Understanding and Managing Taxonomies as Economic Goods and Services
by Denise A. D. Bedford

EDITOR’S SUMMARY
As pervasive as taxonomies have become in knowledge management, their economic value is not often scrutinized, documented or fulfilled. Taxonomies range in cost and sophistication from flat lists through thesauri, semantic nets and faceted systems. Each can be an asset or liability to varying degrees relative to the resources expended to design, manage and implement it. A well-designed taxonomy produces value by enabling consistent information management, facilitating records management and compliance, building domain understanding and saving time through efficient access to information. Yet numerous potential liabilities can rise from a taxonomy generating poor search results or being incomplete, out of date, confusing to users or simply not used. Taxonomies must be viewed from a perspective informed by economics, considering assets and liabilities, actual and indirect costs, external effects on third parties, their value as a public good and the opportunity costs of resources invested. As advocates for taxonomies, we should present the case as knowledge and information economists, fully aware of the range of economic decisions and effects involved.

KEYWORDS
economics of information
taxonomies
costs
index language construction
information production

A taxonomy is a tool or structure that can be used to organize things [1]. There are many types or levels of taxonomies, including
- flat structure such as simple controlled vocabularies
- rings such as synonyms or authority control applications
- hierarchies such as classification schemes,
- faceted structures such as metadata
- network or complex taxonomies such as thesauri or semantic networks.

Each level of taxonomy provides important opportunities for organizations to organize and manage their resources effectively and efficiently. Those of us who have worked with taxonomies tend to think of them as inherently good because we have seen the value they can produce. However, the economic value of taxonomies as products and services is not obvious to everyone. And while many taxonomy efforts produce positive value, there are many that do not.

Taxonomies are no longer unique or rare products. Taxonomies are pervasive to our information environment (see Boiko [2], Fisher [3] and Lambe [4] among others). They are designed into our applications. Taxonomies are components of every web content management system and every document management or records management system. They are built into financial management systems, human resource management systems, enterprise architecture applications, machine translation systems, semantic applications and essential text and data mining applications. They are part of every web application and service we use.
Significant money is being invested in taxonomies today. Organizations pay high fees to engage consultants for advice on designing and implementing taxonomies. The literature is rich with guidance on how to design, construct, implement and manage taxonomies [5][6][7][8]. And the grapevine is rich in discussions of how to convince organizations to reconsider or reinvest in taxonomies after experiencing major disappointments. Given the pervasiveness of taxonomies and their role in the knowledge economy, it is time for us to address the economic aspects of taxonomies. What, indeed, are their economic characteristics?

**Economic Considerations – Value, Production Costs, Externalities, Public Goods and Opportunity Costs**

This article takes as a starting point the assumption that taxonomies are here to stay and constitute formal products and services that are inherent to our advanced information and knowledge environment. As such, they should be treated as formal economic products and services. What does such treatment mean? It means that we have to stop making assumptions and start asking economic questions. We pose the following five questions:

1. **What is the economic value of a taxonomy?** What kinds of assets and liabilities do taxonomies produce?
2. **What does it cost to produce a taxonomy?** What types of costs are there and how does the level or type of the taxonomy affect them?
3. **Do taxonomies produce externalities?**
4. **Do taxonomies have public good characteristics?**
5. **Do taxonomies present opportunity costs?**

When posed, these questions produce different answers for different kinds of taxonomies. Different types of taxonomies produce different assets and carry different types of liabilities. Each type incurs different kinds of costs, each requires different levels of investment and some are easy to maintain while others are far more complex. Some types of taxonomies have significant positive externalities and others have next to none. Let's walk through these questions and consider how the answers vary by the kind or level of the taxonomy.

**Question 1: What is the economic value of a taxonomy? What kinds of assets and liabilities do taxonomies produce?**

Today, economic value is defined as “the worth of a good or service as determined by people’s preferences and the tradeoffs they choose to make given their scarce resources, or the value the market places on an item” [9]. A further consideration is that economic value is not static but changes when the price or quality of other similar items changes. Historically, organizations tend to associate the value of a taxonomy with the costs to produce the taxonomy. This view needs to shift to a true focus on the value that taxonomies produce.

Taxonomies are tools that, when applied wisely, can generate value for an organization – they become organizational assets. An asset is defined as “a resource controlled by an enterprise as a result of past events and from which future economic benefits are expected to flow to the enterprise” [10, p. 984]. When an organization embarks on a project to create a taxonomy, it should understand up front what the economic value of that tool will be. The economic value of a tool is related to what it produces in the future. In order to consider a taxonomy as an organizational asset it needs to be identifiable and controlled, and it must generate future economic benefit. Value is thus seen as future economic benefit and utility. The taxonomy has utility if its usefulness can be identified and tracked and if it offers competitive advantage to the organization [11].

On the other hand, a liability is anything that is owed – it is a deficiency [11]. A liability draws or commits resources from the organization – it creates a debt. A taxonomy can be either or both an asset or a liability. It is only through use, actions and decisions that a taxonomy produces economic value and becomes an asset or a liability. What kinds of assets might we expect from a taxonomy? What kinds of liabilities might be incurred? How do we ensure that we produce assets? And, how do we avoid liabilities?

**Flat Lists as Assets or Liabilities.** Flat lists or controlled lists are the simplest form of taxonomy. They are found in every application and every information environment. When they are well designed, implemented and managed they provide value by guiding selections, advising users as to the options that are available and permissible. They also address quality control issues by...
restricting the permissible values. They can also produce liabilities, though, when they are incomplete, redundant or confounded. The most common reason why a flat list might generate a liability is that it is selected when another type of taxonomy would better suit.

**Synonym Rings as Assets or Liabilities.** Ring structures are important translation and interpretation structures. They allow us to bring together synonyms – true synonyms, close synonyms, misspellings, predecessors, abbreviations, acronyms and so forth. Rich ring structures help to expand the meaning of the words that are being searched. They translate ideas across time, across domains and across vocabularies. When ring structures are poorly defined they can generate liabilities in the form of unintended search results – unintended expansion.

**Classification Schemes as Assets and Liabilities.** Classification schemes are simple hierarchical arrangements used to organize “things” into classes or sets [7][12][13]. As with flat lists and rings, the existence of a classification scheme only means that we have an economic product – it does not mean that it produces value. When a scheme is well designed, implemented and managed, it becomes an organizational asset that produces positive value. For example, a well-designed, implemented and managed classification scheme enables organizations to promote, publish and syndicate information across the business. It ensures legal compliance with record-keeping regulations; facilitates information discovery; facilitates workflow across applications; increases both casual and legal discovery; increases efficiency of access; manages recall by bounding context; promotes the understanding of a domain; provides a structure for navigation and browsing; saves time for those who are looking for information; supports categorization of assets from the business perspective and supports interoperability by providing a view into different functional applications.

Classification schemes, though, can also produce liabilities when they are poorly designed, poorly implemented and not well governed or not maintained. Liabilities result where a scheme’s classes are not aligned with users’ understanding and expectations, where the design inaccurately represents the domain, where the scheme is imbalanced and prevents users from full discovery, where it discourages use through an out-of-date representation, where an ungoverned structure causes distrust or where simple naming conventions prevent understanding.

**Thesaurus or Semantic Net as Asset or Liability.** A thesaurus or a semantic network is a data structure in which nodes represent concepts, and links represent the meaningful relationships between concepts [14][15][16]. The concepts can be used to describe “things” in a way that enhances access. Relationships describe how concepts are related in a particular domain. Taken as a whole, a semantic network or a thesaurus can provide a view of a domain, even as a visual space. We might say that the value of a thesaurus or semantic network is greater than a class scheme because it provides greater understanding of the relationships among concepts. It is also richer than a class scheme in its coverage of concepts. When thesauri or semantic networks are well designed, suited to their purpose and well implemented, governed and maintained, they become institutional assets. For example, they support rich and granular concept-level searching, provide strong support for cross-language search, create a foundational rule-base for automated indexing, enable visual understanding of a domain, provide the basis for inferring relationships among concepts, may increase the relevance of search results and reduce the number of zero-hit results through automated synonym expansion and may guide searching through the exposure of related concepts. These functions are only a few of the ways in which thesauri and semantic networks may add value.

Just as they may create positive value, they may also produce significant liabilities. Where a thesaurus or a semantic network exists, but is not well designed, suited to its purpose or governed and implemented, it may fail to meet the organization’s expectations. Its scope and coverage may not align with the organization’s needs – it may represent what the designer thought should be supported. Its warrant may be too generic to be of value, or it may represent a user community that no longer exists. Its language – the concepts represented – may not reflect the way the organization talks or thinks. One of the most common liabilities we have found is that it incorrectly applies the definitions of semantic relationships and produces unintended consequences when implemented in search. This fault may lead to unintended explosions of relationships and dramatic increases in irrelevant results. Where a
A thesaurus may also suffer from other problems. A scarcity of synonyms can constrain the retrieval of all relevant results while the lack of governance may render the thesaurus out of date and of little value to current information needs. It may also be the case that the business rules for implementation of relationships are not followed and, as a result, the thesaurus is implemented partially or incorrectly. If the thesaurus is developed without a strong governance structure by multiple individuals there may be inconsistent development, which in turn produces uneven or unpredictable search results. Another common liability we have experienced is a thesaurus that is a conglomeration of different types of entities. The confounding of descriptors and identifiers means it is not possible to implement it effectively in search without further reverse engineering.

Perhaps a greater liability results when a thesaurus or semantic network has been created but has not been implemented as a component of the information architecture: a significant investment is made but, for example, the infrastructure to implement it does not exist. A liability exists also when the thesaurus becomes “shelfware” – it is not used in search but only to guide human indexing. This liability is significant because it means there is a zero return on the investment. It also carries significant opportunity costs for future efforts – an organization that has once invested in a taxonomy that has failed to produce results will not likely invest a second time.

Faceted Taxonomy as Asset or Liability. Faceted taxonomies are essential to any information organization effort. They comprise the foundation of any information architecture. Facets represent views or dimensions of a “thing.” Metadata is the most common application of a faceted structure. Metadata is an essential asset of any organization. Again, where it is well designed, suited to its purpose and well implemented and managed, it produces significant value, such as the following traits:

- It allows us to identify and distinguish any kind of institutional asset;
- It facilitates multidimensional discovery of information;
- It helps us to know whether information is classified and who can and who cannot access it;
- It supports interoperability across and beyond holding applications;
- It enables records management, archiving and preservation; and
- It helps us to understand whether something is authentic and to trace its provenance.

It also tells us who owns the intellectual content and who can reuse it and under what conditions. It provides a surrogate description for the “thing” itself and helps us to make decisions. It helps us to track changes over time to maintain access, to ensure the integrity of the information and to ensure that “legal evidence” (records) is maintained according to legal requirements (records management). It also tells us something about the thing’s relationship to others information assets such as whether it is part of a series, a serial or a multimedia object.

Again, when faceted structures are poorly designed, poorly implemented or not maintained they can create significant liabilities for an organization, especially, in this case, due to their extensive use in organizations. Liabilities may result when facets are confounded or fail to provide adequate guidance for use. Liabilities also result where there is scope or scale creep over time – where new facets are added to suit unique needs and the overall structure is corrupted due to lack of governance. Governance groups may be established at an expense to the organization but have no enforcement power. Facets may reflect a national or international standard but have little value to the organization. Facets may be reverse engineered over time resulting in a lack of trust of the integrity of the structure. The faceted scheme may impede effective semantic interoperability across applications – a proliferation of disjointed metadata applications impedes semantic interoperability. Where the structure does not support the full information life cycle – cradle to grave – we may find redundancies, gaps or conflicting representations. These problems require additional investments to fix. One of the most common liabilities is poor quality or insufficient granularity of values. Here the initial investment is made, but the expected value is not realized.

In conclusion, value is a concept that may be either positive or negative. While taxonomies hold great positive potential, they may also generate negative values. Organizations embarking on taxonomy projects should be aware of possible assets and liabilities and approach the work with open eyes.
Economics of Knowledge Organization Systems

BEDFORD, continued

Question 2: What are the kinds of costs involved in producing a taxonomy, and how does the level or type of taxonomy affect them?

Types of costs and levels of upfront and recurring investments in taxonomies are important economic considerations.

Observations on levels. As with assets and liabilities, costs and levels of investment will vary with the type of taxonomy (Table 1). For discussion purposes, let’s assume that we’re making good choices and that the taxonomies we’re creating will be optimized.

As a general rule, a flat list will require the least amount of investment. It is a simple structure to maintain, and the options are limited by definition. Chances are the values are well known, and may even be managed as what we call master data sources. Because they are the least costly, we might also assume that they carry the lowest risk. This assumption may not be a good one, however. Risk depends on the nature of the list and where it is used within the organization. For example, if a flat list is used to support a critical access point in a financial or personnel application, failures could have unfortunate consequences.

The next lowest cost taxonomy would be a synonym ring. While a synonym ring requires additional investment in terms of research and maintenance, the structures are simple to manage, and their implementation may be less complex than other structures.

Hierarchies and faceted structures are two of the more costly structures. Their cost levels may be similar but for different reasons. Hierarchies require a lot of consultation, investigation, research and testing to design well. They require careful planning to implement as navigation and browse structures or as embedded sources for parametric search. There are many different ways to apply hierarchies to “things” – sometimes as encoded metadata and sometimes as values in indexes; sometimes generated by human classifiers and other times by automated classifiers. Managing dynamic classification schemes can also add non-trivial update, management and maintenance costs.

Faceted structures require non-simple applications to implement. They may be simpler to design than hierarchies, but the costs to implement and apply are much higher. Each facet has a different rule set, requires different training or different business rules and may require additional complex application support.

Perhaps the most costly type of taxonomy to develop, though, is a thesaurus or semantic network. Each concept has to be identified and validated; each relationship must be identified and validated. Covering a full domain requires extensive consultation with and feedback from experts. Validation requires additional human labor – on the part of information professionals, experts and application and system developers. Perhaps the most important cost element, though, is the dynamic nature of a thesaurus or semantic network. This type of taxonomy challenges the idea of one-time sunk costs. There may be a critical mass of concepts that may be used to calculate initial investments. However, the recurring costs of managing a truly living thesaurus or semantic network may be as great or greater than any initial investment. A thesaurus or semantic network should reflect the current and past language of the organization. Language changes every day – thesauri require continuous reinvestments.

Production Costs of Taxonomies. Organizations embarking on a taxonomy effort should think in terms of three kinds of costs: (1) direct costs, (2) indirect costs and (3) recurring costs. Direct costs are costs that can be entirely attributed to the production of the taxonomy, for example materials, labor, technology and any other expenses incurred in its production. Other costs such as depreciation or administrative expenses are more difficult to assign to a specific product and therefore are considered indirect costs [11].

Indirect costs are costs that are not directly attributable to the production of the taxonomy but are incurred by the organization. These are costs that may be incurred and shared by other products as well, such as the expenses related to organizational administrative personnel, payroll services, facilities, security and so on. These expenses are likely fixed indirect costs because they may in part be associated with the project. Recurring indirect costs are those that result from providing ongoing support activities such as payroll or human resources. Indirect costs for simple taxonomies such as flat lists or synonym rings will be largely invisible or miniscule because the portion of such resources they require in time and effort is miniscule. This situation does not apply, though, for hierarchies, faceted structures or thesauri. Their indirect costs will be visible because of the size of the efforts. In the case of a thesaurus or a semantic network, indirect costs are
likely to be noteworthy because it is difficult to bundle them into other products. For example, a team working on a thesaurus generally does not work on other products or projects.

Recurring costs are those that will continue after the initial taxonomy has been created [11]. As we mentioned earlier, in order to distinguish recurring from non-recurring costs for taxonomy production, we have to be able to say at some point that a taxonomy is finished. Setting an end-point is entirely possible for flat and ring taxonomies. It is doable for hierarchies and faceted taxonomies. In my experience, it is very difficult to do for a thesaurus or a semantic network.

The actual costs of any taxonomy project are tied to its organizational context and the scope and scale of the effort. It is not possible or advisable to say that a typical thesaurus project can be completed for $100,000 or for $500,000 because there is no “typical thesaurus.” Table 1, though, serves as a guide to understanding what kinds of costs might be incurred and what their relative levels might be.

**TABLE 1. Types and Levels Production Costs for Taxonomies**

<table>
<thead>
<tr>
<th>TYPE OF COST</th>
<th>FLAT</th>
<th>RING</th>
<th>HIERARCHY</th>
<th>NETWORK</th>
<th>FACETED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Research &amp; Development</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Communication</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Technology</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Design</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Implementation</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Indirect Costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative Overhead</td>
<td>Miniscule and invisible</td>
<td>Miniscule and invisible</td>
<td>Visible</td>
<td>Visible</td>
<td>Visible and noteworthy</td>
</tr>
<tr>
<td>Payroll Services</td>
<td>Miniscule and invisible</td>
<td>Miniscule and invisible</td>
<td>Visible</td>
<td>Visible</td>
<td>Visible and noteworthy</td>
</tr>
<tr>
<td>Facilities</td>
<td>Miniscule and invisible</td>
<td>Miniscule and invisible</td>
<td>Visible</td>
<td>Visible</td>
<td>Visible and noteworthy</td>
</tr>
<tr>
<td>Security</td>
<td>Miniscule and invisible</td>
<td>Miniscule and invisible</td>
<td>Visible</td>
<td>Visible</td>
<td>Visible and noteworthy</td>
</tr>
<tr>
<td>Recurring Costs</td>
<td>Nonexistent to miniscule</td>
<td>Nonexistent to miniscule</td>
<td>Low</td>
<td>Low</td>
<td>High – as high as initial costs</td>
</tr>
</tbody>
</table>

**Question 3: Do taxonomies produce externalities? Are they positive or negative?**

An *externality* is defined as an incidental effect produced by economic activities but not accounted for by market systems. Externality is an important economic consideration because it means that the cost or benefit of a good produced occurs to individuals who are not engaged in the market transaction for it. In the case of taxonomies and under this definition, if there are externalities, someone other than the producer or the consumer will benefit from existence of the taxonomy. Positive externalities mean that the result is a benefit, such as education for all. Negative externalities mean that there is a liability, such as pollution, that affects a third party [11].

Each type of taxonomy we’ve considered produces externalities. When they are well designed, implemented and managed those externalities will be positive – to the whole organization. When they are poorly conceived, there may or may not be negative externalities. When a taxonomy such as a classification scheme, a metadata structure or standard, or a thesaurus is available to the public – we can say that the probability of positive externalities is high. However, this availability also means that the producing organization is likely uncompensated for the benefit. This type of externality may be addressed by opening up the on-going development and maintenance of a thesaurus to those who use it (for example, making it an open source, collaboratively developed thesaurus).

**Question 4: Do taxonomies have public good characteristics?**

A public good is a product that one individual can consume without reducing its availability for others to consume. Economists refer to public goods as *non-rivalrous* and *non-excludable*. An easy way...
of understanding this idea is to reference Thomas Jefferson’s quote — “He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper [candle] at mine, receives light without darkening me.”[17] Where taxonomies are publicly available and shared, they have public good characteristics. Use of taxonomy does not diminish its economic value. Rather it increases it, particularly where that additional use serves to improve the original product. Free-rider problems can result, though, and discourage others from contributing to the provision of the good to the extent that it is cost free. We would expect that free-rider problems would be significant for taxonomies that are most the costly to produce and maintain, simply because no cost is a compelling incentive for other people to use it.

**Question 5: Do taxonomies have opportunity costs?**

An opportunity cost is the cost of an alternative that must be foregone in order to pursue a certain project or task. What other benefits might the organization have achieved had it not invested in the taxonomy project? The more complex and costly taxonomy products – hierarchies, faceted and network structures – have very high opportunity costs because there appear to be so many substitutes for creating comparable value. There may not, in fact, be substitutes – in my experience clustering or dynamic classification tools are not equivalent assets to hierarchical taxonomies; however, we have generally not been effective in making those arguments to senior management. Opportunity costs are significant where a taxonomy project has been mounted and has not realized its potential. Management will not be inclined to reinvest but will find another alternative. The more costly the taxonomy effort, the greater the opportunity costs.

**Thinking like an Information Economist**

Most economists focus on traditional micro or macroeconomics. As we speed forward into the knowledge economy, we need to think and talk more like knowledge and information economists. That means that information and knowledge scientists need to understand and advance the economics of information and knowledge. In adopting these mental models and ways of thinking, we can do much to advance the cause of our work and increase understanding of the value of the products and services we create.

Thinking and talking in these terms also helps us to be more mindful of the kinds of economic decisions that need to be made. It is no longer acceptable to simply decide we “should have” a classification scheme or a thesaurus because they’re good things to have. Organizations should only in invest in taxonomies when there is a clear and present value to the organization.

How do we make sure that the taxonomies we create are good investments for the organization? How do we ensure that they have economic value? An economic model should be part of the project plan and the proposal to senior management. An economic model should be front and center in managing the taxonomy. Arguments for enterprise taxonomies should be grounded in economics not just data or information management principles. Any decision to invest in taxonomies should be made with clear economic expectations. Such an argument does not mean only focusing on costs, but rather understanding value as assets and liabilities. If we make the decision to invest, we need to understand the nature of the product, plan for the long haul and invest adequate resources while considering both recurring costs and the types of resources needed to support the product.
Resources Mentioned in the Article, cont.


