

Talking Taxonomies and Metadata with Stephanie Lemieux

Leading Learning Podcast Transcript for Episode 367

Stephanie Lemieux: [00:00:00] Taxonomy is part of helping an organization tell their audience what's important to them, what are they all about, what do they cover, and what language is being used in that organization that crosses a lot of different channels.

Celisa Steele: [00:00:24] I'm Celisa Steele.

Jeff Cobb: [00:00:25] I'm Jeff Cobb, and this is the Leading Learning Podcast.

Celisa Steele: [00:00:33] Managing content well and communicating effectively with audiences are two key competencies needed by a learning business that wants to thrive, and taxonomies and metadata can play an important role in both. Stephanie Lemieux is an information management consultant specializing in taxonomies and metadata. Educated as a librarian, Stephanie is now president and principal consultant at Dovecot Studio, which helps organizations optimize the way they structure and manage their mission-critical content. Stephanie and Jeff talk about what taxonomies are and the negotiation that has to take place between an organization and its audience to create an effective taxonomy. They also discuss the role of taxonomy in search, findability, integration, analytics, and personalization, and they talk about how content has gotten more targeted and more granular in Stephanie's years of work. They also touch on artificial intelligence and how AI and machine learning can both benefit from and contribute to tagging and taxonomies. Jeff spoke with Stephanie in April 2023.

Jeff Cobb: [00:01:51] Can you tell us a little bit more about the work that you do?

Stephanie Lemieux: [00:01:56] Absolutely. So this is a relatively new-ish profession in the sense of the digital world. Taxonomies and metadata became very much in vogue as e-commerce took off and people became much more familiar with needing to look through complicated structures of products and use filters to find what they were looking for. That translated pretty seamlessly over into the corporate world, where there is also tons and tons of content.

Taxonomies are really all about controlling and creating structures around terminology so that people can apply labels, apply structured content, and be able to better find that content when they're working on things. We help our customers—who span a very wide range of different types of industries—understand their audience's language, their information needs, and how to make the connection between the two. Taxonomies are really part of the whole interface between the collection and the user, helping to negotiate the conversation between the two, and they get implemented in things like content management systems, search interfaces, and lots of other different types of use cases.

Jeff Cobb: [00:03:15] I think most Leading Learning listeners are going to be familiar with a term like *taxonomy*. Maybe they've been through some sort of taxonomy project before, but I don't want to make too many assumptions. I know I don't consider myself an expert in taxonomy, but I think of it as having to do with tagging, with coming up with the right terms to describe content information, and then being able to use that in meaningful ways. Can you talk a little bit more about exactly what a taxonomy is and what goes into creating a taxonomy?

Stephanie Lemieux: [00:03:46] Yes, definitely. We still have lots of people who ask, "What is a taxonomy?" It's a word that can sound complicated, but, really, all it means is a structure of words. What's important about those two things is that it's not just any structure; it's a structure that has been agreed upon by a community or a set of people who share a mental model. The "words" part is that it's agreed upon and controlled words. So we are choosing what words we are going to use to describe our content and organizing those words into a structure that makes sense to the people who are going to be using the content. Most organizations have more than one taxonomy. So, if you think about in an e-commerce perspective, for example, you've got the type of product that it is; you might have the size, the color, and the brand. Any time where you have terminology that needs to be harmonized and you need to have a single way of talking about things so that you can get everyone to the same place, that's where taxonomy comes into play.

Stephanie Lemieux: [00:04:54] And the "structure" part comes into play when you are exposing that taxonomy to people and asking them to navigate, so you can put those terms into a hierarchical order if there's a grouping that you want to use to get all the appliances together and to get all the shoes together. You can do lots of grouping, but you can also have different vocabularies playing at the same time. We call this a faceted taxonomy. This is becoming more and more popular outside of e-commerce, so in a corporate context as well.

Jeff Cobb: [00:05:29] You made an important point there around this needing to be a set of language, basically, that people have to agree to and that the people who are involved have to agree to. Can you talk a little bit about that process of getting a group to the point where you've defined the terminology and have gotten people to agree that this is, in fact, what we need to be using in describing whatever it is we're talking about, as you said—the shoes or other products?

Stephanie Lemieux: [00:05:57] Yes, and that's probably the most fun and interesting part of taxonomy work, actually, is that whole negotiation process. If you think about an organization, this can be any kind of organization, any kind of corporate entity, and its audience trying to have a conversation together. If I'm a professional association of tax accountants, let's say, I'm creating content, I'm creating articles, I'm creating learning material, and I want to publish this content to that audience. I need to use terminology, and I need to organize this content in a way that makes sense to my audience. Getting agreement within the association of the people creating the content is one part of it. But then marrying that up to the terminology that makes sense to the audience and getting some harmonization between the two lenses is where it gets a little bit crunchy but also very interesting.

Stephanie Lemieux: [00:06:54] A big part of our work is understanding what kind of content is being created, and what kind of language is coming out of that content naturally by the content creators? What kind of language is used by the organization and do they want to use to portray the kind of content that they have or the kind of areas that they cover? But then, very importantly, what kind of language is the audience using and expecting? What kinds of things are they searching for in Google? What kinds of things are being talked about in the industry at large? Because you don't want to be making up terminology that makes sense to no one, and you don't want to be invisible on your little corner of the Internet. There's that negotiation that has to happen between those three lenses.

Jeff Cobb: [00:07:43] It sounds like you probably do have to do a certain amount of outreach to your audience to understand how they're thinking about what you offer. If they were going to go search your Web site or whatever, what would they actually put in to potentially get them to whatever you're offering?

Stephanie Lemieux: [00:07:59] Yes, absolutely. We start most of our projects with a little bit of user research. That may be direct user research, where we actually talk to a focus group or run some interviews with some representatives, folks who've volunteered to chat with us. This can be pretty easy if it's a membership-type organization. They can just pull some members. It could be customers. It could be anywhere where you have access to a group of consumers of the

content. But you can also get indirect user research by looking at things like search logs. If you have a search engine on your Web site, if that's how people are interacting with your content, people are typing in things, and what they type in is usually pretty reflective of what they're thinking about and how they might be talking about a particular subject. You can also glean that context from even forum posts or other ways that your user base is interacting with you or interacting with the rest of their peers or their industry in public spheres.

Jeff Cobb: [00:09:04] How does taxonomy relate to what I think is a bigger, more umbrella term like "information architecture" on the one hand and what I think is a narrower term—you can correct me if I'm wrong—around something like "digital asset management," which I know a lot of organizations are very concerned about?

Stephanie Lemieux: [00:09:24] I like how you're putting it into a hierarchy here—broader and narrower terms. That's perfect taxonomy.

Jeff Cobb: [00:09:29] There you go.

Stephanie Lemieux: [00:09:30] Yes, it's definitely part of information architecture. So, if information architecture is all about structuring and organizing information in some kind of presentation vehicle, whether that be a Web site or information portal internally or any content-consuming application, how do I chunk and architect information in a way that makes sense to people? How do I present things to show them what's more important and what's less important? How can I give them navigation options? How do I present to them a menu that makes sense to them? The taxonomy can be a very important part of an information architecture. If you've got navigation, the taxonomy can show up as part of your navigation. The taxonomy can also feed into how and where content gets presented in an interface. So, if you've got content that has taxonomy applied to it through tags, then you can say, I want to put a box on my landing page that shows the most recent news articles that we've published on these topics or for this country, depending on where the user is coming from.

Stephanie Lemieux: [00:10:42] The taxonomy can help direct content and help get people to get to the content, and it's also very often seen in search information architecture as well. You'll see it as faceted filters or as other kinds of related content as part of page design. Now, when it comes to digital asset management, that's one application of taxonomy. Digital asset management is just the management of videos, images, and other media objects, and their taxonomy takes on even more importance because those media assets don't tend to have a lot of text native to them. They're visual imagery, or they're audio files. You have to put more effort

into the structure of data that you put around those assets to make those assets findable and put them into a context where they're reusable by the folks that need to use them in content.

Celisa Steele: [00:11:48] As someone who listens to the Leading Learning Podcast, you should know about the Leading Learning newsletter, which you can subscribe to at leadinglearning.com/inbox. The newsletter is inbox intelligence for learning businesses and helps you understand the latest technology, marketing, and learning trends and grow your learning business. Best of all, it's a free resource. As a subscriber, you get Leading Links, our monthly curated collection of resources to help you grow the reach, revenue, and impact of your learning business; the podcast digest, a monthly summary of podcast episodes released during the previous month; plus periodic announcements highlighting Leading Learning Webinars and other educational opportunities designed to benefit learning business professionals. Subscribe, for free, at leadinglearning.com/inbox. And, if you're already subscribed, point a colleague to leadinglearning.com/inbox.

Jeff Cobb: [00:12:49] I'd love to hear about what a relatively typical client engagement looks like for you. I assume a client has some information content challenge that taxonomy, applying metadata, and all of those things are going to help with. What are some common situations that you might walk into?

Stephanie Lemieux: [00:13:09] We often get presented with some kind of technology use case. We've just bought a fill-in-the-blank platform, and it needs metadata, and part of that metadata is probably going to be some kind of taxonomy. We get pulled into projects like implementing a digital asset management system—as you were just talking about—or implementing a new Web content management system; doing a Web site redesign—that's also a common use case; a new search engine. It could be more internal-facing things like intranets, [which] is also a big deal. Then sometimes it's more functionally based. So we want to do a better job of findability across multiple applications, or we need to do integration between our customer relationship management system and our knowledge base. In order for those two systems to talk to each other, they need to share metadata, share tags, and be able to pass information back and forth. Another big one is personalization. We want to understand more about who's consuming our content, and then we want to translate that back into targeted e-mail campaigns or show them personalized content once they're signed in.

Stephanie Lemieux: [00:14:32] All of that is relying on these metadata structures and controlled vocabularies being applied to content behind the scenes, but it touches multiple applications. I would say those are the two big types of projects that are either single system—we're

implementing a thing, and we know we need taxonomy because that's part of the metadata and part of the user experience for that specific tool—or taxonomy is more of an enterprise architecture component, where the taxonomy is serving more of a function of unifying metadata across lots of applications or enriching the organization's ability to do something with the data or with their content, such as personalization, analytics and reporting, or any of these wider functions.

Jeff Cobb: [00:15:27] I suspect our listeners are probably going to fall in both of those camps pretty commonly. On the one hand, they might have, say, a learning management system or a learning content management system that is that single platform that needs some organization that goes with that. But then that's usually part of a suite of different platforms that an organization will have. They might have a membership management or CRM-type system. They might have publications in another system. They need some way for, for example, a user on their site who maybe isn't in the learning management system right now or isn't in the publication part of the site, but, if they are going to search, it needs to bring up the relevant content for their search from wherever it happens to live. That sounds like the type of scenario that would be in your sweet spot.

Stephanie Lemieux: [00:16:13] Absolutely. I think that you're touching on a really important part of taxonomy, which is that taxonomy is part of helping an organization tell their audience what's important to them, what are they all about, what do they cover, and what language is being used in that organization that crosses a lot of different channels. If I, as a user, am coming to your Web site, then I have to log in and use your knowledge base, and then I'm applying to go to one of your conferences, I'm choosing my tracks, or I'm filling in my personalization—here are the topics that I'm interested in; please e-mail me your digests every few weeks. These are all different channels and touchpoints, which, behind the scene, for the organization, may have lots of different applications sitting behind them. But the user has a reasonable expectation to be able to learn and understand the language being used and not have to learn a new language with each new channel that they are interacting with you. Any time we do a project, even if it is a single application being replaced, if it's a new CRM or a new LMS or whatever, we don't want to look at taxonomy as a siloed piece of architecture that's just for that one application.

Stephanie Lemieux: [00:17:37] Even if that is the project, we will always be looking at taxonomy across all of those different touch points with the user because we don't want the user to have to learn a new lingo with each different application. If we're going to call it a "car" versus an "automobile," we want to try to be consistent as an organization, not only for the user's

experience but also for our own capability as an organization to understand our interactions with that user, gather data about it, and be able to do something with that data. If you're gathering analytics, for example, and you're gathering analytics from your CRM, from your LMS, from your conference management tool, and you have some from your membership management tool as well, if they're all using completely different metadata structures, you're going to have more work ahead of you to be able to consolidate that information and get meaningful insights from it. Whereas, if you do a little bit of legwork upfront and do a little bit of harmonizing of the language and metadata architectures being used, then you can really level up your ability to understand and react to your user base.

Jeff Cobb: [00:18:55] I think that's just such an important thing. We see it happen so often in organizations—I'm sure you've seen it too—where people are in their particular silo, and they've got their particular project or initiative. They're really hunkered down on it and trying to make it work. But, yes, they're calling something one thing, and another part of the organization working on the same type of stuff is calling it something completely different. And, as you said, that's going to impact the user experience dramatically, but it's also impacting your experience as an organization and what you're able to do. Everybody likes the idea of being able to use all of this data that we can get now. But, if the data is all over the map in terms of using different languages for the same thing, you're going to have a really hard time doing anything with that data.

Stephanie Lemieux: [00:19:34] Yes, absolutely. It's so funny how taxonomy projects are this weird moment where sometimes teams within the same organization meet each other for the first time and shake hands like, "Oh, I didn't know you were working on that. Here's what we're doing." There's this wonderful exchange of ideas and knowledge that happens as part of a taxonomy project, which is really fun to be a part of.

Jeff Cobb: [00:19:58] A tip for anybody listening out there: If, for example, you are implementing your learning management system or your learning content management system, and you're applying all of this terminology based on your competency model or whatever it is, make sure you're talking to those other parts of your organization and that they use the same terminology for those things so that you can provide that great user experience and, really, a personalized user experience in the end to get the member, the user, or the customer to the content that they really need.

Jeff Cobb: [00:20:37] A lot of what we're talking about here is just around content and information. Content, obviously, has been important since the beginning of the Internet. It was

important before that, but it became such a focus so early in the Internet. It's grown exponentially in importance over time. You're always hearing, "Content is king." It continues to be a mantra. I'm wondering, in the work that you do, what have you seen change, particularly over the past five to ten years, about content and how you have to think about content? Anything that's really different now than what it was ten years ago?

Stephanie Lemieux: [00:21:14] Yes, I think it has changed quite a bit, actually. We went through this whole microcontent phase, and I think that fad has gone a little bit, or we're talking about it differently. But social media came in and disrupted a lot of things. I think we're creating content differently than we used to, and content is a lot more targeted rather than this large-audience, long-form content, one size fits all. It's a lot more granular, it's a lot more personalized, and there's a lot more needed architecturally to make those things happen. It's no longer these monolithic Web sites, for example, with the same pages that everybody sees. It's a lot more, again, targeted and personalized, and the content is being managed in a much more atomic way. Everything's getting carved up into smaller pieces, and those pieces are being reused across lots of different channels. We have all these systems now that are trying to help that process of multi-channel content dissemination. And the taxonomy has become a big piece of that as well. So being able to write an article or a piece of learning content in a way that is now granular and even broken up with XML and tagged at the paragraph level, that's a lot more popular than it used to be.

Stephanie Lemieux: [00:22:49] It used to be that you would only get to that level of detail if you were writing these 300-page technical manuals. But now people are doing that more and more for regular Web and learning content in order to be more efficient but also to provide a more nimble user experience, one that can more efficiently cross channels. We're seeing this even more so now with AI becoming part of the picture and seeing chatbots and things like that. In order to make a chatbot work, you have to have a robust content collection, and that content collection has to be granular and divisible into chunks that can get tagged or can get some metadata so that a bot knows what to do with that sentence or with that small paragraph.

Jeff Cobb: [00:23:42] I'm not sure if this is the right way to think about it or not, but it seems like it's almost as much about context as it is about content because you've got to match content and context to get the maximum results.

Stephanie Lemieux: [00:23:56] Yes, context absolutely is such a key thing, more now than it used to be. Before, it used to be a little bit more "one size fits all," as we were just talking about. But, if we've learned anything about people and language, it's that everyone is different,

everyone thinks differently, and communities, ideas, and trends can be so much more small-scale, especially [on the] people and interests' side. So we have all of these communities of interest popping up, and people can be interested in very unique things. Understanding those interests at that more granular level can help you target content at the right level for the person. That's probably been one of the areas that we've seen the most growth: using taxonomy, information, and architecture a little bit more generally to help identify what people are interested in and be able to match the organization's content to those interests, not only in terms of size and structure but also at the right channel.

Jeff Cobb: [00:25:05] It seems to me that this is so core to realizing the potential for personalized learning, just-in-time learning, and all of the things that have been buzzwords in the learning and development field for years now. You've got to be able to match content and context very flexibly and very rapidly to make that happen. You mentioned AI in this. I'd be interested to know what possibilities is AI now creating? How might it be a tool in your work going forward? I'll say, in the back of my mind—and I know just enough to be dangerous about this—I'm thinking about things like the Semantic Web, which seemed unrealizable to do what's necessary to have the Semantic Web before AI came along and potentially made it possible to help do all of the tagging and everything that would be necessary around that. Like I said, I'm probably speaking about things I shouldn't speak about, but I'll bounce it back to you. What is AI going to help us do?

Stephanie Lemieux: [00:26:01] I think AI is going to take a lot of the grunt work away from some of the process of setting up these architectures, certainly. We've been using ChatGPT even in our taxonomy work to ask it questions—because it has basically read the Internet—to say, "What would you call this group of things but not this and maybe this?" It does a pretty good job of trying to give us answers about potential terminology, given that it has, as its context and pool of resources, the entire Web. Which is not to say that I would want ChatGPT to create a taxonomy for me, but it's been an interesting input into our process. Generally speaking, though, the part about AI that's fascinating is that everyone seems to think that it's so magical, and, when you do interface with something like ChatGPT, it can feel very magical. Behind the scenes, there is so much that's happening that is based on very rich, robust semantic frameworks that have been in the works for decades. So you've got Google, who's been working on this huge knowledge graph behind the scenes. Wikidata and Wikipedia have an enormous knowledge graph behind the scenes. A knowledge graph is really just an architecture that is identifying different kinds of entities and creating relationships between those entities, and that is a necessary architecture for these types of AIs to have gotten where they have today.

Stephanie Lemieux: [00:27:48] It's the same for internal organizations and their internal content. They have to put some legwork into their semantic framework in order to make things like chatbots and other types of internal AI work for their content. I just want to make sure that everyone understands that this isn't a magic bullet. I'm not just going to point ChatGPT at something, and it's going to solve my universe. ChatGPT is being trained on what's available on the Web. Most of the time, if you're looking at internal content, you have to help any kind of AI, machine learning, or other graph-based technology to understand who you are as a business, what's important to you, what are your entities, and what are the relationships between them, in order to do a good job with your own content. Will AI solve everyone's problems? No, I don't think so. In order to solve some of our problems, we still need to think about what kind of semantic framework is necessary to put that AI on top of. It doesn't absolve us of the work of having to do a little bit of taxonomy and metadata.

Jeff Cobb: [00:29:15] AI is getting all the buzz right now. It's getting all of the focus, particularly ChatGPT. But are there other factors out there that you're looking at, whether technological or otherwise, that you think are really going to influence, shape, and impact the future of information management and content management, those areas?

Stephanie Lemieux: [00:29:35] We've been using an enormous amount of machine learning-based things in various parts of our work. Machine learning-based search has been a lot more prevalent, where it's taking relevance to the next level. It's not just relevance based on what you've typed but also who you are and what other people have typed ahead of you. So enterprise search and even site search—not talking about Google here, but in terms of organizations and their content—that has gotten so much better. It is relying on taxonomies and metadata to help that machine learning learn more about the context of the users and their needs. I've done a couple of projects now where we've been working with knowledge bases and call center agents that are trying to find content while they're on the phone. The search is learning as they search and becoming even more and more powerful in understanding what kind of client they're even talking to, what kind of question they're trying to answer, or what topic a ticket is open for. Search and machine learning is definitely a big area.

Stephanie Lemieux: [00:30:49] Machine learning and tagging are also more and more prevalent. We have a lot more clients that have large volumes of content and not a lot of time and energy to apply metadata. So, yes, they want taxonomy, and, yes, they want good metadata and good, structured content, but they don't want people to spend an enormous amount of time doing that. In the old days, I think auto-classification or the automatic tagging of content used to be mostly newspapers because they were publishing hundreds of articles a day. But I'm seeing

more and more auto-classification happen within organizations because it's become a little bit more easy to implement and more accessible. So I think those are the two big areas that we've seen a lot of machine learning happening in terms of taxonomy and metadata.

Jeff Cobb: [00:31:41] Right. Switching gears just a little bit before we wrap up here. Thank you for a great conversation around taxonomy, information architecture, and all of the areas that you work in. We'd love to ask about your own approach, though, to lifelong learning—since this is a podcast focused on learning—and find out how do you approach lifelong learning? I'd love to know, too, if you have practices that relate to what is often characterized as personal knowledge management (PKM) to organize your own information over time. Do you have tools that you tag your own information that you've collected with and make it possible to retrieve and revisit, that sort of thing? But, just in general, how do you approach lifelong learning?

Stephanie Lemieux: [00:32:26] I have to say that I'm very blessed and lucky to be in the discipline that I'm in because, as a taxonomist, every project is so different, and I get to learn about an entirely new sphere of the world. One project will be with neurologists and talking about brains, and then another project might be about taxes, and another project might be about fashion. It's been such a great experience to be able to constantly be exposed to completely different domains of knowledge. And then, from a taxonomy and information architecture perspective, that domain has evolved so much in the last 20 years. It's fun having to keep up, especially with this whole AI thing that started maybe ten years ago. AI started way before that, but it became the hot topic not that long ago. So I've been trying to stay up on that. I use podcasts and blogs mostly. I go to conferences as well. There have been a lot more open courses that have been available, so I can stay up on top of the technical parts of it, learning new coding languages and things like that. Not that I'd do an enormous amount of coding as a librarian.

Stephanie Lemieux: [00:33:52] Yes, I would say that's been my learning approach. It has been just trying to keep on the pulse of what's happening, regularly expose myself to new professional conferences, and take a course here and there. And then, from a personal knowledge management perspective, I have to say that it's a little bit more challenging as a small business. We're not an enormous amount of people, and we don't need these giant multimillion-dollar software packages, which is where you have a lot of those good tagging mechanisms. For individuals and for small businesses, it's a little bit more sparse in terms of options. We've actually had to create some of our own information architecture and hack things like the Google Suite to do what we want to do in terms of managing our own small business information and personal information as well. I'd love to see the industry get better at individual stuff, but that seems to be focusing more on the automatic, "Don't worry about it-

Google will tag your photos for you, Google will tag your documents or tag it behind the scenes," so you don't need to. I think that would be an area that I'd like to see improved.

Jeff Cobb: [00:35:12] The scary side of the whole Google world is that kind of thing is happening in the background. In a certain sense, it's Google and these other large companies that are providing the automated structure to what we're doing. I use Evernote personally to track a lot of things, and it just occurred to me as I was listening to you talk that I have, occasionally, over time, tried to take advantage of tagging things in Evernote, but I don't think I have consistency even in my own mind around what I call things. I just tag this stuff all over the place and will probably never be able to find things that should be completely related, but I've called them something completely different.

Stephanie Lemieux: [00:35:51] I feel like those tools don't set you up for success, too, because I think they assume that the vast majority of people are not going to put that effort into coming up with their own personal little taxonomies. They don't tend to be very good at prompting you and helping you follow on and be consistent in your own information organization. So, yes, it's probably not just you.

Jeff Cobb: [00:36:16] Yes, I suspect I'm not alone in this.

Celisa Steele: [00:36:27] Stephanie Lemieux is president and principal consultant at Dovecot Studio, an information management consultancy with extensive experience using taxonomies, metadata, and other knowledge organization systems to enhance content findability and usability.

Jeff Cobb: [00:36:42] In the show notes for this episode at leadinglearning.com/episode367, you'll find a link to Stephanie's LinkedIn profile and the Dovecot Studio site.

Celisa Steele: [00:36:52] Jeff and I would be grateful if you would rate the Leading Learning Podcast on Apple Podcasts or wherever you listen, especially if you find the show valuable, because ratings help improve our findability.

Jeff Cobb: [00:37:04] And please spread the word about Leading Learning, whether in a one-on-one conversation with a colleague or a personal note or on social media. In the show notes at leadinglearning.com/episode367, you'll find links to connect with us on Twitter, LinkedIn, and Facebook.

Celisa Steele: [00:37:20] Thanks for listening, and see you next time on the Leading Learning Podcast.
[music for this episode by DanoSongs, <u>www.danosongs.com</u>]
This transcript accompanies the episode of the Leading Learning Podcast
available at <u>www.leadinglearning.com/episode367</u> .