Why Taxonomy Development Is Graceless, And What To Do About It

By Patrick Lambe

Why do many enterprise taxonomy projects fail? In this thoughtpiece I explain why taxonomy projects are extraordinarily fragile in relation to the quality of the initial evidence gathering stage. I outline a methodology for improving the robustness and effectiveness of an enterprise taxonomy design, through a systematic knowledge audit at the project's outset.

Gracelessness and Requisite Variety

In the data storage domain, we use the concept of graceful and graceless degradation in the quality of data retrieval. Graceful degradation is common to analogue storage media such as vinyl records, audio tapes and filmstrips. It means that the quality of data retrieval degrades proportionately with the quality of the storage medium. Parts of the media might be damaged or degraded, but information can still be retrieved from the undamaged portions.

Graceless degradation relates to many digital storage media. In such media, data is stored randomly, but some sectors contain the addresses of the various data elements. When this kind of storage medium degrades, a small error to the address sectors means a sudden and total collapse of all data retrieval as distinct from gradual deterioration.

It turns out that enterprise taxonomy development is graceless in relation to how the quality of outcomes map against the quality of the evidence sampling.

Here's why. In much enterprise taxonomy development, the primary goal is to serve the information requirements for tasks and decisions across the enterprise. This means the taxonomy must somehow represent and reflect the diverse ways a wide variety of user communities think about, and organize, their information content.

This is a bigger challenge than it may sound. Human groups – even in the same basic organisational function – excel in being different from each other in their classification and labelling decisions, once they are no longer in close proximity and interaction.

Ashby's Law of Requisite Variety is a well-established principle in information science and cybernetics. It states that any system designed to serve a given environment must have a comparable or greater level of variety and complexity compared to that environment. The system cannot be simpler in its capabilities than the range of requirements that users may place on it.

In taxonomy and metadata terms, the taxonomy content and structure must match and accommodate the diversity of the needs it is required to serve. Failure to capture evidence of information needs, uses, gaps and potential flows at the requisite level of granularity and variety means a complete failure of the taxonomy capability – impacting search, discovery, navigation and sensemaking.

Poor taxonomy design from inadequate evidence sampling does not produce partial failure – it produces complete failure. This is why it is graceless.

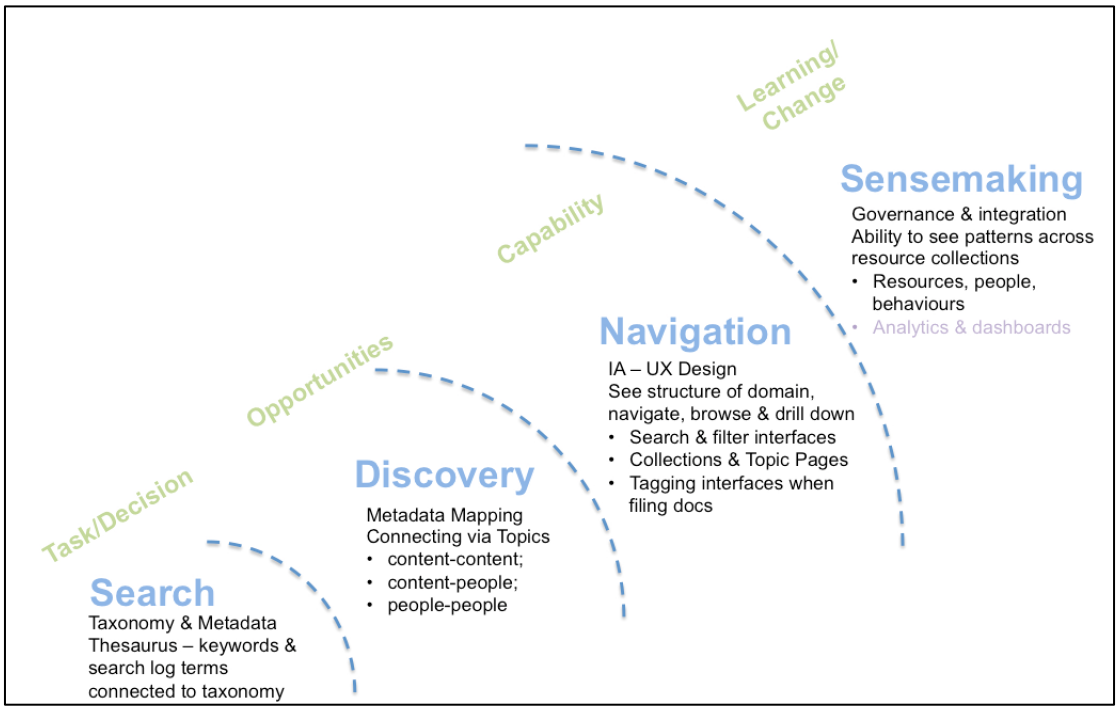
This means that evidence gathering and needs analysis must be comprehensive, complete and sufficiently granular for any benefits to be realised. (By contrast, other aspects of a typical knowledge and information management project are more graceful – in culture analysis or analysis of current intranet usage for example, the quality of the insights scales more proportionately to the quality of the samples).

This is why the quality, comprehensiveness and granularity of the initial evidence gathering and needs analysis stage are so critical. Skimping on or glossing over the initial evidence sampling has enormous impact on the quality of all the downstream outcomes.

Why is it so Important to Get the Taxonomy Right?

So what are the consequences if we don't get the initial sampling right, and the taxonomy turns out to be mismatched to the user communities and the information content it's meant to serve?

The following diagram attempts to illustrate this.



The Taxonomy Capability Hierarchy *Source: Patrick Lambe, 2015*

Many people who think about searching and exploiting an information or knowledge base, think simplistically in terms of Search and Find. User types in keywords. Search engine matches keywords with relevant content. Hey presto, relevant results turn up in the search results page.

Fewer people recognize that it's useful to have a taxonomy or thesaurus at the back end, so that the search engine can use it to resolve variant keywords meaning the same thing, to the same basic concept, which is represented in the taxonomy and connected to content tagged with that content. Related content can be pulled together even if it's tagged with different keywords, and even if the users are using different keywords to search for it. So far so good.

In fact, however, information seeking and using in enterprises is much more sophisticated than this simple model. Information capabilities go far beyond simple search and find. The Taxonomy Capability Hierarchy diagram above illustrates this.

Let's look at the levels in the hierarchy. The foundational element in the hierarchy is support for **Search/Find** against specific searches through a search engine, in relation to specific tasks and decisions across the organisation. Somebody is doing a work task, they need relevant information resources or templates or help, and they conduct a search related to that task.

Search/find should retrieve relevant information resources to support those tasks and decisions. The taxonomy, metadata and supporting thesaurus are the core elements supporting this function as we explained above. This benefit is primarily about delivering efficiency and effectiveness benefits against **known tasks and activities**.

However there are several more levels of benefit from an organisational taxonomy capability.

The next level of benefit after Search/Find is **Discovery**. This means that staff engaging in a search will find useful information resources that they had not been aware of. This depends on content and people being consistently tagged by taxonomy topics, so that they can be associated and connected by the system, resulting in the capability to connect content with content, content with people, and people with people. If multiple information and data systems are involved, then it requires cross-system metadata mapping, so that the different information platforms can be connected/cross-mapped via the common taxonomy and metadata schema. This capability delivers **unanticipated opportunities** from the information infrastructure.

The third level of benefit is a set of **Navigation/Mapping** benefits. This goes beyond support for day to day tasks and activities, and delivers an ongoing capability in using the information infrastructure in flexible and novel ways. This is the point at which the taxonomy design and integration with the search engine, starts to integrate with the Information Architecture (IA) and User Experience (UX) design of the information platform.

Information Architecture should enable users to see the overall structure of different knowledge domains within the organisation from perspectives that make sense to them, to navigate them, browse them, drill down and filter resources (using the taxonomy) to those that are most useful at any given point in time. The IA/UX design delivers this through (among others) the search and filter interfaces, resource collections and topics pages, and tagging interfaces when filing documents. This capability starts to make the information environment capable of exploitation in much more open and flexible ways.

Note that each level of benefit is dependent on the strength and robustness of the prior levels. Discovery depends on the taxonomy/metadata in a strong Search/Find capability, Navigation depends on strong Search/Find capability and strong Discovery/Connection capability across platforms.

The final and highest level of benefit is at the organisational level. Once the information infrastructure is connected through common taxonomy/metadata and IA/UX design, and provided there is strong governance in place, then it should be possible to be able to sense

useful and meaningful patterns about information importance, use, and behaviours. Dashboards and analytics would help in this, if you can make them available via your platform's reporting tools. This is an organisational learning/**Sensemaking** capability.

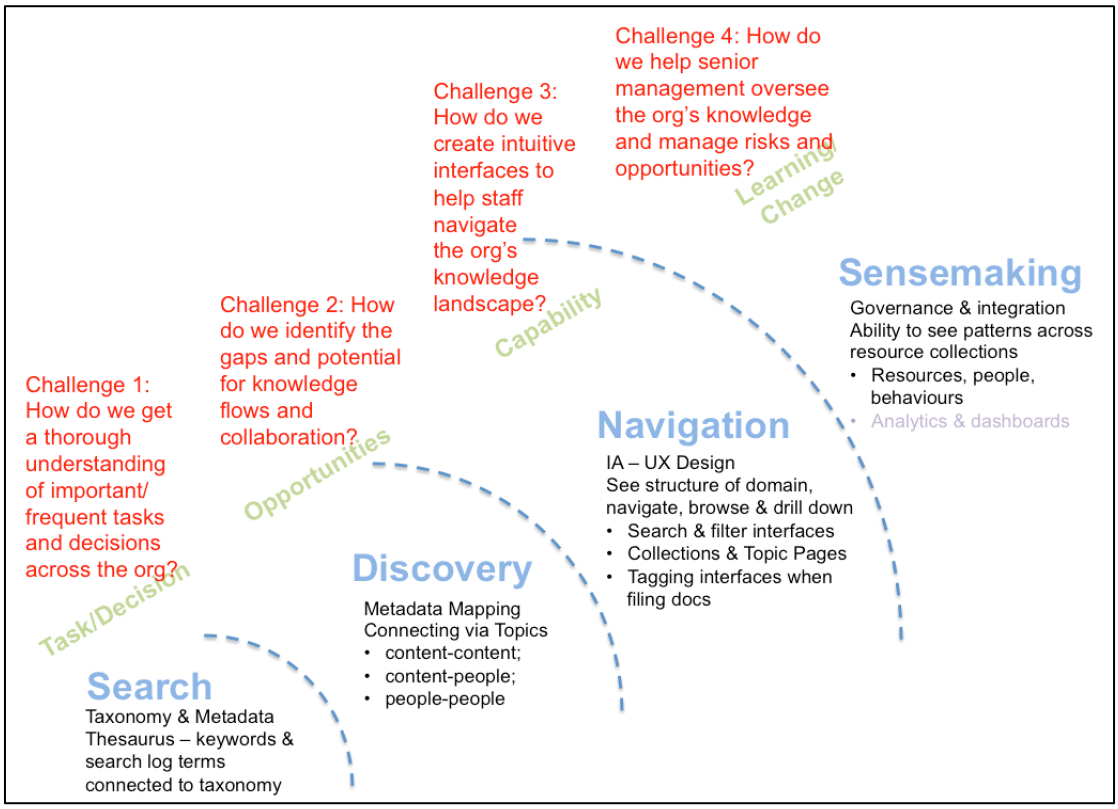
The Taxonomy Capability Hierarchy tells us two things. This first is that enterprise taxonomy work is about more than enabling simple search/find functions. It can – and should – have much greater organisational benefits.

The second thing it shows, however, **because** of the dependency of each level of benefit upon the prior level, is that if you get the first level wrong (taxonomy and metadata to support everyday tasks and decisions), everything else fails. Gracelessness cascades through the whole organisational system.

This is why it is so important to get the first stage right.

The Four Challenges of Developing a Taxonomy Capability

We need to meet four types of challenge, one for each level of the Taxonomy Capability Hierarchy. This is captured in the diagram below.



Challenges to the Taxonomy Capability Hierarchy *Source: Patrick Lambe, 2015*

The first challenge, at level 1 of the hierarchy, is how to get a thorough, sufficiently granular understanding of how tasks and decisions across the enterprise need to be supported by information at this point in time. This involves understanding how information resources and the tasks they support, are thought about, and described by the various user communities.

The second challenge, at level 2 of the hierarchy, is how we design for novel and improved information flows, even where the possibility of those information flows is not currently recognised. How do we enable discovery when most people work in silos and don't see what information resources exist outside their normal work horizons?

The third challenge, at level 3 of the hierarchy, is how to create interfaces that make sense to a wide variety of user communities on the ground. This has a critical dependency on solving Challenge 1.

Finally, governance does not happen unless it is sanctioned and protected from the top. How do we help senior management understand their role in protecting a strong, coordinated and integrated approach to the structure of the knowledge environment? Without this, different parties will take control of different portions of the landscape, which will devolve again into silos and special interests.

Again, it's clear that addressing each challenge is dependent on addressing the previous challenges.

So How Do We Get the Taxonomy and Metadata Right?

Our methodology for taxonomy development, grounded in a knowledge audit/ knowledge mapping methodology, is evidence-based (as distinct from opinion-based) and is organized around the key functions, needs and potential knowledge flows of the organisation.

In an evidence-based approach to taxonomy design we talk about the importance of **warrant**. Warrant simply means that there is evidence for a particular pattern of information usage. It is used to ensure that the taxonomy is hospitable to known needs. There are three types of warrant.

- **Content warrant** provides the evidence base to justify the inclusion of terminology and organising structures. It is collected by analysing the way that information resources are currently being labelled, described, and organised.

- **User warrant** provides the contextual understanding about the way that users interact with information resources (the important or frequent tasks and decisions that information resources need to support). It guides design decisions, and provides the means to evaluate the effectiveness of the taxonomy in achieving its purpose.

- **Standards warrant** is important where information and data need to be exchanged between different agencies, and there is a requirement to use specific vocabularies for this purpose.

Now of course we can conduct a content audit of the existing information resources and repositories to be covered by the taxonomy. This is in fact an important step in the content warrant stage. However, most information collections in an enterprise are accretions built up over years, and the labelling and organisation decisions were not only mostly ad hoc, but were likely influenced by many different factors at different times. A simple content audit does not help us see beyond this historical baggage to what is important and salient now.

Nor does it give us adequate insight into the contexts of the current user communities and their primary needs.

This is why we use a **knowledge audit**, which has at its centre a knowledge asset mapping activity.

The term "knowledge audit" generally refers to a composite set of activities, which may look at any combination of knowledge management processes, knowledge management gaps and pain points, knowledge behaviours, knowledge asset maps, and knowledge flows, actual and potential. When we use the term in this paper, we refer to knowledge asset mapping, knowledge behaviour analysis, and knowledge management pain points analysis. In this paper, the knowledge asset mapping component is the most critical for developing a taxonomy management capability.

Remember Ashby's Law of Requisite Variety? We need to create knowledge asset maps at sufficient level of granularity that they help us understand typical and important tasks and decisions across the enterprise. Typically we build knowledge asset maps at department level, but in large, diverse departments, we may go down another level to units or teams, just to capture the diversity of activity that takes place. Where there are several small departments that do pretty much the same thing, we may go up to divisional level, if that is the best unit for discriminating one group of activities from another.

Once we've decided at what level to map, we facilitate workshops, where 2-3 knowledgeable representatives from each workgroup are gathered to identify their key business activities, and then to map against them, the key knowledge assets (both tacit and explicit) that are required for those activities to happen. They also map knowledge assets that are produced by the business activity, and knowledge gaps. We identify where the knowledge assets come from, and where they go. These knowledge maps are produced by all of the workgroups in the enterprise.

Once they have been created and validated, they are exposed to all the other workgroups in the enterprise. We call this a subscription exercise. Workgroup representatives then browse the maps created by their colleagues in other workgroups and identify knowledge assets that would be useful to them in their activities. This is the most powerful step, because it creates a visibility into knowledge assets that exist elsewhere in the organisation that they had not been aware of previously. The knowledge assets that are identified most frequently as being useful outside their producing workgroup suddenly become visible as highly valuable, sharable knowledge assets.

We have done three important things here:

1. We now have a current snapshot of how activities and knowledge assets are described, organized and used across the enterprise. This helps us cut through the historical baggage that the content audit gives us, to see what is most current and relevant now.
2. The knowledge maps are organised around important business activities, so we have a context for how they are used and applied. This gives us good insight into both content and user warrant.
3. The most frequently requested knowledge assets in the subscription exercise highlight the knowledge assets that are most important for the potential knowledge flows in the organisation – i.e. to support knowledge discovery. This means we now

have a basis in our taxonomy and information architecture design to foreground the highly sharable knowledge content, and background the rest of the knowledge assets (i.e. reducing the "noise" of less generally important knowledge assets).

Of course we do other important things. We do our content audit. We visit workgroups to see and understand their information practices in context. We look at common knowledge behaviours and pain points to understand where the most important improvement areas need to be. But the foundational activity – the one that gives most evidence and insight for the taxonomy and information architecture design, is the knowledge asset mapping activity of the knowledge audit. Get that right, and everything else will follow.

Patrick Lambe, February 2015.

ADDITIONAL RESOURCES

Poster on Taxonomy Development:
http://www.greenchameleon.com/gc/guide_detail/how_to_develop_a_taxonomy/

Poster on Knowledge Audit and Knowledge Asset Mapping:
http://www.greenchameleon.com/gc/guide_detail/poster_on_conducting_a_knowledge_audit/

Find out more about our knowledge asset mapping tool at http://www.aithinsoftware.com