Bibliography for Tagging - Metadata Discussion

prepared by LWM Technology Services
Printed: 03/16/06 14:28

Anand, Vivek. ***The Myths about Meta Data***, by Vivek Anand and Mark Robinson DM Direct 04/08/2005. http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1025066

> "In business, the devil is in the details. That is why companies consider meta data (data about data) as a way to get a handle on the details of their business. However, if it isn't done right, metadata might as well be meta-Seinfield - data about nothing. A serious investment can be wasted without really finding out if the data you don't know can indeed make the data you do know better.
>
> "By definition, metadata is selected or summary information about data i.e., name, length, valid values or description of a data element. Meta data is stored in a data dictionary and repository. It insulates the data warehouse from changes in the schema of operational systems....
>
> "Meta data is the glue that binds enterprise information together. The arteries of a successful metadata solution will reach across most of an organization's functions. With numerous beneficiaries, metadata will have many champions throughout an organization. Although, a metadata implementation needs a dedicated administrator to coordinate the interests of multiple stakeholders and maintain the repository, upper management should champion its active maintenance to assure its success. If such coordination is possible, then everyone in the organization will become the masters of their data's domain."

Borenstein, Joram. ***Integration Theory, Part 2***, by Joram Borenstein and Russell Ruggiero. 3p. DM Direct Special Report 09/06/2005

http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1036180

> Part 1 of this Integration Theory series appeared in the August 12, 2005 issue of DM Direct and is available at http://www.dmreview.com/article_sub.cfm?articleId=1034584.
>
> "Back in late 1999, XML was considered an immature technology with a very low adoption rate among both vendor and developer communities. Today, things have changed a great deal regarding mainstream adoption of XML. A case in point is the recent announcement that the next version of Microsoft Office will save files in XML by default. And while 1999 may be looked at as the pivotal year for XML adoption, we now believe a similar scenario regarding the Semantic Web is developing in 2005. As with any revolutionary or groundbreaking effort, adoption is predicated on a number of factors that include guidance, education, and the ability to deal with setbacks. In any event, the Semantic Web is a compelling concept that is building critical mass at a very impressive rate." Includes sections on:
>
> - The Semantic Web
> - How Ontologies Relate to the Semantic Web
> - Popular Search Technologies
> - The Semantic Web
> - New Meta Data-Focused Solutions
>
> "The Semantic Web will most likely be built in stages by various communities of practice, which will act as the building blocks for this effort. Rather than just presenting a concept, we must leverage open standards, software, and hardware solutions, combined with accepted best practices to help solve business problems.

Burriesci, Jeannette. ***When Your E-mail Reads Your Mind***. 5p. Intelligent Enterprise 07/01/2005
http://www.intelligententerprise.com/showArticle.jhtml;jsessionid=XBOTTU4TCMAV0QSNDBGCKH0CJUMEKJVN?articleID=164301176

> "By making Office into an alternative presentation for back-end systems and by exposing enterprise and systems metadata to Office, your corporate developers, systems integrators or application vendors can make your users' e-mail read their minds, too. Ever since Microsoft introduced its Information Bridge Framework (IBF) last year, many developers have been busy taking advantage of the relative ease with which they can now integrate back-end applications into the Office front end. Enterprises have been clamoring for Office integration from packaged application vendors and system integrators for a long time. In the past, developers had to slog out hard-coded integrations. IBF makes this integration quicker, less expensive and more flexible."

Carty, Stu. ***Enterprise Meta Data Management - Fast Forward to 2005***. 3p. DM Direct Newsletter 08/26/2005 http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1035614

> ' "Enterprise meta data management means collecting and cataloging the contextual information about every aspect of the enterprise: data, information, systems, applications, processes, stakeholders, business rules, architectures and infrastructure," says Alan Perkins, chief solutions architect for ASG-Rochade and ASG- becubic at Allen Systems Group Inc. (Naples, Florida). "Enterprise meta data provides the answers to: What is it? Where is it? How did it get there? Why do we have it? Who owns it? and Who is responsible for it? It also answers: Who can change it and what is the process for doing so?" This statement, which is relevant today, could have also been made in the 1980s.'

Cass, Stephen. ***A Fountain of Knowledge; 2004 will be the year of the analysis engine*** 8p. IEEE. Spectrum 01/04/2004.
http://www.spectrum.ieee.org/WEBONLY/publicfeature/jan04/0104comp1.html

> Open commercial platform that will crawl the Web, XML tag the content, preparing it for data mining and semantic-like searching.

Dutra, Jayne. Developing and Applying Controlled Vocabularies in Large Organizations for Increased Business Value, Jayne Dutra, JPL Information Architecture, NASA Taxonomy Manager, [for the] Semantic Technology Conference. 37 slides. 03/01/2005
http://pub-lib.jpl.nasa.gov/pub-lib/dscgi/ds.py/Get/File-215/NASA_Tax_Sem_Tech_Conf.ppt

> Best Practices increase interoperability and extensibility:
> Faceted Classification Schema
> - Facets give flexibility and power
> - Modular in nature for easier maintenance
> - Can tag what is appropriate to the use case
>
> Polyhierarchy
> - Concepts can appear more than once
> - Enables knowledge discovery from multiple viewpoints
> - User-centric organization

Duttaroy, Arup***. Five Ms of Meta Data*** 3p. DM Direct Newsletter 04/15/2005.
http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1025568

> Meaningful, mature, manageable, maintainable and migratable. Arup Duttaroy currently serves as the OLAP technical lead and is an active member of the national Data Warehouse practice of Covansys, an IT Services organization. Duttaroy has been one of the key players providing full life cycle project execution and consulting services for various clients in government, retail and automotive industries.

Farrell, Vickie. ***The Need for Active Metadata Integration: The Hard Boiled Truth*** 3p. DM Direct Newsletter 09/09/2005.
http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1036703

> "This lack of what Gartner calls "semantic reconciliation" among data from different sources is inherent in a diverse, dynamic and autonomous organization. It's exacerbated by the fact that different tools (ETL, data modeling, BI and DBMS) follow broad industry standards, yet have their own proprietary extensions that create semantic problems across tools....
>
> "The problem is with the data meaning, or metadata. The purpose of metadata is to standardize legal definitions of data. There are tools available to address technical or syntactic metadata; that which defines things like the location, syntax, structure, source and value of the data elements. However, they tend to work only within their own environment and don't facilitate sharing among tools, data sources, applications or services. They don't even begin to address "semantic metadata. Unlike data management, metadata management is not an area where the tools available meet the needs of business and IT users."

Grey, Denham. ***Classification – does it work?*** 03/19/2006. Knowledge-at-work [blog]
http://denham.typepad.com/km/2006/03/classification_.html

Lamont, Judith. ***Search Plus: Far-reaching and Versatile in the Enterprise*** 3p. KMWorld, v. 12, i. 7 07/01/2003.
http://www.kmworld.com/publications/magazine/index.cfm?action=readarticle&article_id=1560&publication_id=1

> The push to increase interoperability of search across multiple data repositories is finding universal support for use of XML tagging as a unifying data formatting solution. Composite application solutions are becoming popular with distributed searching high on the list. Susan Feldman of IDC predicts the merger of content infrastructures into technologies that support structured and unstructured content, search, format conversion, security and categorization, and text mining. Enabling technologies will be XML, WSDL, UDDI, and SOAP.

Lamont, Judith. ***Unlocking enterprise data: Metadata holds the key.*** KMWorld 04/01/2005
http://www.kmworld.com/publications/magazine/index.cfm?action=readarticle&Article_ID=2084&Publication_ID=132

> "If you ask people to describe their pain,. Carty adds, they say they want a .Google-like product that can search their corporate information assets, as well as a .MapQuest. To see how information is interconnected and interrelated.. Metadata management helps achieve both of those objectives. Organizations will continue to struggle with an increasing volume of data stemming from compliance requirements, business operations or new input from technologies such as RFID. Getting a handle on metadata and using it effectively will become more critical, and a proactive role in exploring options will offer better results than waiting for problems to develop."

MIT. Libraries. ***Metadata Reference Guide; A guide to metadata by the Metadata Advisory Group of the MIT Libraries***. MIT 01/01/2004.
http://libraries.mit.edu/guides/subjects/metadata/index.html

- Selected metadata standards
- Metadata mappings (crosswalks)
- Suggested readings
- Metadata glossaries
- Metadata projects
- Metadata Advisory Group, MIT Libraries

Moulton, Lynda. ***Knowledge Management Section: Enterprise Knowledge Requires Professional Stewardship***. 3p. SLA. Library Management Div. 12/01/2005
http://www.sla.org/division/dlmd/IMPACT/Fall2005.htm#km

If you consider the foundation of library and information science that formed the basis for my graduate education (over 30 years ago), it was divided into three distinct disciplines that correlate to a more generalize business model today. Here is how that foundation breaks down:

1. Collection development that emphasized the scope of content, published and unpublished (internal) that would constitute a body of literature to support the work of our constituents. This requires knowledge of publishing, scholarly societies, news sources, government agencies and their publications, the fundamentals of business operations and where, within the operation, valuable internal material is likely to originate. To be effective today, content management initiatives must cover a similar scope plus the enormous expansion of media types that content encompasses. No other professional discipline is trained in the scope of content and media as librarians are.

2. Cataloging, classification and indexing were requisite components that address the issues of thesaurus/subject headings, controlled vocabularies, automated indexing methods (e.g. B-tree indexes, notched punch cards), bibliography, and broader philosophical models for organizing collections (e.g. Colon classification, faceted classification). Currently, technologies abound that automatically categorize, index and extract metadata. Content management systems (CMS), Document Management Systems, and Auto-categorizing engines contribute some tools that can aid with the more structured approaches. However, they all depend on information science expertise to bring quality and discipline to the indexing process that characterized human (original) cataloging and indexing practices in the past. It is increasingly common for members of our profession to play a role in developing and implementing the systems in this market.

3. Research and retrieval were the broad activities for the services librarians brought to our constituents for elevating collections to a point of usefulness in the workplace. Using our investigative methodologies and retrieval techniques to navigate and mine content from a myriad of disparate indexes and catalogs, we delivered the resources our constituents needed to get work done. Search technologies have invaded every aspect of 21st century workers domains, migrating from integrated library systems, to embedded search in virtually every software application, to the Internet free search engines. Paid search has been eclipsed in visibility by numerous free search options. However, it is important for us to educate and remind the consumers of content that much of the really valuable content is still invisible without subscriptions to value-added and peer reviewed materials. As searchers we still excel at discovery and retrieval, and our competency in evaluating content and establishing relevance is better than any rule-based system now in commercial deployment.

Murray, Phil, ed. ***Knowledge Organization - the Best Kept Secret of the 21st Century.*** 38p. The Barrington Report on Advanced Knowledge 05/w0l1e/2004.
http://www.kmconnection.com/

The recent emergence of hundreds of companies selling technology and services to address knowledge-organization activities is just one indicator of the growing awareness of the impact of KO.

NISO Framework Advisory Group. ***A Framework of Guidance for Building Good Digital Collections***. 24p. NISO 01/01/2004. http://www.niso.org/framework/framework2.html

"This Framework has two purposes:
First, to provide an overview of some of the major components and activities involved in the creation of good digital collections.

Second, to provide a framework for identifying, organizing, and applying existing knowledge and resources to support the development of sound local practices for creating and managing good digital collections."

"The Framework is organized around indicators of goodness for four core entities:
- **Collections** (organized groups of objects)
- **Objects** (digital materials)
- **Metadata** (information related to objects)
- **Projects** (initiatives to create or manage collections)"

OASIS. ***OASIS Open Document Format for Office Applications (OpenDocument) TC*** 2p. OASIS 05/23/2005. http://www.oasis-open.org/committees/office/faq.php

"OASIS Open Document Format for Office Applications (OpenDocument) is a standardized XML-based file format specification suitable for office applications. It covers the features required by text, spreadsheets, charts, and graphical documents."

"OpenDocument has its roots in the OpenOffice.org XML file format. …OpenDocument was approved as an OASIS Standard in May 2005."

Shaw, Tony, ed. ***Conference Trip Report; 2005 DAMA International Symposium & Wilshire Meta-Data Conference***, [held at] Renaissance Orlando Resort at SeaWorld, Orlando, Florida. May 22-26, 2005. 55p. Wilshire Conferences 05/01/2005. http://www.wilshireconferences.com

Silver, Bruce. ***Content in the Age of XML*** 5p. Intelligent Enterprise 06/01/2005. http://www.intelligententerprise.com/showArticle.jhtml?articleID=163100779

"Can you manage documents with the ease and automation of data? Is there a payoff in a structured approach? Will compliance demands usher in a new era? Look ahead to the future of content management....

"But for XML content to become pervasive, content must be authored in Microsoft Office and other standard tools. New versions of Office do support XML, but the feature hasn't been heavily marketed by Microsoft or widely used. Companies such as i4i have added XML-authoring plug-ins for Word, but users will still face the hurdle of following highly structured documents templates. "The capability to "autotag" Office documents and HTML based on heading styles and advanced heuristic techniques turning ordinary content into XML under the covers is well-established in the Web content management world. This capability could be brought to bear on the problem by leading content management vendors ... if they wanted to promote an XML revolution. But they don't."

Sutija, Davor Peter. ***Desperately Seeking Relevant Meta Data***, by Davor Peter Sutija with contributors: Pal Roppen, Torstein Thorsen and Bjorn Olstad of FAST. 4p. DM Direct Newsletter 09/23/2005. http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1037824

"The database administrator's frustration often starts in identifying where to look for the information one needs. The problem of data location becomes more complex when multiple information repositories are involved. These data stores often contain key pieces of information, but not the entire puzzle - the data is fractured. How much time could be saved if one had a tool that identified, connected, and offloaded relevant information into a searchable repository with authoritative and complete meta data?

"The technology needed to solve this kind of data fragmentation needs to be agnostic with regard to data format and must be capable of accessing best-of-breed systems - both enterprise knowledge management and data warehouses. One solution to this issue

of data fragmentation and accessibility can be found in the form of enterprise search technology. Enterprise search technology empowers users to intuitively find the exact information they are seeking by automatically enriching and cleansing data as it is indexed. “This article will illuminate how knowledge management systems have evolved into enterprise search platforms (ESP), and how an enterprise search can broaden access to all company data by offloading databases, thereby bringing intelligence directly to business users. “Document Management is all About the Meta Data...”

Zeller Jr., Tom. ***Beware your trail of digital fingerprints*** 2p. NY Times 11/07/2005.
http://www.nytimes.com/2005/11/07/business/07link.html?pagewanted=1&th&emc=th
“Technically, metadata is sort of the DNA of documents created with modern word-processing software. By default, it is automatically saved into the deep structure of a file, hidden from view, with information that can hint at authorship, times and dates of revisions (along with names of editors) and other tidbits that, while perhaps useful to those creating the document, might be better left unseen by the wider world.” Article goes on to describe and recommend ways to foil the capture of metadata.

Zhao, Xu. ***Metadata management maturity model*** 3p. DM Direct Newsletter 10/07/2005.
http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1038827
Ad hoc, Discovered, Managed, Optimized, and Automated models are described.

See: http://www.dmreview.com/assets/article/1038827/10.07%20zhao%20table1.gif for table format.