The practicalities of effective enterprise metadata and taxonomies: Their governance and ethics

Carol Owens

led the launch of the BBC's innovative work on metadata in 1998, after a long career in editing, operational management and internal consulting with the Corporation. Her team won the Royal Television Society Technology Innovation Award with the BBC enterprise data model SMEF™ in 2000, and Carol was made a fellow of the Society in 2001. She also chaired the European Broadcasting Union's P/Meta project to define a business-to-business content exchange scheme. Both SMEF and P/Meta are still in use. She and her team transferred to Siemens in 2004 and in March 2008 Carol moved to Serco Solutions as Head of Media Strategy to start a new sector.

Keywords: *enterprise metadata, taxonomies, governance, projects, ethics*

Abstract This paper summarizes a panel discussion held at the 2008 Henry Stewart Digital Asset Management Conference in London. It not only captures experience from people who have been involved in enterprise metadata management for many years, but also offers some novel insights from a relative newcomer to the field. The discussion covers the use and governance of metadata and taxonomies in digital asset management, the motivation of users to create the data and some ethical implications of description.

Journal of Digital Asset Management (2008) 4, 372–376. doi:10.1057/dam.2008.40

This paper summarizes a panel discussion at the 2008 Henry Stewart Digital Asset Management Conference in London, chaired by Carol Owens of Serco. The panelists were:

- Alan Irons, Project Manager, The Salvation Army;
- Judi Vernau, Information Architect, Metataxis Ltd.,
- Madi Weland Solomon, Independent Taxonomy/Metadata Consultant, Chair, AMIA Metadata Joint Subcommittee;
- Rana Ghosh-Roy, PhD, Director, Custom Software Development, Capgemini.

INTRODUCTION

Metadata is a key element of any digital asset management project, system or operation, from the small scale to the extremely large. Enterprise metadata describes the approach to information management used across complex multi-business-unit organizations or affiliations. As such, is not necessary for localized implementation, but nevertheless some of the techniques and disciplines can be shared.

For those of us who have been involved in enterprise metadata for a long time, it is

important to remember that people are still discovering the topic for the first time and need help in understanding the issues and approaches to successful implementation. For example, in the case of the panel chair's own prior experience, September 2008 marks the tenth anniversary of the launch of the BBC's Enterprise Data Model.¹ When work began in 1998 with a modest innovation project, the question was whether it was even possible to employ formal information management techniques in the media industry. It was, and is, but the discipline has to be accompanied by ceaseless user-friendly explanation and promotion in every individual project and business unit.

The value of an enterprise approach to metadata is its ability to enable data exchange and sharing between business units and systems, and also an incremental program of development. The enterprise model defines all the data in play across the organization, with its relationships, rules and local synonyms. Each business area or project will adopt a sub-set of that data relevant to its own activities, and

Carol Owens
E-mail: carol.owens9@btinternet.com

elaborate it through its own detailed analysis. The output may be used to select or develop an information system and to design interfaces to other systems and groups. Any updates revealed by the project can be assessed for their usefulness to the organization as a whole, and if appropriate, can be incorporated back into the main model.

Two or five years later, say, another project can be initiated using the enterprise model as a starting point, with confidence that its data specification will be consistent and compatible with both the main model and other compliant systems. Integration should be straightforward from a semantic point of view. The technical means of integration may mirror this, as in a Service Orientated Architecture (SOA), but continuing governance is required to maintain the enterprise model over time and so get most value from it.

Over the last decade, the BBC enterprise model has been managed and supported by a central team, now under the sponsorship of BBC Future Media & Technology. It continues to evolve as new projects work with it all the time, including the development behind the BBC iPlayer. Each project team seems to discover metadata anew, but they build on the foundation of the enterprise model and its ten-year depth of shared knowledge and insight. They can also understand the data presented by other projects and systems that they may need to connect to.

An enterprise approach to metadata may further involve management of any taxonomies in use: the sets of agreed values that end users may see in drop-down pick lists, for example. Taxonomies or cataloguing languages can be large or small, can contain 100,000 terms or just 100 (or even 10) and are useful so long as they are understood and adopted. They can be domain standards, mandated or *de facto*, as can the metadata models to which they provide reference data sets. The relationship between a taxonomy and an enterprise metadata model is that the former provides a set of standard values for attributes in the latter. For example, in the media domain an attribute “Genre” may have an agreed list of options such as Factual, News, Drama or Comedy. The list has to be agreed on and enforced to ensure consistency in data management and cataloguing.

The final two topics in the panel discussion were governance, how to ensure that the enterprise metadata approach and taxonomies are adhered to in projects and daily operations, and ethics. This was an interesting revelation for some people in the audience, raising issues around the point of view of the annotators of digital assets. How do we protect the subjects referred to in the document or media asset from later reinterpretation in a different context, which may misrepresent the original situation? This is not about intellectual property rights (IPR), but verges on libel: yet another challenge for metadata managers to address.

ENTERPRISE METADATA

In our panel session, only a quarter of the audience identified themselves as using an enterprise approach to metadata. The issue of organizational scale is one reason, but the span of common concepts and data understanding often needs to go well beyond immediate departmental boundaries into upstream and downstream areas, and also into a virtual enterprise of partners, suppliers and customers. All parties need to be clear about what the information they are exchanging means, and when applications or web portals are created with wide access across boundaries, a common underlying semantic is crucial even if the user interfaces are configured into local terms and friendly layout.

The resisting factors to enterprise thinking may be silo-mentality, localized funding arrangements that work against collaboration, or just a pragmatic recognition that some organizational entities, such as government departments, will always be different. Judi Vernau, who is an expert on government e-standards, finds that so long as the concepts are the same, the actual terms can be different and can be related as synonyms. In SOA terms, everyone can speak their own language at home, but employ a lingua franca on the Enterprise Service Bus or Model.

Another of our panel members, Alan Irons of The Salvation Army, had project managed a purely local project to digitize still photographs, but believes that International Headquarters have expressed an interest in a worldwide approach. He could therefore see the value of an

enterprise approach, and also a standard taxonomy for the charity's specialist terminology so it could be translated into different languages. For example, The Salvation Army has a role of "Songster," which is specific to it and wider than just a chorus member. This term exists in each country in the local language, and somewhere there needs to be a repository of every version so that HQ can track its usage. Alan is a convert to a formalized approach and can see the benefit from his own experience.

This is refreshing for information management professionals like Rana Ghosh-Roy of Capgemini in the face of frequent resistance to what appears to some practical people to be a highly abstract and academic discipline, for which one needs a PhD. This scepticism about enterprise metadata is a Luddite response, but the defense must be to explain the issues in plain language, and demonstrate the benefit through effective delivery of solutions that work and connect.

There is also a choice to be made between the enterprise approach as a *post hoc* mapping activity and as an active design authority. This will depend on the nature of the organization, and the level to which it is possible to compel or persuade staff to adopt standards and create good metadata. At the BBC, it is very much an active principle, with even the Director General mentioning metadata in his speeches. No one can ignore metadata any more, but there is still the challenge of getting people to create it. Enterprise standards can be imposed at system design level, but communication and perhaps incentives will be needed to motivate human data entry. The best argument may be "if you want to find your content again, you'd better call it something sensible that everyone can understand."

TAXONOMIES

Taxonomies are familiar to professional librarians and cataloguers, and exist too in specialized fields such as natural history. The challenge for digital asset management professionals is to persuade their sponsors and users to agree to the definition of standard lists, and then implement them. It is particularly important in the visual arts, whether static images or audio-visual content, as the assets are not self-describing.

Even with image-recognition software, it is not yet possible to type in "polar bear" and find all the content without invoking some textual tagging.

To people new to metadata and cataloguing, the notion of a taxonomy is strange and often counter-culture. The staff describing photographs for The Salvation Army project were encouraged to look at them from a searcher's perspective, and it would have been useful to have access to records of real requests to understand the common concepts and phrases in use. This could have provided the basis of a simple taxonomy to assist both description and retrieval.

Professional librarians in media and publishing are totally happy with highly complex and extensive taxonomies, but for ordinary users, the formal terms may need to be fronted by a thesaurus function that translates plain language into "legal" terms. These taxonomies are intended for description and retrieval and to facilitate reuse.

Judi spoke about taxonomies that have been developed to describe business concepts and attributes primarily for search purposes, to help people navigate company intranets and files. Various vocabularies have been published, which may be useful, but should always be tailored to fit the individual business. The objective, as ever, should be to assist understanding and retrieval, otherwise the digital assets are not assets: if they are impossible to find they have no value.

GOVERNANCE

Implementing any metadata project or standard requires good governance to be in place in order to ensure delivery and also continuing adherence. Oddly, it seems that there is still resistance at senior management level in some organizations to supporting and funding an enterprise approach or even a metadata standard, as they do not have to deal with the outcomes of bad governance. This makes life hard for the metadata professionals and business areas that depend on good data.

Once a project has been approved, it is vital to talk to all stakeholders in advance and to understand their objectives, culture and constraints. It is also important to have a project manager who understands the need for metadata definition and builds it into the project plan.

Too often, it is seen as an offshoot of Business Analysis, and is not supported with enough time or the appropriate tools. It should be a clearly defined workstream within the plan, and its outputs should be used to support the delivery lifecycle into development, testing, data migration and implementation.

Rana Ghosh-Roy advised taking a BOSS approach to metadata definition, covering data to support business, operations, systems and security. He also recommended identifying business champions and using analysts trained in target system capabilities.

After the project has completed and the system has gone into business-as-usual service, governance needs to continue to ensure that data is entered and maintained correctly. Madi Weland Solomon emphasized the need to identify leaders to “steward” performance and ensure that the terminology fits the user context. It is very important to have the correct taxonomy and for data entry to be easy and clear.

Even in telecommunications companies where staff are more technical, a system with 70 metadata fields may only get three or four of them populated: enterprise search has to rely on synonyms. An audience member suggested rewarding annotators for good metadata and maintaining a constant marketing-style communications program to remind people what metadata is for. In government circles, data compliance is built into peoples’ role specifications: the stick approach, not the carrot.

Governance at the enterprise level requires the skill of a diplomat, in Madi’s view. Change is hard and sometimes money is the only way of getting things changed. Making projects and business areas to work together requires active design and sometimes pure bribery. The best argument for an enterprise approach is to demonstrate real value. A member of the audience cited an example of a company that used enterprise metadata management to reduce its month-end accounting period close from ten days to three. That is an obvious benefit.

ETHICS

Metadata is commonly used in the identification and protection of intellectual property (IP), but

its role in protecting personal privacy should not be overlooked. Alan Irons explained that The Salvation Army photographs of people used in their campaigns had to be carefully managed, to ensure that they are not used inappropriately in the future or in a way which might cause distress to individuals or their families. This presents a different moral perspective on protection than simply IPR and payments.

Judi observed that the Freedom of Information Act had woken up information holders to the need to set rules about what should be accessed, and to what level of detail. The requirement to respond within 20 days has focussed attention on the need to find and retrieve data quickly and accurately, down to parts of individual documents.

The perspective of the person doing the annotation is important, in the way they interpret the media or document. For example, a member of the public may send in some user generated content of people at a festival, tagging it “Glastonbury” or similar. A television news or current affairs producer may use it in an item about drug use, which is how it will be catalogued for the archive. This raises an interesting ethical point about misrepresentation, which may be accidental but could nonetheless lead to distress and litigation.

The question of the bias of the metadata creator from a political or moral standpoint was a new one to most people, but in a world where content is ubiquitous and virtually uncontrollable (in spite of IP protection), it may become an increasing problem. In managed environments it may be enough to maintain an audit trail of annotation updates, but in the world of the web and folksonomies, anything can happen.

KEY LESSONS FROM THE PANEL MEMBERS

Finally, the panel speakers gave their top tips for successful enterprise metadata implementation and management.

- Alan admitted that he had been naïve at the start of the project, but had quickly learned that one had to fully understand what was needed in

practical terms, to learn from mistakes and to make sure of real buy-in.

- *Judi* recommended being as specific as possible about metadata creation. Guidance is not enough: all elements need to be defined clearly and without ambiguity.
- *Madi* warned people not to underestimate the impact of migration. Technology changes every few years, so a plan is needed to protect and preserve the archive. Metadata must be seen as a company asset, a machine serving the business.
- *Rana* urged listeners to enter the metadata definition as a deliverable item within a project plan, with its own start date, end date and resources. As senior managers often focus on deliverables, any problems would be

escalated immediately, as the impact would be significant.

Metadata is for life, not just for projects.

Acknowledgments

I would like to thank the panel members and the audience for their contribution and interest, and the organizers of the Henry Stewart Digital Asset Management Conference for the opportunity to chair this session.

Note

- 1 The BBC Standard Media Exchange Framework or SMEF™, now known internally as the Logical Enterprise Data Model or LEDM. It can be accessed via: www.bbc.co.uk/guidelines/smef.