing relevant recommendations. However, to effectively connect people to content, we need to connect content to content first.

3. What We Did

To manually associate related content for any sizeable collection will be very labour-intensive. Text and data mining (TDM) technologies can automatically, efficiently and accurately identify these associations. TDM is akin to sieving through all the content, and applying well-established information retrieval theories to identify similar content.

NLB implemented TDM on its digital content, and identified over 1 billion similarity associations. We can now bring together the full spectrum of resources available in NLB regardless of institutions, formats and language and present them in a way that encourages a deeper understanding of the precious memories of the nation.

It is also what knowledge seekers have been dreaming of. They can spend less time gathering the pieces, and focus on digesting and analysing the 'dossier' of relevant information to derive new insights.

4. Challenges and Lessons Learned

The first hurdle that we needed to cross is the technical one. As we planned to use TDM for large data sets, we looked for solutions built on the popular Hadoop big data platform. The expertise in these technologies was lacking locally. We have our own Solutions Architect to work on proof-of-concepts to pick up the skills.

The next challenge came when we need to scale the TDM to process a data set of over 10 million items. It proved too much even after we scaled our Hadoop cluster to 13 servers. We went back to the drawing board, and figured that we could split the data set into smaller clusters of related ones before the TDM processing. This 'divide-and-conquer' approach worked very well.

5. Impact and Benefits

With over 1 billion associations identified via TDM, a massively inter-connected network of knowledge has been formed.

The related resources identified via TDM are made easily available to the users. The users are now able to view and explore related resources within and across collections and media formats. Information is no longer viewed in isolation, but seen as a part of the larger context. We call this contextual discovery.

The initial results have been promising. The level of access to the resources has increased.

6. Next Steps

We are exploring several possibilities to bring contextual discovery to the next level:

- *Cross-institution*. With NAS coming under NLB, we now have an expanded collection of digital resources on Singapore. We will be working on linking NAS and the National Library resources.
- *Cross-language*. There are 4 official languages in Singapore: English, Chinese, Malay and Tamil. We are considering the use of machine translation to provide cross-language recommendations.

12) MARK GLIKSON

Deploying a Semantic Operating System in a small financial services company

1. About the Case Organization

O-Pulse provides a portal for investment advisors to manage client relationships, compliance, and communications with portfolio managers. Currently in private Beta, O-Pulse was founded by a leading third party money management & financial advisory firm based in North Carolina.

2. About the Challenge

Typical of many conventional Enterprises today, the founders of O-Pulse, in their existing financial advisory firm, recognized the need to transform their traditional offline organization into a modern digital enterprise. Management faced increasing pressure, with a noticeable uptick in client attrition to competitors with even the most basic online client portals. O-Pulse was established to create an online portal better servicing their existing advisor clients, and the wider US-based independent investment advisor market.

The v1 O-Pulse portal was designed to facilitate complex targeted content distribution and syndication between portfolio managers, investment advisors and their clients. News, articles, portfolio commentaries and financial updates could be curated, shared and syndicated through one central dashboard with full compliance reporting. A key requirement was to enable white label deployment of the O-Pulse portal to any number of organizations under a SaaS subscription model.

O-Pulse engaged with multiple development firms only to be faced with exorbitant open-ended quotations. All proposals utilized a common “silo” build methodology; inflating costs throughout the development cycle with limited asset re-use and extensive

infrastructure configuration & maintenance.

Recognizing the need for agility to remain competitive in today's marketplace, O-Pulse also demanded a high degree of architecture flexibility and extensibility not possible in traditional modes of development, enabling freedom to deploy, test and iterate over time to grow a truly transformational product. O-Pulse did not have the financial and human resources to outsource the project under the traditional development model, recognizing this model would also fail to deliver their true agility requirements.

3. What We Did

Starting with a Semantic O/S, O-Pulse was able to leverage schematized data structures, defining real-world entities, things and objects semantically without the need for database setup or manipulation. A semantic operating system works natively with semantic details (about things and the relationships between them), enabling complex operations without the need to structure databases or build significant amounts of boiler-plate operating code to support them. User friendly Semantic OSes have natural language interfaces to allow users to interact with the system. By simply describing the model of each users journey and key interactions between them, O-Pulse had already designed their systems complex information architecture: Advisors and Clients become Contact nodes; market portfolios become Products; information & content become Articles. Connections between these nodes defined key relationships. A Client is connected to one or more Portfolio products; an advisor is connected to all of their clients.

The remaining application build involved assembly of a combination of pre-fabricated and custom-built modules, with the bulk of assembly work completed by a UX professional. Like all other assets, modules are represented semantically in the O/S, enabling re-use, extensibility and interoperability. V1 of the O-Pulse portal was assembled entirely with pre-fabricated modules, together with custom CSS styling.

Future versions of O-Pulse will require new custom modules to grow the available module library, including API integrations with financial institutions and custodians to pull portfolio data. Custom modules such as this will be isolated and passed to specialized developers for build, becoming permanent reusable assets upon completion.

Leveraging inbuilt platform capabilities, O-Pulse is able to deliver its white label portal to clients without having to engineer a subscription or deployment system. The entire O-Pulse portal was assembled with zero infrastructure or database setup, no IT Pro's or database administrators, and complete flexibility to modify and customize in future without refactoring.

4. Challenges and Lessons Learned

O-Pulse's seemingly unique workflows around content management and distribution required highly granular control over outcomes triggered by specific actions. Gumbuya's challenge was to implement a platform-level framework that supported the unique requirements of O-Pulse, yet remained applicable to a broad range of scenarios. By decomposing complex interaction flows to their core fundamentals, the key composite actions were identified. A decision framework was then constructed enabling triggering of these actions in a custom sequence, controlled by the end user through the platform's visual assembly environment. As a result, O-Pulse requirements were met and a key piece of platform functionality was delivered.

As a SaaS product, O-Pulse required a mechanism to deploy new user accounts, and display user-specific content within each client's portal. Gumbuya was able to achieve this by implementing a user-customizable rights-access framework in its graph-based environment, enabling role-driven access to specific groups of graph nodes.

5. Impact and Benefits

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.

The Semantic O/S removed the need for any infrastructure setup, ongoing maintenance or associated IT Pro personnel. Combined with the use of pre-fabricated modules, built-in subscriptions and SaaS deployment capabilities, O-Pulse was able to significantly reduce cost, time & complexity. In addition, founders could remain directly involved with the assembly of their portal under a semantic build model using fundamental real-world concepts and language they understood.

Leveraging the underlying graph architecture, with a fundamental separation of content, styling and structure, O-Pulse achieved their goal of building a platform with a high degree of extensibility and flexibility to adapt & change to meet evolving client needs.

O-Pulse has already identified a number of areas for product expansion, which can be seamlessly integrated without reworking the underlying system architecture.

6. Next Steps

O-Pulse is currently in private beta, adding additional integrations and product features while collecting user feedback, with a phased release scheduled over the coming months.

O-Pulse is planning to continually grow the platform, with a number of significant product extensions scheduled in the post-release roadmap. The focus for v2 is on integrating portfolio aggregation and reporting, progressing towards creating a true "one-stop" dashboard for investment advisors.

13) NOR AZLINAYATI ABDUL MANAF

KRSTE.my: Using Linked Data and semantic search to build a one stop knowledgebase of Science and Technology R&D in Malaysia

1. About the Case Organization

The Malaysian Science and Technology Information Centre (MASTIC) was established in 1992. MASTIC is a division under Ministry of Science, Technology and Innovation Malaysia (MOSTI) located in the Federal Government Administrative Centre Putrajaya, Malaysia. MASTIC is responsible to provide reliable, up-to-date and comprehensive science, technology and innovation (STI) information through efficient information management system.

2. About the Challenge

- KRSTE.my should be a knowledge management system, a hub for R&D in the domain of Science, Technology and Innovation in Malaysia.
- KRSTE.my Portal should serve as a Single Point of Access to the wealth of knowledge artefacts from various internal and external systems sources.
- KRSTE.my should organise the nation's research knowledge : collect all research results, funding, researchers, etc, in a national repository and classify them according to MRDCS (Malaysian R&D Classification System)
- KRSTE.my inventory nation's research capability : Develop a national inventory of research capability (competence) of researchers and institutions
- KRSTE.my instrument for effective research policy : Provide overall view of funding, research areas and accomplishment for policy making on innovation
- KRSTE.my should turn information silos into an integrated knowledge resource

3. What We Did

KRSTE.my leverages on the power of Semantic Web Technology through the use of ontologies and structured knowledge bases, Linked Open Data (LOD) and knowledge graphs, graph search and graph databases. KRSTE.my is a collaborative tools:

- to manage Science, Technology & Innovations project implementation
- reduce project redundancy
- preserving institutional to avoid "reinventing the wheel"

Innovations about this effort are:

- Ability to perform Semantic Search based on concepts rather than keywords.
- Ability to perform question answering using natural language.
- Ability to perform Researcher Profile Analytics using the existing data in the knowledge base, such as analysis on Researcher Collaborations, Experts Identification and Researcher Conflict of Interest.
- Provides user access to published knowledge base in the form of open linked data which goes beyond KRSTE.my knowledge base by linking KRSTE.my users to other publicly available knowledge repositories in the world within the Open Linked Data Cloud

4. Challenges and Lessons Learned

Challenges:

- Increase researcher registration and participation
- Continuous updating information from the researchers

During development of KRSTE.my, around 93 organizations were engaged to enquire relevant source of information on R&D activities in Malaysia. The information provider comprised of:

- Universities and other Institute of Higher Learning;

- Government Agencies including Research Institutes;
- Private Companies; and
- NGOs

MASTIC is always working on novel ways to promote KRSTE.my especially among the researchers in Malaysia through seminars and workshops.

5. Impact and Benefits

Impacts:-

- Availability of strategic and reliable S&T Info that provide informed decision by policy makers;
- Enable networking, collaboration among knowledge workers;
- Repackaging of information for specific needs that can reduce search time for information and learning material;
- Online streaming of information which offers real time information from source
- Context based search which reduce search time; more accurate search to meet with specific work needs;

Benefits:-

- Policymakers and Planners (Public & Private Sectors) - better decision making and projecting future venture.
- Business community (Venture Capitals, Entrepreneurs, SMEs) - product and services innovations.
- Research community (Education & Industry) – research extensions and product and services innovations.
- Education community (Private & Public Institutions) – research extensions.
- Individuals interested in S&T and R&D particular on indigenous knowledge in S&T and R&D

6. Next Steps

MASTIC is currently working on several ways to promote KRSTE.my especially among the researchers in Malaysia. Some of the main MASTIC initiatives are:

- KRSTE portal can be accessed through Institutional of Higher Learning, Government Agencies, Library and other relevant organization website & portal;
- Continuous engagement with relevant organisations and other potential information provider through various meetings, discussions and forums;
- Organise road shows involving relevant organisations and researchers to get information on KRSTE.my as well to capture updated information on R&D activities; and
- Taking part in many of its Ministry events organised by cluster agencies as a platform to promote KSTRE.my to the public especially researchers.

14) MARK GARLINGHOUSE

Integrating internal and external knowledge sources to accelerate innovation in an R&D function for a food and beverage manufacturer

1. About the Case Organization

A Fortune 500 food and beverage manufacturer with a team of about 1000 R&D professionals based in 12 facilities around the world restructured their R&D capability as a consequence of a global re-organization. One of several conclusions reached as a result of a review R&D capabilities was that the R&D repository and R&D content assets were woefully underutilized. The organization was centralizing the R&D capability in order to better coordinate global research activity

2. About the Challenge

The client desired to deliver better R&D outcomes in the following ways:

- provide better access to existing R&D content and avoid duplicate research
- provide consistent access to content sets representing both internal and external sources: past R&D results, authoritative content (both proprietary and open source including patent data, scientific journal literature, standards and regulatory information), and partner content where available
- provide more effective search results to both information professionals as well as lay searchers

The client had an existing R&D repository with high quality content that was difficult to retrieve. It was where "good research was left to fade away." Each R&D facility made their own decisions about licensed information resources and data structure leading to inconsistency, extra expense, and difficulty sharing across businesses.

The client had about 1000 R&D professionals supporting a diverse product line of food and beverage products. Innovation in the industry comes not only from new and enhanced food and beverage products but also from innovation in packaging.

3. What We Did

After working with the client to clarify desired outcomes and project deliverables, we recommended a solution that provides more effective access to existing content and identifies relationships between internal and external content through semantic indexing. This enabled identification of any content relevant to research interests in the context of those research interests.

With an inventory of the existing client content to include in the system and clarity about the research focus, we 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. Implementation required server integration and an ODBC interface to leverage legacy content. The solution leveraged existing and newly created term dictionaries to enhance context for the semantic search engine.

Recognizing the challenge of any technology solution is more in adoption than implementation, we invested significantly in change management necessary to drive adoption. This included identifying a team of internal champions: power users, mentors, knowledge base designers, quality control agents and analysts.

4. Challenges and Lessons Learned

We learned a great deal while implementing and tuned the implementation where necessary working as partners with the client.

We deliberately planned to invest heavily in implementation and support of change management activity. We invested as partners with the client to ensure that they understood and derived full value from the system. With the client's internal champions, we drove continuous education about the differences between open source intelligence research tools and authoritative, consistent research tools. This was a user mind set change going from a focus on searching for sources to a focus on searching for answers.

We reinforced the importance of repeatable outcomes in R&D and usage behaviour evolved away from a default to Google searches. Client created case studies and use cases helped to drive adoption and foster cultural change about R&D practices.

We had and relied heavily on senior management support. We benefited from significant frustration from the previous R&D content systems, but we still had to actively manage expectations and to clarify optimal results. We experienced significant up take at launch and then more gradual continued.

5. Impact and Benefits

Use of the solution enables users to spend more time on research and less time looking for information. Anecdotally, we know that researchers identified content from both internal and external sources that were inaccessible prior to the system implementation.

The project leaders achieved the original goals. They are happy with the outcomes so far and we continue to work with them to quantify outcomes. System usage is the easiest metric to track, and anecdotes about avoiding research rework demonstrate the most value. (For example, just one research project avoided represents significant financial savings.)

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.

6. Next Steps

We continue to work with the client to ensure effective adoption and support the on-boarding of new R&D staff. The client monitors the availability of additional datasets to include in the portfolio and monitor query results to ensure they are providing quality results. We expect to roll out the system to users outside of the R&D teams in the future.

15) PATRICK LAMBE & LOH YU GENE

Using a graph database to build a highly scalable and flexible knowledge mapping tool with taxonomy management system connectivity

1. About the Case Organization

Straits Knowledge is a small consulting firm that specialises in knowledge audits, KM strategy development and taxonomy development. Part of the knowledge audit methodology is a collaborative knowledge mapping activity.

2. About the Challenge

Our methodology for knowledge mapping for the past decade has been paper-based. We have a high burden of transcription of knowledge asset descriptions into large and complex Excel tables, checking for accuracy, and manual analysis of the knowledge maps. Analysis of the knowledge maps needs to be highly flexible because different organisations suggest different questions to explore.

An additional use for the knowledge maps is to provide raw evidence for the first draft of a faceted taxonomy for KM purposes (the maps represent a comprehensive description of key organisation activities and associated knowledge assets, in the language of the user community). Extracting this language from the excel spreadsheets has been laborious and burdensome.

Prior to the development of the graph-based system, which we call Aithin™, the work involved in transcribing, checking and analysing the maps was very expertise and labour-intensive. In addition, the maps were very difficult for the client to handle and analyse, and troublesome to update.

3. What We Did

We decided to go with a graph database solution for the following reasons:

- The maps needed to be highly scalable – we work with a wide range of organisations, from small to very large.
- Making the transition from a paper-based system tightly controlled by us to a system largely controlled by the client meant we needed to take an iterative, prototyping approach to development. Not all the issues and requirements could be anticipated in advance, and graph databases are intrinsically flexible in this way. The main thing we had to be clear about was to get the data model correct – after that, development could be very flexible.
- Making the application available to clients meant we could not predict all of the possible types of reports that might be needed. Graph databases, given a robust initial data model, are extremely versatile in the kinds of reports and analytics they can power.
- We needed to be able to offer project-based hosted licenses as well as permanent installed licenses – this meant the solution had to be browser-based, lightweight, and extremely fast.

The solution was innovative not only because this solution is unique in the market, but because it is leveraging the graph database in a number of different ways, including the capability to export data that can be consumed for taxonomy development.

4. Challenges and Lessons Learned

Most of our challenges related to being inexperienced in software development. We were very fortunate to be introduced to a development company that was familiar with knowledge management and taxonomy solutions, and so understood the purposes for which the software was to be deployed – and fortuitously, also had expertise in graph database development.

The flexibility of the graph database software also meant that we did not have to have all our requirements tightly tied down at the beginning, we could quite quickly overcome the usability and functionality issues we encountered along the way, simply because of our inexperience at the beginning.

While the cycle time from beginning of development to release of a working beta was longer than expected, much of this delay was caused by our need to develop helpware to guide our clients through the knowledge mapping methodology on the screen.

Our main advice would be to find a software development partner with a strong track record in developing graph solutions, and who understands, or is prepared to understand, the core tasks you are trying to perform.

5. Impact and Benefits

Key benefits include:

- Rapidly reducing the consulting time taken to collect maps, validate them and produce the analysis
- Enabling mapping activities to take place remotely (on the hosted version we can monitor and provide feedback on a mapping

session as it happens live, without having to be physically present)

- The data in the application can be rapidly analysed and is capable of being exploited in very flexible ways – the time taken to extract candidate terms for a taxonomy is also vastly reduced

6. Next Steps

Our first version was deliberately very simple, so as to make it accessible to clients using it for the first time. In our first handful of projects we have picked up a number of requirements to improve the usability and the interactivity of the analytics, and we are now working on version 2 with those features added. In the next version after that we plan to build a connector that will allow us to do direct export of map data into a taxonomy management system such as Synaptica®, to shorten and simplify the taxonomy development process.

KNOWLEDGE ORGANISATION COMPETENCY FRAMEWORK

This competency framework was developed by Matt Moore in consultation with Patrick Lambe. It is intended to provide a simple self-assessment tool for practitioners working with Knowledge Organisation Systems (KOS) to identify their areas of strength and opportunities for improvement. It is indicative rather than exhaustive and we expect it to be developed further over time.

Rate yourself on the form overleaf, using the scale below:

- No Experience = I have no prior knowledge of this activity.
- Basic Understanding = I have an understanding of the concepts at work here (e.g. academic study or peripheral involvement in a project) but I have not successfully undertaken this activity.
- Undertaken Successfully = I have successfully undertaken this activity at least once.
- Undertaken Repeatedly = I have successfully undertaken this activity multiple (more than 3) times. I may coach others in how to undertake this activity.
- Innovating = I regularly undertake this activity and have developed new tools and techniques to improve its efficacy.

Once you have completed your self-assessment, we suggest that you identify the areas of greatest weakness **and** of most interest to your current role, and develop a simple action plan. Feel free to contact Matt Moore or the conference organisers for suggestions on self-development opportunities!

Matt Moore *matt@innotecture.com.au*

Dave Clarke *dave.clarke@synaptica.com*
Patrick Lambe *plambe@straitsknowledge.com*
Maish Nichani *maish@pebbleroad.com*

After the conference, we will send a link to this self assessment as an electronic survey, and if you have rated yourself Undertaken Repeatedly or Innovating, and would like to volunteer to help others, you'll have an opportunity to do so there!

My area of greatest interest are:

My action plan to develop these areas is:

	No Experience	Basic Understanding	Undertaken Successfully	Undertaken Repeatedly	Innovating
1 KNOWLEDGE ORGANISATION SYSTEM (KOS) PROJECTS					
1.1 KOS business case creation					
1.2 KOS project management					
1.3 KOS stakeholder mapping and engagement					
2 USER ANALYSIS					
2.1 Developing user segments and personas					
2.2 Conducting user observation and interviews					
2.3 Facilitating user workshops and focus groups					
2.4 Developing and testing prototypes					
3 CONTENT ANALYSIS					
3.1 Conducting content inventories					
3.2 Conducting knowledge audits					
3.3 Modelling data structures					
3.4 Analysing content semantics					
3.5 Running statistical tests					
4 MANAGING SYSTEMS					
4.1 Managing Content Management Systems (CMS)					
4.2 Managing Relational Database Management Systems (RDBMS)					
4.3 Creating SQL queries					
4.4 Working with graph databases					
4.5 Managing Hadoop installations					
4.6 Using data visualisation tools					
5 DEVELOPING KNOWLEDGE ORGANISATION STRUCTURES AND FRAMEWORKS					
5.1 Developing and implementing taxonomies, thesauri or controlled vocabularies					
5.2 Developing and implementing ontologies					
5.3 Developing and implementing metadata schemas and standards					
5.4 Working with text analytics and autoclassification					
5.5 Working with enterprise taxonomy management systems					
5.6 Working with Linked Data					
5.7 Integrating taxonomies and metadata with search tools					
5.8 Building search based applications					

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES