Conquering chaos via smart content management — Interview with Seth Earley of Earley & Associates

Seth Earley

is a recognized thought leader in the areas of content management, digital asset management, search, taxonomy and metadata. He has been implementing content management and knowledge management projects for over 14 years and has been in the technology field for 20+ years. He hosts a series of conference calls and forums on search, content strategy and digital, is founder of the Boston Knowledge Management Forum and co-author of *Practical Knowledge Management* from the IBM Press. He is a former adjunct professor at Northeastern University, where he taught graduate courses in E Business Strategy. Seth has developed search, content and knowledge strategies for global organizations and has developed underlying taxonomies for a diverse roster of Fortune 1000 companies. He is a popular speaker and workshop leader at conferences throughout North America, speaking on intranet design, knowledge management, content management systems and strategy, taxonomy development and other related topics.

Keywords: *content management, taxonomy, metadata, metadata standards, digital asset management, search strategy*

Abstract Managing content — whether documents, transactional data or digital assets — is about providing *content in context*. Users can't find what they need for three reasons: (1) It's the nature of the beast — information and systems evolve and tend toward a disordered state. We need to expend energy to clean things up and organize them from time to time.

(2) In most organizations governance processes around content management, digital asset management, search, taxonomy and metadata are immature. Therefore, there is no clear ownership of content and repositories.

(3) Content is not "selectively managed," that is, little consideration is given to the continuum of content value. People lump collaborative workspaces (which are messy and chaotic) into the same problem as document and content management (which should be a more orderly process around capturing and reusing high value finished content). These are two ends of the spectrum.

In this interview we discuss a number of issues around content management, taxonomy, tagging, metadata and search, and provide some ideas on how to tackle the chaos to create business value.

Journal of Digital Asset Management (2008) 4, 318–330. doi:10.1057/dam.2008.16

MM: We're here with Seth Earley of Earley & Associates. Seth, would you give us a professional background in terms of your involvement in all things DAM?

SE: Do you want my professional background? You don't want my sordid history?

MM: Either will do.

SE: I've been in technology for over 20 years. I've been building content and knowledge and digital asset management systems for the past 14 years. I've focused a lot on content and knowledge processes in the context of metadata standards and taxonomy development. Our

initial focus was on IBM tools and we have evolved into doing more design and strategy than technical implementation.

Much of our work is helping organizations sort out their content and DAM projects and problems. We recently did a project for a global publishing organization with over 100 imprints. They had a very diverse organization where we worked with divisions and groups on how to implement metadata standards. We helped to develop their strategy around leveraging evolving standards to manage content and assets across processes and business units.

Seth Earley
Principal, Earley &
Associates, P O Box 308,
Concord, MA 01742, USA
Tel: 781 444 0287
E-mail: seth@earley.com

We work with a variety of organizations. Pharmaceutical and life sciences, manufacturing and technology, telecommunications, financial services and the public sector. For instance, we're doing a large project for the Department of Labor on taxonomy, information architecture, metadata standards and search strategy.

We see very diverse types of customers and problems, but all have the same core problems around organizing content for specific audiences and processes. This entails content management strategy development and establishing metadata standards and enterprise taxonomies, and then coming up with tactical plans to operationalize this to meet business goals.

We've been doing this for about 14 years now. We have an 18-person consulting firm based here in Boston, Massachusetts. Our consultants are dispersed throughout the US, and Canada and the UK.

MM: Would you start with an overview of how you see the state of the industry for DAM and content management? What are some of the key trends?

SE: That's a great question.

Having been in the industry for quite a while, I've seen an evolution of maturity around content and knowledge processes. Back in the old days when we started, it was before internet applications were really popular. It was at the beginning of that phase.

We used to do a lot of this work with Lotus technologies. At the time, people were trying to get their minds around managing content and knowledge assets, and managing the things that they thought were valuable in the organization.

The industry went through lots of mistakes while getting through the learning curve. There was a huge proliferation of applications with inconsistent organizing principles, fragmented islands of knowledge and disconnected processes. Organizations did the best they could, but tools were immature and approaches were immature.

We used to say at that time that Lotus Notes did poorly what nothing else could do at all. That was the state of the industry. The technology wasn't there. People had a desire to start managing intellectual assets and support collaboration. IT departments were trying to get their minds around content and images and other unstructured information. This was very

different than what they were traditionally charged with — the hard, structured, transactional data.

The industry went through an evolution — actually a transformation with internet technologies — and in a way, the maturity curve started again.

It seemed that some of the lessons learned around internally focused technologies — the client-server technologies — was lost, as a new group of people were called on to try to make sense of the web. A lot that was learned about application design and development, organizing principles and usability was lost with web technologies. This is the way it had to be. We were coming at the problem of information management from a new perspective and that did mean throwing out many of the rules. Unfortunately we threw out some of the good rules as well.

When I taught application development, most of our classes were comprised of non-technical people — administrators, business people and so on.

After learning how to make better use of those collaborative technologies, organizations threw much of that out and there was now a whole new technology shift with web technology. That started off with static pages and lists of links with no clear organizing principles. A lot of organizations got stuck there.

You can still go into companies these days and find that they're not managing their content and they're not managing their assets. They're doing things like using file servers and having folders labeled, "Important Stuff," or "Joe's Assets," or whatever it might be.

There are certainly pockets of maturity in this very broad mix. Market conditions have caused some organizations to move faster through the learning curve. So there are lots of organizations that are still at the earlier stages and are immature, and other organizations that have been forced into maturity.

Something that comes to mind is newspaper publishers. If you look at the lifecycle of information for newspapers, it's very, very short. A news organization has a fast "clock speed." Those guys have had to learn to become much more mature.

Other types of industries, maybe not so much. There are some financial services industries that are very good at managing structured data because that's what their business is about. But they're not very good at managing unstructured data.

The same thing goes for organizations that are managing their marketing initiatives. Creative processes have traditionally been unstructured, ad hoc and collaborative. Creative people are very resistant to adding structure in workflow and process and organizing principles.

There is an overall movement along a learning curve. We're seeing a general improvement in maturity levels. But it's really spotty in some industries. It depends upon the business model. It depends upon the type of application. It depends upon the technology.

Some organizations are doing well by moving along this learning curve and reaping the benefits. Other organizations are very immature because the business drivers have not been there to push them along.

We're starting to see bigger and bigger shifts in the industry, where there's not an opportunity to gradually learn. There are sudden big, big shifts in how people are either delivering or consuming information or repurposing information that is forcing industries that perhaps were a little bit slower to learn, to learn a lot faster. We're seeing a lot of changes in the industry, and a lot faster acceleration along the learning curve.

MM: I'd like to run a scenario by you. I think you'll be able to respond with really concrete prescriptives.

I'd like to start with the idea that markets have fundamentally undergone and will continue to undergo a structural change, as characterized by what we can now call self-directed, highly social networked customers.

As a function of that, people today don't simply wait in their car or their living rooms to have markets market to them. They in fact are going out and engaging brands and markets on their own.

In fact, marketers now must think in terms of an engagement of theater, where these self-directed customers show up at various levels of knowledge, process maturities, sophistication, buying process.

SE: Right.

MM: These engagement theaters — almost by definition — need to have a lot of different methods of engagement.

We think about a digital supply or value chain working from the brand space of these self-directed consumers that use — increasingly — social media and web rating things, by which to understand what to buy and how to buy it. They show up in an engagement theater, which you could characterize as information architecture, navigation, brand, search, user-experience.

Then if you dip into that a little deeper, you find a content-optimization group. And a content-optimization group thinks in terms of what I call contextual consumption.

SE: Exactly. Yes.

MM: As we have these various types of stakeholders engaging, how do we optimize content for some number of meaningful interactions and engagements?

SE: Yes.

MM: So, content optimization is really about tagging content and creating many-to-many relationships of reusable content parts — or what I call “intelligent content subassemblies.”

SE: Yes. Absolutely.

MM: Intelligent content subassemblies usually now live in some sort of an XML database. So there's a fairly sophisticated text-mining and classification of these subassemblies into what some people refer to as “ontologies.” And logical collections of “stuff.”

That then results from what I'll call a “content refinery,” where you're now creating those subassemblies from your publishing workflows. So if you're a magazine publisher, your traditional workflow has been K4 or Quark Express. So how do I take those pages and now deconstruct them into intelligent reusable subassemblies? So there's a whole refinery part.

SE: Yes.

MM: Hence, the content refinery.

What you're really doing is normalizing these little subassemblies down to their lowest level of reusability.

SE: Sure.

MM: Usually it's at that point in time you're also applying management information and rights information. That then gets all the way upstream to the content operations, or the creative composition or the editorial operations.

Increasingly, editorial operations have to do with user-generated content — including user-generated video. So we think we have this multi-channel, multi-media editorial operations that feeds the content refinery. That's a way of breaking it all down into intelligent subassemblies. Then content-optimization, which is really packaging it for contextual consumption. That then provisions that to an engagement theater, which is the user-interface, user-experience, navigation, sophisticated search. That's ultimately about providing meaningful points of engagement to consumers as they come in from brand space. That's our concept of a digital supply chain.

Can you share with us some of the things that you've found to be essential and critical toward making that vision work?

SE: That's a great question. I think you answered a lot of it in the question. So much of what we talk about is "context" and "content": understanding the *specific* need of a *specific* audience in the context of *their tasks* or *their objectives*. Part of that is really understanding and characterizing the audiences. Who are they? What's their experience? What's their background? Things like personas are very helpful in doing that.

Understanding exactly what their task is, and what the steps to their task are, and what their line of thinking to their task might be. We need to anticipate what it is they're looking for and the decision-making process as they're purchasing a product, or consuming a service or finding an answer. We need to be able to present the content they require — or give them a path to the content as they require it.

As you mentioned, there's a continual fragmentation of attention. There's a continual fragmentation of delivery mechanisms and consumption mechanisms. People spend much less time in any one medium and any one channel and any one area. Many times they use their social reference groups in order to help guide them.

You mentioned social media and you mentioned user-generated content. When I look for a restaurant in a new city, I'll look for the reviews. I'll look for what other people say about this place. I'll try to relate something to those particular user reviews so that I can say, "OK, this sounds reasonable to me."

I think we're going to start to see a greater use of things like social network analysis in conjunction with content processes. But the biggest piece of this is to have consistent mechanisms for parsing the content, and creating the content object models that will make sense. As you say, content has to be in much more atomic and granular forms. We refer to that as micro-content, and ways to use that content as a micro-content strategy. So that we can say, "Let's take a piece of content that may be traditionally consumed in one mechanism, and start parsing it out and breaking it down so that I can segment it and present it to consumers maybe in a slightly different mechanism." Or when they're engaging in a different task. Or when they're using a different vehicle for consuming it.

Maybe they're consuming things on their mobile phones. Or maybe they're doing it through a news feed. Or maybe it's through some type of a group discussion and so on.

The key to this is to have a process. First, understand that audience and understand what they want. We can't tag every piece of content for every possible purpose in advance. We don't know what that will be. We have to have some understanding of the audience, what they want, and what they're trying to accomplish. Then we can start to build the content object models that will allow for this.

One of the challenges that organizations have is they don't know what those models will be. They don't know how consumers are going to be consuming content. It's very difficult to come up with the metadata standards and strategies and the taxonomies and the content-object models and the workflows that will allow them to take this material and break it up in new ways for new mechanisms for consuming it when those mechanisms and models are not yet defined.

MM: Seth, you bring up a point that I'd like to exploit here. As we start thinking about contextual consumption, that then entails developing an explicit model, which you characterize as a persona or goal or mode — a decision-making process or task, and therefore, content requirements. Right?

SE: Yes.

MM: In the course of developing context for consumption, inevitably you begin reverse

engineering, desire and intent of consumers. In the course of mapping desire, awareness, understanding of consumers, you inevitably start to reverse-engineer psychology and psychological (gifts) from not, “What content do we have or need,” but, “What do people need as human beings?”

Doesn't this then suggest that the underlying principle of contextual consumption is a much more robust, humanistic model of need, desire and aspiration? And isn't in fact what we're doing recreating the major passages of life — from adolescence to elder years — and then understanding the psychological drivers or the psychological impulses that drive consumption?

SE: Right. Yes. I think that makes a lot of sense.

That comes back to building out representations of audiences. We just went through this with a large project where we surveyed 13,000 people. We did focus groups and contextual interviews. We segmented these different audiences to understand their drivers and what was important to them given their goals and their life stage.

When creating personas you are trying to pull all of those factors together. The persona is basically saying, “Who is this individual? What do they care about? What's important to them? What's their experience? In what stage of their life are they? What are their aspirations? What are they trying to accomplish?” From that, you can start to say, “Well, this is how I believe this person is going to make decisions. This is how they're going to think about things. These are their values.” This helps to infer and build out content needs for various tasks.

But I think you're right. You're building out a psychological profile of people. You're building it out based on their lifestyle and life stage. Their experience, their culture, education and so on. All of those things are incredibly important when we try to characterize an audience, and then build some mechanism for delivering what they want, when they want it — and trying to anticipate what needs they will have as they traverse content, consume services or purchase products.

I think you've hit it spot on.

MM: That would suggest — and this is the classic conversation that I encountered 20 years ago in artificial intelligence and expert systems. That specifically is, “Are we going to have an

engineer attempt to model consciousness?” Or, “Are we going to have philosophers and metaphysicians give a logical structured framework for consciousness.” Right?

Of course, Hubert Dreyfus — the professor at UC Berkeley — wrote the great book on “Why Computers Can't.” That was an utterly devastating, authoritative, conclusive destruction of the idea of artificial intelligence.

SE: Sure. Yes.

MM: The point here is, are you using any particular models or modeling — psychometrics or connotive capabilities? Like the Myers-Briggs test? Or Jungian Archetypes? Are you using any kind of established persona methodologies?

SE: No, but those are some interesting areas for research and exploration — perhaps doing a Myers-Briggs on your persona — or correlating that with user personality profiles. So you would know your persona was an “ENTP” — Extraverted, iNtuitive, Thinker, Perceiving, and so on. This would be an interesting approach.

The challenge is that many organizations don't even build out basic personas.

There are some organizations that want to understand in-depth consumer behaviors and understand the psychology behind decision-making. That could be mapped back to mental models and content needs. The mental model could guide development of the engagement theater — the various ways to reach that consumer.

There's a lot of room in the industry to begin to do this, and to get into even greater depth for audience characterization and analysis in building those mental models.

MM: This really leads into another conversation in terms of our engagement theater and what's going on in brand space. An important distinction that came from the work of Adrian Slywotsky of Harvard Business School, and then later Mercer Management Consulting, was the distinction of a share-determining market sector.

He makes the case that generally in large markets you have a group of customers that are your most profitable. You have a group of customers that are your most loyal. Sometimes there's a heavy overlap and sometimes there's not. Then there's a small, generally overlooked portion of your customers that, while small, determine tomorrow's share of market.

For example, for consumer electronics, the share-determining market sector tends to be the bedrooms of teenage daughters. And it's no surprise that the iMac and then later the iPod really kind of own the teenage girls' rooms. As these teenagers age each year, they take with them the brand loyalty that ultimately disrupts an entire market.

Another example of that would be lifelong habituated use of tobacco products. It begins with 14-year old boys. The whole point of the Marlboro Man was to position an aspirational character — the Marlboro Man — to a 14-year old boy who is dealing with that, needless to say, highly conflicted psychological state of living in his father's house by his mother's rules.

So the Marlboro Man is really, "I am independent. I can do what I want." Therefore, smoking Marlboros is a demonstration of that aspiration. Right?

So the idea of creating and understanding where your new customers come into a market — and really creating special interactions with them — are especially important. It seems to me that publishers — magazines, newspapers and such — have a particular challenge today. The kids today seem to consume media in a fundamentally different way. It tends to be multi-task. And they tend not to subscribe to magazines and newspapers anywhere near as they used to in the past.

From a content strategy and from a micro-content strategy, what have been some of the things that you've seen in the publishing industry to really engage the youth market?

SE: Well, that's a very challenging market. I think one of the things that people have realized in trying to engage that market is that they're much more savvy and sophisticated and cynical. If cynical began with an "S," you could say the "Three Ss." Unfortunately, it's SSC — Savvy, Sophisticated and Cynical. It still has a nice ring to it.

They don't take well to marketing pitches. They don't take well to sales pitches. They are very aware of being manipulated.

Some of the things that are working are, again, reference and peer groups. Viral types of efforts through social networks. Humor works well with them. Irony, wit, honesty. Being clever and direct — not manipulative.

You can't simply sell that market. You can't simply market to them. You have to engage them in a way that they find honest and sincere and informative. They want to learn something, but not in a preachy way, not in a lecturing way.

They're inquisitive, and they want to learn from their peers. They want to know what's going on. I think that it just changes the message. It changes the medium. It changes how marketers are going about trying to reach that audience.

Again, they're savvy, sophisticated and cynical. They're much farther ahead in their years and their maturity than previous generations. They're also able to consume information much, much more quickly. Their clock speed is faster and their attention spans shorter. They can multi-task in real time.

My 13-year old stepdaughter might have several instant message windows going — having conversations with lots of people — while she's watching television and doing her homework and watching an online video. How much she actually processes might be another story. But people have never even had the opportunity to deal with that kind of information overload until now. It's astounding how much information children are learning to consume at earlier and earlier ages.

That just means that the information experience is more fragmented. You have to reach them from more channels. The network of social connections is where they want to spend their time. They don't want to waste time, and want to trust their sources. The way they trust their sources is by looking at their reference groups.

I think it's a very challenging perspective for marketers to engage.

MM: Let's use this to shift back into the digital operation of publishers and marketing organizations. In particular, I'd like you to address the development of faceted taxonomies and things related to that — be they a thesaurus or organizational ontologies.

Before we go into that, could you give us a working definition of those three disciplines?

SE: Sure.

First of all, taxonomy is simply a system for organizing information. It can be a traditional hierarchical type of arrangement of subjects and topics. It can be a whole/part type of

arrangement. Many times people misinterpret the idea of taxonomy. They think of any classification structure being a taxonomy.

The one thing I like to say to people is that a taxonomy is not the same as navigation. You can have different navigational constructs that leverage classification. Without getting into a long explanation of that, I like to use the example of something that we built for a sales organization.

We had something called “Sales Tools.” It was a place in their navigational hierarchy. Under that, there were white papers, research reports, presentations, specifications and so on. I always like to ask the question, “Where in our taxonomy do white papers, presentations and specifications live?” Those do not live in anything called “Sales Tools.”

“Sales Tools,” was the name of a node that was a navigational construct. “Sales Tools” does not live in our taxonomy. All of those things I just mentioned were doc types. That’s my example of saying that a taxonomy is not the same as navigation. Navigation is context and audience specific. It is also a representation of an index — not a taxonomy. A taxonomy is a system for classification; an index is a content access structure. It is specific to a body of content. A taxonomy is reusable; an index is not. You can’t take a back-of-the-book index and drop it into another book.

A taxonomy can *inform* and *influence* navigation, but it’s not the same. Most people don’t make that distinction. They just think of taxonomy as a navigational hierarchy.

When we think about a faceted taxonomy, we’re suggesting that there are lots of different ways to classify information. I could have doc type, which is one of the facets. I could have “audience,” which is another facet. I could have “industry,” which might be another facet.

Then I can apply terms from those facets to content. I can apply metadata from those facets and present that as multiple navigational constructs. A faceted classification can present different ways of organizing content for different audiences.

MM: Seth — in that context — a facet really represents one logical point of view for how they see and categorize a bunch of stuff.

SE: Yes.

MM: A way of looking at things in a repository is doc types.

SE: Yes.

MM: But there might be another way of speaking of it in terms of subject matters. Right? Then another might be in terms of appropriateness in a buying process.

SE: Yes. That’s right.

MM: Presale, close of sale.

SE: Yes.

MM: So facets represent a logical collection of names from a particular user point of view or an engagement perspective. Is that a fair characterization?

SE: That is. But it’s even more interesting than that. That’s looking at it in a singular perspective. But we can also look at it as an intersection of points of view and engagement perspectives.

If I want to look at the specific white papers for this product and this industry, now I’m using those three coordinates — those three facets — to zero in. You can think of it as navigating a three-dimensional space. One coordinate is the doc type. Another coordinate is the industry. And the other coordinate could be the product or the application.

You can get to a very precise location of information by leveraging those facets. But it’s not limited to three. It’s as many dimensions as you like — as many metadata fields as is feasible. You can apply several facets to content and be able to navigate extremely precisely.

This is where we start to get into guided navigation or “faceted search.” Go to the PC connection websites when you want to look at things according to a certain type of computer. Desktop versus server versus laptop. You want to look at a certain monitor size. You want to look at a certain speed processor. A certain amount of RAM. You’re navigating in an N-dimensional space, to very precise information.

We can do the same thing with any kind of asset and any kind of information. This is where we’re starting to see very, very successful companies like Endeca — that’s one particular vendor who has really capitalized on this. It’s giving people new power to navigate to that content. But it also means we have to tag and organize that content with the right metadata. It means we have to have controlled vocabularies.

It means we have to have metadata standards and taxonomies.

That's one aspect.

The other question you had was the thesaurus. What is a thesaurus? Well, a thesaurus is a taxonomy on steroids. A thesaurus has all of the hierarchical relationships (whole/part and parent/child). But it also has related terms or "*associative relationships*" and synonyms or "*equivalence relationships*."

Vermont is also called the Green Mountain State. That is an "equivalent" relationship or synonym. If my search engine leverages synonyms, when I do a search on the Green Mountain State I'll pull up information on Vermont.

I also know that Montpelier is a place in Vermont. That's a whole/part relationship. So there's a parent-child or a hierarchical relationship.

What about "maple syrup?" "Well, that's not another name for Vermont. And it's not a place in Vermont. But it's somehow conceptually related to Vermont. This is an associative relationship." It's a "see also" term. Associative relationship types in thesaurus structures give us the power to provide *content in context*.

We can say, "Here is an occupation that I'm interested in if I'm looking for a job. But I want to know what certifications go with that occupation." Or, "Here are certifications. I want to know what training courses give me that certification." Or "I'm an insurance processor and I'm processing a claim. I want to know what associated policies go with this particular claim." And so on.

Associative relationships can present related products. This can be used for cross-selling or up-selling. There are endless possibilities when presenting thesaurus structures, and leveraging thesaurus structures in content-management applications and search applications.

MM: So the thesaurus really becomes the framework for identifying and modeling contextual consumption.

SE: Absolutely. 100 per cent. You have content and a process. You can say, "I'm a consultant. These are the steps of my process. I first have to research opportunities or I have to look into the solutions for the customer."

Well, what information do I need at that process step? It's content related to a process. We

can have a process taxonomy and a content taxonomy and relate the two of them together. That's what's giving us the context for content.

If I'm doing a search on a website, I could be looking for anything. If I know something about that user, I can say, "Oh. You may also want to think about this. Here's something you might be interested in." That's from a thesaurus type of relationship.

You also asked about ontologies. Essentially, an ontology represents a domain of knowledge. It's all the collections of taxonomies and thesaurus structures and all of those relationships between them.

You can come up with any relationship type. You can come up with any associative relationship type that you'd like between any two terms. If you can come up with a concept between them, that's your associative relationship. There are limitless ways to describe knowledge domains and concept relationships.

I can relate anything to anything contextually. Again, that's the ontology. Ontology says, "How do we represent ideas and processes to one another?" If it's an insurance company, it's products and markets, risk areas and regions, policies and claims — all those different things. If it's a pharmaceutical company, it's drugs and brands, chemicals, interactions, diseases, biochemical pathways, proteins and so on.

We can look at any domain of knowledge and say, "What are all the facets?" "What are all the controlled vocabularies?" I can use those to zero in on content very precisely, but then I can also present related content for whatever my task is — for whatever my audience is. For whatever my process is.

That's the exciting thing. I think that's where organizations have not really leveraged this kind of power. That's where I think you're going to start to see more powerful capabilities, as people begin to understand semantic relationships and how you can manage those outside of specific bits of content so that they can be leveraged across multiple systems and content repositories.

Right now you can do this in content systems. You can take a piece of content and say, "I'm going to tag this piece of content with this related product." But then if that changes, that's very focused on that one asset.

If you can abstract from that and say, "I'm going to manage my semantic relationships

externally,” and then feed that to multiple systems, or feed that to multiple search engines and be able to change that as my business conditions change — that’s going to be some tremendous power.

The organizations that learn how to manage that will have a very strong competitive advantage in any marketplace. Search engines are beginning to do this, but most are missing the point of context specific to a domain of knowledge — they try to make this work for all domains. In order to truly make it work you need to map semantic relationships based on your information, processes, business model, user audiences, etc.

MM: Seth, to that end, what systems are people using to externally manage those ontologies and taxonomies?

SE: I have to say that, in full disclosure, there’s a sister company of ours called Wordmap. They actually do that. That’s full-disclosure commercial interest there.

There are also ontology editors. I’m trying to think of a name. There are some open source tools. There are tools from a company called “Access Innovations”; their product is called “Data Harmony.” There are tools from a Dow Jones company which is Factiva. There are some of the large metadata management vendors, such as ASG, that are beginning to get into that space. It is a part of metadata management, but there are two ends of the spectrum on metadata management.

There is highly structured metadata and data modeling. That’s very much on the technical side. Then on the other side, there’s the semantic perspective — semantic management. It’s the business side of metadata

The challenge is that on the structured modeling side, people tend to think of things in more black and white terms with little ambiguity. I have heard some data modelers say, “this just a metadata management problem. It’s simple.” Just give me the dropdowns, the list of values, the reference data. But it’s not that straightforward because reference data — the terms you see in drop down boxes on web forms — can be difficult to derive or may change very quickly. The data model itself can be concrete, but the terms to populate it can be variable, changing and ambiguous.

For example, the values of stock number or price or name and address can be defined and

are unambiguous. There can certainly be problems with data consistency and quality, but for the most part what the values *should* be is clear. That’s relatively easy to model.

But what about solution? What’s a solution? Is a solution a product? Is it a service? Is it a combination of a product and a service?

At Motorola, where we’ve done work for the last three years, we’re wrestling with that all the time. How do we define these things? These are based on market interpretation. These are based on customer needs. These are based on changes in technology. So there’s a lot more interpretation and evolution and change in that area.

MM: It seems, Seth, that you’ve set up a nice conceptual two-dimensional model here. The baseline would be from ambiguous data to unambiguous data. Right?

SE: Right.

MM: Then the vertical axes would be static and dynamic. Stuff that’s always changing.

SE: Yes.

MM: As we look at that quadrant of ambiguous/unambiguous static and dynamic, it seems to me that you’d need a different set of practice tools and accountabilities for managing various aspects of that metadata landscape.

SE: That’s very true. And you need different people to consider that. There are different sets of problems. You’re right. You can think of it as structure and lack of structure, chaos and control. There’s a bunch of different ways to think about that.

They become the purview of different parts of the organization. Again, there’s the structured, unambiguous data which could be changing very quickly. Transaction processes are very fast moving but unambiguous. It’s definable, and there are mechanisms for modeling and managing that information. If you consider folksonomies and social tagging — those are very ambiguous and dynamic, and subject to interpretation. That type of metadata is also typically applied to lower-valued content, whereas structured metadata is applied to higher value information.

A pharmaceutical company might have an FDA-validated process. That’s a very, very strict process. Your drug cannot go to market unless you meet that certain well-documented guidelines and processes. Documentation needs

to be tightly controlled and well managed. In that case it would not be a good idea to allow free-form tagging of the content. You would not want to use a folksonomy. You would not want to use social tagging and uncontrolled vocabularies. The process and documentation need to be controlled and managed. You would never just let anybody tag anything with anything. It's very locked down in terms of your workflow and in terms of tagging and business process — controlled vocabularies, roles, definitions and security.

On the other hand, a discussion forum or a blog or wiki does not contain as much mission-critical information. It doesn't matter as much what people do there. It's not as strategically important.

If the goal was to organize best practice documentation or example case studies, these types of documents represent methods and approaches that embody the intellectual value of the organization. That kind of high value content requires a review process, a vetting process, editing, and approval and so on. Higher value content justifies the use of more structure.

Some people think, "Well, why do you need taxonomies? And why do you need tagging? We're just going to get Google or Thunderstone. We're just going to get a search appliance and plug it in."

It's ridiculous to say that because you're not thinking about the continuum of value of information. And it would be the silliest thing that, "Well, you know what — forget our account data. We're not going to use account numbers anymore. We're just going to use a search engine." It's ridiculous to say you don't use structure and you won't do any kind of tagging with metadata. Search engines are getting better, and they work by deriving metadata about content. But they cannot infer intent and they do not know what is important to you. Search appliances are not very good at distinguishing fine-grained differences of information in narrow domains.

Again, it's a continuum of processes. And I do like your idea of the static dynamic — unambiguous and ambiguous.

MM: But more to the point here, it seems to me that as we create this metadata landscape — for lack of a better term — using this two-dimensional model — it seems to me you'll

have zones within that model for which you have a very specific investment requirement and payback.

SE: Yes.

MM: In order to bring governance of a certain level of transparency and repeatability to this particular area of stuff, it will require investment in this text-mining thing or in this auto-classification thing. And it will produce these kinds of gains in terms of productivity, time-to-market and increased sales.

SE: Right.

MM: Yes. Fabulous. This sets up another question that I wanted to have, if we follow on to this. That's the notion of the cost of developing metadata, and the payback of it. Can you give us some frameworks for understanding the cost to develop faceted taxonomies, navigational hierarchies, thesauri and then ultimately process taxonomy and context taxonomies that then support an ontology.

SE: Sure.

MM: Let's stipulate that the business model of this customer's ideal place to start relies heavily on search optimization and federated search across some multiple number of web properties.

What's the best place to start really developing organizational ontologies with process taxonomies, context taxonomies, and therefore really smart navigational hierarchy? As opposed to, "We'll get around to it."

SE: It has to be driven by something that's customer-facing. If you get too far away or get too abstract and say, "We're going to do this across the entire organization," but you're not looking at specific applications that you can influence, then that's not going to be a good place to start.

At the same time, you do need to take an enterprise view. Because the first thing you're creating is a domain model that says, "What are the overall organizing principles, so that I don't leave anything out?"

If you start too focused on just an application, you might limit future options or not allow for adaptability. The balance has to be correct — if you start too broadly, you're not going to have the applicability, too narrow and you lose adaptability. It's also important to choose something that can be measured, such as a customer support system. You can measure direct impact on things like time-per-incident or

support call deflection. You can measure customer satisfaction. You can measure e-commerce activities and conversion rates. Web metrics can determine user behavior and the impact of changes to both navigational hierarchies and classification structures. There are many things that can be measured. Findability can be directly measured through usability studies and based on use cases. I'd say to get something that's customer-impacting, and that's measurable, and that has a direct effect on the bottom line or the top line. In other words, you're either going to see some revenue increase or you're going to see some costs decrease tied to the project.

You can find places where you can get measurable results.

MM: In your experience, as we start talking about really developing an enterprise taxonomy and a navigational hierarchy, would you say that it trends more toward pre-sales — buying and facilitation, or post-sales problem-solving and solutioneering?

SE: A lot of the work we do these days trends toward pre-sales. E-commerce sites can be directly impacted in very substantial ways. I'd say that's a pretty solid area to get measureable results. We typically measure website usability and product findability before our work and after. Then the results are obvious and data driven.

MM: Is there a minimum number of SKUs or products that a company should probably be marketing such that they would have a sufficient level of complexity and revenue to justify the investment in developing this?

SE: No. You can always have an impact on findability and conversion. This starts to dovetail into usability and user experience. There's always an opportunity to improve, understand more about your audience and how they're going about their purchase tasks, about their mental model, and translating that into better organizing principles on your site.

MM: In terms of a framework of "Good-better-best"— what would be the typical kinds of investments a client would make in terms of developing an effective metadata ontology and so on?

SE: There are a couple of elements. One element is the initial derivation. Depending on where the organization is on the learning and experience curve, this can be very resource

intensive. After the initial development of a framework, there is implementation and operationalization, which includes maintenance and upkeep. The organization needs to devote resources to managing things on an ongoing basis.

In order to overcome organizational inertia and to assess all of the areas where organizing principles need to be leveraged, in most cases outside help will be needed. It's just not possible to do this along with your day job.

This may sound self-serving, but in order to do this correctly you really do need methodologies and approaches along with a team that has done this before.

I'd say it could be for somewhere in the range of \$100,000–\$150,000, depending on scale and scope.

MM: So that \$100,000 would represent one or two independent consultants working a period of how many months?

SE: Typically an initial project is in the eight-week range for two consultants. This includes workshops and education, an assessment of information architecture and current metadata, review of controlled vocabularies, stakeholder interviews, content analysis, review of existing functional designs, gap analysis, and metadata and vocabulary development and recommendations.

More extensive projects around operationalization, governance, DAM and search integration, and full information architecture can be a multiple of that figure but would typically start in the \$200k range.

MM: Out of that couple hundred thousand dollars, does that entail any procurement of new technology?

SE: No. That's pretty much all consulting services. There might be some technology in there. It's possible to get a metadata or taxonomy management tool with that figure, which helps with implementation.

MM: In terms of a moderate range engagement, which would be startup and probably some ongoing operational support — what are we talking about?

SE: I would say between \$100,000 and 200,000.

MM: Then to have Seth and Team come in and really transform our content operations into a digital supply chain in much the same manner that we've talked, and really creating contextual

consumption models, and really helping us re-engineer our micro-content strategies and all that?

SE: You're probably going to make at least a \$500,000 investment, which would be extended over a couple of years. Some of our current clients are in multi-year engagements. . It would be very difficult to transform an organization in less time because changing how people work along with the systems they use simply takes time and cannot easily be compressed. That kind of transformation would not be a short-term engagement. To get started, more energy is expended at the start — educating people, assessing systems, identifying opportunities and gaps, developing a strategy — and execution and operationalization might be done internally or with minimal outside support.

It's also important to understand that there needs to be internal ownership of these kinds of initiatives. An external consultant can't do it all themselves. We hear clients say "you're the experts; just tell us what we need," But part of the solution is having people go through the process and making our recommendations relevant and actionable. Your mileage will vary, as they say. You have to engage internal change. You have to build the governance processes. You have to build standards. And you have to really operationalize it. I think that's where the longer-term efforts go to. I'd say \$500,000 over a year and a half or two years is not an unreasonable amount to spend on something like that.

MM: As we start to move toward the conclusion of this interview, are there any other topics that we might've started but didn't fully develop that you'd like to circle back on now?

SE: I think that the whole idea of optimizing search and all the related content mechanisms is something that we can emphasize. I like to say search is an application and not a utility, and organizations need to treat it that way by developing use cases and user scenarios, performing task analysis, understanding processes and mental models. This might be a topic for another discussion, but in a nutshell, search is part of your digital asset management strategy. Search is part of your metadata strategy and your taxonomy strategy. Search is another information access mechanism. And of course the lines between navigation and search are

blurring, which means it's even more important to get the information architecture right and to look at things holistically.

When you fully leverage metadata and implement things like faceted search and guided navigation — those are search, but they look like navigation. Stored queries look like navigation.

The idea of plugging in a search appliance and putting up a search box, and not thinking it through — not designing it as an application — is naive. That's not to say that there isn't a place for that. It's just like we said — there's a continuum of content value. But there's also a continuum of structured versus unstructured search, and different mechanisms are appropriate for different purposes and asset sources.

So search needs to be thought of in terms of various flavors: old fashioned full text indexing for web pages and documents, use of different mechanisms to access more structured information like business intelligence and transactional data, faceted search using metadata, entity extraction, clustering and categorization, federated search across repositories, semantic search, social search, and so on. There are lots of different mechanisms for doing search. That needs to be thought through as part of the DAM strategy.

MM: Sure. Cool.

Well, that sounds like a great place to conclude our interview. Now I'd like to address the Henry Stewart DAM MOM Symposia.

SE: Sure.

MM: Seth, as I understand it, you're going to be sitting in on the "Expert Experience Panel" on Monday, the 12th of May, at the Henry Stewart DAM and MOM Symposia.

What are some of the things that you anticipate developing? Or ideas that you'd be introducing there on the expert panel?

SE: I think that a lot of the things that come up on that type of panel will relate to the concepts of findability and usability and providing content in context. That's one of my typical areas that I like to speak about and talk about and work on. The role of various types of organizing principles in providing content in context.

How do you find information internally if you're a user of a DAM system? How do you

present content to users in ways that they want to consume that content, if you're exposing that to a public-facing type of application? And what are all the principles that you need to think about in order to do that? What are the pieces that you need to get right? How do you analyze your processes? How do you characterize your audiences and think about structuring your content in order to do that?

Those are the types of questions that typically get asked in those kinds of sessions.

MM: Great. That sounds like a great session.

Seth, would you share with us the skills portfolio or curricula vitae of someone who says, "You know, I really think that metadata and taxonomy and corporate ontologies and semantic web is really a career path for me." What would be some of the things that you would want to have as skills or industry experiences?

SE: That's a great question. In fact, I have an informational interview later on today with a library of science student from one of the local universities. Certainly a library of science degree is helpful.

It's funny — there's a colleague who works at another consulting firm that does not have a library of science degree. None of their people have library of science degrees, and they're saying, "The last thing you want to do is have a librarian do your taxonomy."

He's pretty much offended most of my staff with that statement. Most of my staff have either Masters or PhDs in Library Science. We kind of want to say, "Wait a minute. You mean the people who've been organizing information for thousands of years have no understanding of how you should organize information?"

When you think about how web development people and content development people and even data architects don't necessarily realize that there's a tremendous amount of knowledge and expertise that can be leveraged by the discipline of library science. Library science is foundational.

Even some of the newer things that we're seeing on the web, such as faceted search and faceted navigation, were developed in the 30s and 40s. Ranganathan was a pioneer in that area.

He was a person who came up with what he called "colon classification." That's now being exploited on the web.

A masters in library science is a terrific place to begin that process or to add to your knowledge. It's also important to realize that some of the newer library of science degrees really are talking about information access and information science. Courses in usability, web development and user interface design are very important.

Understanding that user perspective and how organizing principles are surfaced to the user is very important. That in addition to understanding basic theories of classification and content organization.

That combination is what we look for when we interview people who are interested in working for our organization. We like to see practical experience in the industry, usually in content management, knowledge management or DAM function. Usability is also very important. Some combination of courses in usability or web design, information architecture, library science or computer science is a good starting point.

MM: Excellent.

Any other sorts of things you'd look for or that somebody should consider having as an experience?

SE: Graphic design is always good to have some exposure to. It's helpful if you can think of ways to bring together the structured and unstructured, or the creative and the logical. Any time you can think left- and right-brained. English degrees are quite good. Especially English with a technical degree. That shows that you can think technically yet communicate with different types of people. So communication skills and writing skills are always very important.

MM: The ability to simplify, simplify, simplify?

SE: That's very important. I've found that one of the really great skills that someone can bring is the ability to take a complex subject and simplify it, and explain it in business or layperson terms. To speak very simply. I think that's an absolutely essential skill to have. That kind of ability is very valuable.

MM: Great. Thank you.