INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
Intell. Sys. Acc. Fin. Mgmt. (2009)
Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/isaf.309

A COMPARATIVE ANALYSIS OF THE EVOLUTION OF A TAXONOMY FOR BEST PRACTICES

DANIEL E. O'LEARY*
University of Southern California, 3660 Trousdale Parkway, Los Angeles, CA 90089-0441, USA

SUMMARY

Taxonomies play an increasingly important role in knowledge management of business best practices, providing a basis by which to index, find and communicate knowledge. However, knowledge continues to evolve over time. As a result, taxonomies must also continue to evolve as organizations innovate and change.

Reportedly, firms customize best-practice taxonomies to meet their unique organization needs. Accordingly, we might expect organizations to generate dissimilar best-practice taxonomies. However, taxonomies must also reflect the state of knowledge in the area being categorized, and thus are likely to be similar in many ways in different organizations.

The purpose of this paper is to study how taxonomies change in different organizations and how they stay the same. In order to explain the parallels in organizational taxonomies, the notion of 'knowledge artefact efficiency' (or knowledge efficiency) is suggested to capture the concept that new knowledge is rapidly adopted by many organizations in their knowledge management systems. Copyright © 2009 John Wiley & Sons, Ltd.

1. INTRODUCTION

Best practices are among the "better" ways that organizations perform particular processes. Because of the importance of best practices, firms capture those best practices in knowledge bases, indexed using 'taxonomies'. Standard taxonomies allow organizations to communicate with other organizations in order to benchmark their performance. However, organizations also have motivations to change the taxonomy to meet their unique industry requirements, firm-specific requirements and to innovate. In addition, knowledge changes over time; and if the taxonomy is to be useful, then it also needs to change. Accordingly, there are pressures to maintain the same taxonomy, but also pressures to change it.

1.1. PURPOSE

The purpose of this paper is to study how taxonomies change in different organizations and how they stay the same, in order to understand knowledge change and innovation better. Accordingly, this paper provides a comparative analysis of taxonomy evolution in two different organizational settings. In this paper I find that, given the same starting point, two organizations can follow apparently radically divergent paths over time in the development of taxonomies. As an example, structurally, using the same starting point, namely a taxonomy with 271 items, one organization developed a taxonomy with 217 items while another developed a taxonomy with 520 items.

* Correspondence to: D. E. O'Leary, University of Southern California, 3660 Trousdale Parkway, Los Angeles, CA 90089-0441, USA. E-mail: oleary@usc.edu

However, I also find that there are some similarities, particularly semantic similarities, in the two evolved versions. Those similarities appear to be driven by general changes or stability in best-practices knowledge. These similarities lead to the notion of 'knowledge efficiency', where organizations rapidly adopt knowledge innovations into their knowledge management systems. In addition, the paper tracks best-practices taxonomy knowledge change and innovation and relates it back to an overall metric of knowledge change, namely numbers of Google pages.

1.2. Importance of the Research

Understanding knowledge change and innovation is a key issue of knowledge management and business processes. The notion of knowledge efficiency is an important characteristic of knowledge that applies to knowledge management in general, but seems particularly critical in the evolution of taxonomies, and ultimately ontologies and vocabularies.

Further, knowledge of best practices is an important asset. Organizations need to be able to communicate internally about those best practices and externally in order to benchmark with other firms. However, that knowledge of best practices needs to evolve to meet firm- and industry-specific needs, as well as in response to changes in knowledge innovation. Knowledge bases need to evolve over time to accommodate new knowledge and retire old knowledge. Unfortunately, there is limited knowledge about evolution of knowledge bases and taxonomies. For example, it is not clear what drives the need for change or when such taxonomies should be updated. This paper analyses these issues, with the investigation of evolved versions of two organization's taxonomy frameworks.

1.3. This Paper

This paper proceeds in the following manner. Section 2 examines taxonomies for best practices, providing some background, uses and some discussion of a particular framework, the process classification framework. Section 3 analyses the need for change and evolution in the taxonomy. Section 4 lays out the empirical approach and summarizes the available data. Section 5 investigates differences between the frameworks. Section 6 presents some assertions as to similarities of evolution change and discusses the results for the assertions. Section 7 provides a summary of the paper, an analysis of drivers of evolution and discusses some extensions.

2. TAXONOMIES AND TAXONOMIES FOR BEST PRACTICES

Taxonomies are hierarchical organization and categorization of information used in a range of activities, including knowledge management and artificial intelligence (e.g. Chandrasekaran, 1983). In particular, Bailey (1994: 6) defines a taxonomy as a classification of empirical entities. In practice, firms use taxonomies as a way to facilitate access to chunks of knowledge (e.g. Simon, 1985), where the chunks are indexed using the taxonomy. That knowledge may be articles, flowcharts or any of a wide range of other formats.

Best practices are also referred to as 'leading practices'. Best practices are knowledge about processes, and can include flow charts, summaries of key learnings, case studies and other materials.

Taxonomies for best practices are used by organizations to organize information about the best practices and keep track of best-practices information, in order to provide a 'roadmap for

DOI: 10.1002/isaf

knowledge' (Leandri, personal communication). Using the roadmap, knowledge is captured and put into the different categories of information so that others may find and access that information. In addition to the storage of information about best practices, such taxonomies provide other capabilities.

- At least one firm (Muzikowski, personal communication) uses a process classification framework to facilitate workers' understanding of how what they do interacts with other systems. Facilitators guide workers through the framework, in order to help design new systems.
- Further, classifications of best practices also have been referred to as providing an 'International Business Language' (Price Waterhouse, 1998). Using this common 'language', firms can talk to and compare themselves with other firms about their business.
- In addition, if a number of firms all use the same taxonomy, then they will be in a position to summarize results in similar manners and benchmark against other firms.

2.1. APQC Process Classification Framework

Arthur Andersen and the American Productivity and Product Center (APQC) developed the 'process classification framework' (PCF) in 1992, finishing with a published version in 1996. The PCF was used as the basis to organize the best-practices knowledge.

> The process classification framework is the primary framework that Global Best Practices uses to organize best practice tools and information—it is, in essence, the table of contents for the knowledge base. (http://www.globalbestpractices.com/SiteDocs/default.asp?navid=13)

One of the unique aspects of the original process classification framework is that it uses a structure similar to Porter's value chain (Porter, 1985) to organize information about best practices. Porter's value chain (Figure 1) partitions an organization into nine processes. However, the process classification framework (Figure 2) partitions an organization into 13 processes. Porter's value chain differentiates between *primary* activities (e.g. operations, service, and marketing and sales) and *support* activities (e.g. human resources). Similarly, the process classification framework differentiates between

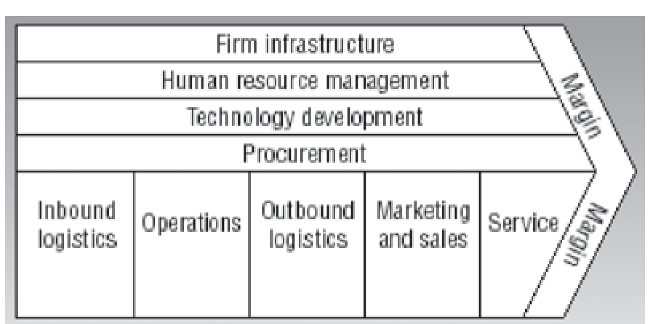

Figure 1. Porter's value chain (*Source*: Porter, 1985)

DOI: 10.1002/isaf

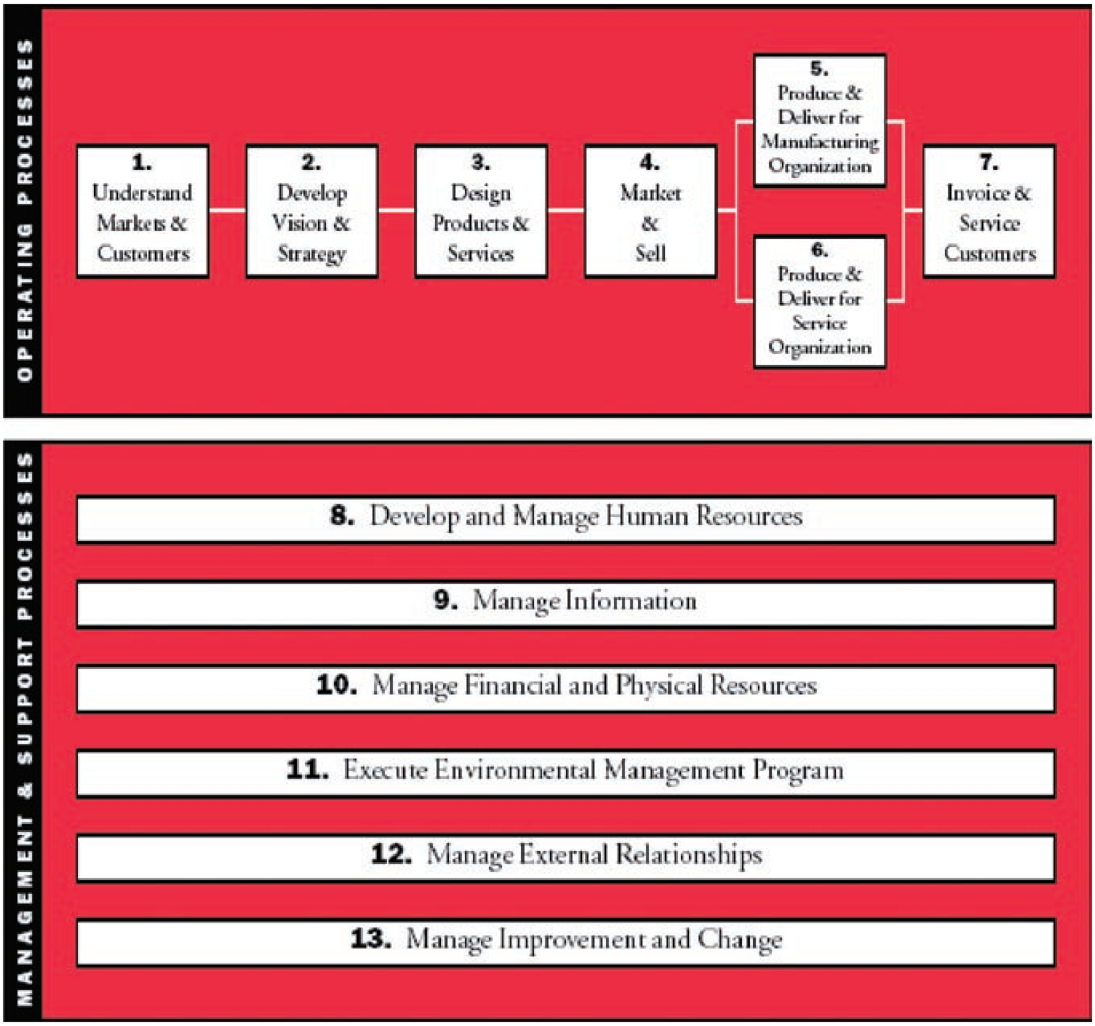

Figure 2. APQC and Arthur Andersen process classification framework (*Source*: APQC, 1996)

operating processes (e.g. market and sell) and *management and support* processes (e.g. develop and manage human resources).

The taxonomy has multiple levels of concepts. In most of the original taxonomy there are three levels, starting with broad concepts and drilling down to greater detail, e.g.:

2.0 Develop vision and strategy
  2.1 Monitor the external environment
    2.1.1 Analyse and understand the competition.

Best practices and metrics can be stored at each level of the framework, with the framework providing the indexing necessary to categorize the best practices.

DOI: 10.1002/isaf

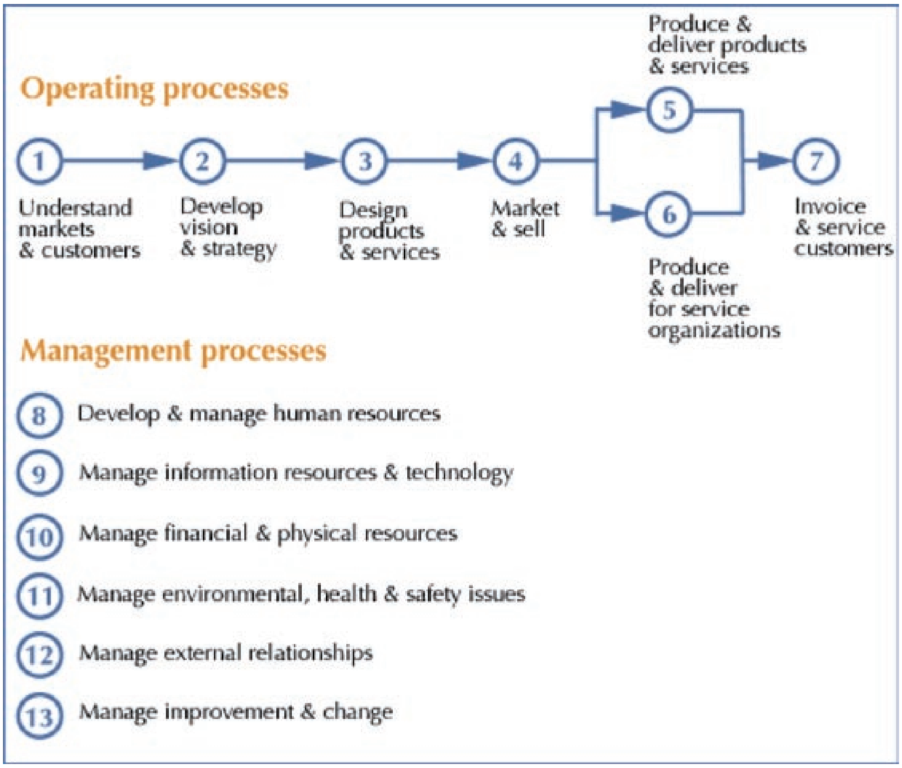

Figure 3. PWC process classification framework (*Source*: PWC, 2008)

2.2. PWC Process Classification Framework

After its development, Arthur Andersen used the framework as a basis for providing an Internet-based knowledge base of best practices (Figure 3). The PCF was maintained to provide their clients with information about best practices. Clients paid a yearly fee for use of the knowledge base. PriceWaterhouseCoopers (PWC) took over the Internet presence (http://www.globalbestpractices.com/Best_practices/) after the demise of Arthur Andersen.

2.3. XYZ Taxonomy

XYZ is a pseudonym for the real company's name. XYZ is a Fortune 100 manufacturing firm. They have developed their own version of the PCF to meet their particular needs. For example, XYZ uses the taxonomy as an 'enterprise architecture' (Bates, personal communication). Accordingly, they use the architecture to help bring together the organization's knowledge, systems and processes.

2.4. Taxonomy Evolution: Some Previous Research

There has been limited research in the analysis of taxonomy evolution and related issues. Feiler (1993) investigated the evolution of terms related to 'reengineering', gradually expanding an existing

DOI: 10.1002/isaf

taxonomy. Alberdi and Sleeman (1997) developed a system designed to facilitate change and revision in a taxonomy. Van Deursen and Visser (2002) suggested that Wikis could be used to facilitate change in a taxonomy. Sherif *et al*. (2004) examined how an initial taxonomy can be developed to facilitate communications. In the closest related research, O'Leary (2007) examined different empirical approaches to capturing and measuring evolution of a taxonomy.

3. TAXONOMY EVOLUTION AND CHANGE

There are pressures for both innovation and change and for stability in best-practice frameworks. On the one hand, if taxonomies are stable, then firms can use the taxonomies to facilitate benchmarking and other interorganization communication. On the other hand, over time, organizations find better ways to perform certain processes and business changes. Changes in processes can require changes in taxonomies, because what is being categorized changes.

3.1. APQC Expectations of Evolution

What does APQC expect in terms of organizations making changes to the taxonomy? In some cases organizations change the APQC framework terminology in order to meet industry-specific terminology better. This type of change likely would result in a framework with similar structure and number of items, but with the terminology slightly different. Additionally, organizations get the basic structure from the APQC framework and then flesh it out to meet their unique organizational requirements. This type of change likely would result with increases in portions of the framework that do not provide sufficient detail. Finally, organizations may build more detail into the framework in order to define the work in their area, enhancing the framework. These rationales (Hasanali, personal communication) generally suggest development of larger, possibly more complex evolved versions.

3.2. Evolution to the PWC PCF

When the original PCF was developed, there were plans to evolve it over time:

> The Process Classification Framework is an evolving document. The Center will continue to enhance and improve it on a regular basis. (APQC, 1996: 3)

Further, as noted by PWC, the framework has changed over the years since development of the original model:

> We have adapted the framework over the years to reflect changes in the business community. (http://www.globalbestpractices.com/SiteDocs/default.asp?navid=13)

The revised PWC version has been driven by needs and expertise of Arthur Andersen and PWC and their clients. In particular, as noted by Leandri (personal communication) the PWC version is 'market driven', as PWC tries to provide one-stop shopping for knowledge about best practices. Formerly at Arthur Andersen and currently at PWC, case writers would ask individual experts (individual agents) in the firm to provide their input regarding the PCF. The case writers would summarize their remarks. As a result, individual agents in different functional areas also drove changes.

DOI: 10.1002/isaf

Further, discussion with the Partner in Charge of the Project indicated that many of the changes reflect changes in our state of knowledge. For example, human resources got more ‘granular’ (more detail) and information technology has changed over the recent years (Internet, World Wide Web, Enterprise Systems, etc.). As knowledge changes, the taxonomies that categorize them need to change.

It could be argued that PWC also has an incentive to reduce the size of the framework. Much of their framework is populated with diagrams and key performance indicators. Maintaining those materials can be costly. Everything being equal, the fewer the categories, the less the amount of maintenance that needs to be done, given that market needs were met. Accordingly, this would suggest that PWC requirements could lead to a decrease in the number of concepts.

3.3. Evolution to the XYZ PCF

At XYZ, changing the process classification framework also was decentralized to individual functional areas, placing the responsibility for change with different agents. One of the key decisions made by XYZ was to drop an entire category, ‘6. Produce and deliver for a service organization’, because XYZ saw themselves as a manufacturing organization, not a service organization. XYZ sought changes for their taxonomy from process owners (individual agents), but coordinated those changes centrally.

4. METHODOLOGY

The methodology used in this paper was to gather multiple taxonomies and compare them. The taxonomies developed by APQC (1996), Price Waterhouse (1998) and XYZ’s versions provide the data used in this analysis.

4.1. Approach

I used a two-pronged approached aimed at capturing both semantic and structure information about the taxonomies. In particular, in order to capture semantic information, for each category and subcategory of the revised versions, the corresponding category from 1996 was matched. In a few cases this meant mixing second level in one version and third level with another version. The similarity of the categories was categorized according to one of the following.

Identical:	Wording used was identical
Very similar:	Wording is virtually the identical, but there are additional descriptors, e.g. ‘Assess technology innovations’ versus ‘Assess *new* technology innovations’
Similar:	Basic concept is roughly the same, e.g. ‘Monitor competitive offerings’ versus ‘Determine customer reactions to competitive offerings’
No matching concept from old to new:	No concept in the old framework that matches the new framework
No matching concept from new to old:	No concept in the new framework that matches the old framework

What was the nature of the matches that were not identical? The 'Very similar' matches would include or exclude a word that would not change the meaning. For example, the original category '2.1.4 Assess new technology innovations' morphed into '2.1.4 Assess technology innovations'. The very similar matches also included changes where one word is replaced by another very similar word. For example, the original category '1.3.1 Determine the weaknesses of product/service offerings' morphed into '1.3.1 Determine deficiency of products and services'. Further, the 'Very similar' category also included aggregations or disaggregations. For example, the original '10.5.1 Ensure tax compliance' was broken down into '10.5.1 Manage national tax compliance', '10.5.2 Manage local tax compliance' and '10.5.3 Manage international tax compliance'. For the 15 changes rated 'Very similar', four included a word change, nine included additional or fewer words and two were aggregations or disaggregations.

Two categories of nonmatches were generated. Those concepts that were only in the old framework were captured as 'old only'. Those concepts that were only in the new taxonomy were captured as 'new only'. Redundancies (aggregations and disaggregations) from old to new and new to old were also captured 'More than one new' and 'More than one old'. The resulting data was analysed and summarized in a number of different ways (e.g. by category, changes and matches).

In addition, different structural characteristics of the taxonomies were analysed and compared (e.g. number of items and number of paths).

Finally, outside information about the amount of knowledge in the world (from the World Wide Web) was also generated. In particular, the number of pages relating to particular taxonomy terms among the 13 top-level concepts (Figure 2) was captured in order to estimate the change in knowledge in business processes over the relevant time-frame.

5. STRUCTURAL DIFFERENCES BETWEEN THE FRAMEWORKS

The number of items in each of the three taxonomies is summarized in Figure 4 and Table I, by the 13 top-level categories in the APQC framework. The correlation coefficients in Table II suggest that the number of subcategories in the APQC version are not correlated (Table III).

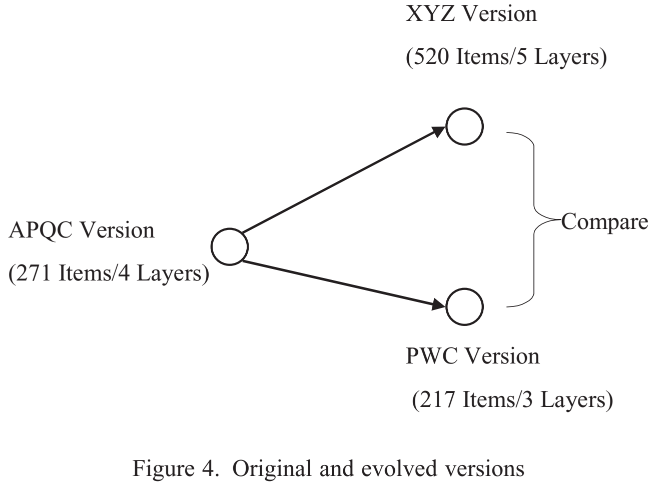

Figure 4. Original and evolved versions

DOI: 10.1002/isaf

Table I. Relationship between three taxonomies: number of items by top-level category

APQC	PWC	XYZ
16	12	12
16	19	25
24	22	37
13	15	75
24	22	49
14	13	0
11	10	14
45	21	48
35	12	137
30	31	50
11	15	27
8	9	32
24	16	14
271	217	520

Table II. Correlations between number of items in taxonomies

	APQC	PWC	XYZ
APQC		0.291	0.246
PWC			0.006
XYZ			

Table III. Statistical significance (*t*-test)

	APQC	PWC	XYZ
APQC		0.167	0.209
PWC			0.492
XYZ			

The complexity of trees (the graphic representation of a taxonomy) is typically measured using three different metrics: size, depth and number of paths from the root to the ends of the branches. Using these criteria, the PWC taxonomy is the least complex. It had 217 categories and subcategories: there were 13 top-level, 60 second-level and 144 third-level categories; in addition, there were 163 paths in it. The APQC was between the others in complexity: it had 271 categories and subcategories and had four levels. The XYZ taxonomy was the most complex. It had 520 categories and subcategories: there were 12 top-level categories, 67 second-level, 218 third-level, 210 fourth-level and 15 fifth-level subcategories; in addition, there were 364 paths from tree root to each of the branches.

The number of paths in each of the trees is summarized in Table IV and the correlations between the number of paths are given in Table V. The correlations of the number of paths between, APQC, PWC and XYZ are not statistically significant (Table VI). The two organizations apparently have evolved structurally significantly different taxonomies from the same starting point.

DOI: 10.1002/isaf

Table IV. Number of paths through taxonomy tree

Category	APQC	PWC	XYZ
1	10	8	8
2	13	15	19
3	19	16	24
4	10	12	52
5	19	17	29
6	10	9	6
7	7	6	0
8	37	17	41
9	27	9	109
10	24	25	30
11	10	11	16
12	7	7	20
13	18	11	10
	211	163	364

Table V. Correlations between number of paths in taxonomies

	APQC	PWC	XYZ
APQC		0.351	0.300
PWC			0.012
XYZ			

Table VI. Statistical significance (*t*-test)

	APQC	PWC	XYZ
APQC		0.120	0.160
PWC			0.484
XYZ			

6. SIMILARITIES OF THE FRAMEWORKS: ASSERTIONS AND ANALYSIS ABOUT CHANGES IN KNOWLEDGE

This section analyses various aspects of the original and revised frameworks, which allows us to assess the assertions associated with evolution of a taxonomy from one version to another. Although the PWC and XYZ taxonomies appear structurally quite different from each other, this section investigates some assertions of similarity. In particular, it appears that the taxonomies are similar in a number of ways, apparently because of knowledge efficiency.

It has been suggested that the financial markets, such as the New York Stock Exchange, are very 'informationally efficient'. Information about financial instruments, such as stock, is rapidly embedded in the price of the instrument. As information about or related to a firm and its performance is revealed, the impact of that information is embedded in the price of the instrument. This has led to the so-called 'efficient markets hypothesis' (e.g. Fama, 1965).

DOI: 10.1002/isaf

In the same sense, I make a similar hypothesis that knowledge management artefacts, such as taxonomies, are also 'efficient', referred to as the 'efficient knowledge artefacts hypothesis' or simply just the 'efficient knowledge hypothesis'. Information will be embedded into those knowledge artefacts, such as best-practices taxonomies, as or 'if' they evolve along with the evolution of knowledge in the area. If that hypothesis is true, then organizations that develop taxonomies at the same time for the same purposes are likely to have much of the same information in them (i.e. *match*). Similarly, organizations that start with the same taxonomy and develop a taxonomy at the same time, for the same purposes, are likely to make similar *change*s in their taxonomies in response to general changes in knowledge.

These similarities are summarized as a set of assertions that will then be analysed empirically. Each of these assertions is made assuming all other things are equal. Through each of these four assertions, it is assumed that innovations and changes in knowledge can drive changes in taxonomies.

6.1. Assertion 1: Knowledge that Stays the Same in the Two Evolved Taxonomies is Correlated

In order to address this issue, I will analyse the extent to which the number of subcategories in the new evolved PWC and XYZ versions stays the same as in the original APQC version. The matches by category are summarized in Table VII for both APQC versus PWC and APQC versus XYZ. The total number of matches, 159 in panel A (APQC versus PWC) and 169 in panel B (APQC versus XYZ), are very similar, in spite of the fact that the XYZ taxonomy was so much larger than the PWC taxonomy (520 to 217 items). This suggests that the existing knowledge in the APQC framework used by both PWC and XYZ are similar.

Rather than just compare the aggregate similarities, the total number of matches in the top level categories in Table VII, panel A and panel B were compared. The resulting correlation coefficient of 0.790 was statistically significant at 0.0006 using a *t*-test. These results suggest that, although the resulting taxonomies are different, there is a core of categories that was kept by both PWC and XYZ.

As a result, what I find is that knowledge stays the same in both organizations is highly correlated. One explanation is that both organizations recognize that particular correlated knowledge has not changed. Thus, both the resulting taxonomies reflect the current state of knowledge about best practices, consistent with the notion of knowledge efficiency.

6.2. Assertion 2: Knowledge that Changes in the Two Evolved Taxonomies is Correlated

In order to investigate assertion 2, I will analyse the changes in the taxonomies to determine the extent to which changes from the APQC version to the PWC and XYZ versions are correlated. Changes include addition of new categories and deletion of old categories.

The category changes are summarized in Table VIII. As with the matches, the number of old categories that were removed was similar in the PWC and XYZ frameworks: 113 and 102 respectively. Further, the total numbers of old items not used by PWC and XYZ were compared and I found a correlation of 0.9004, significant at the 0.001 level, using a *t*-test.

As noted in Section 5, the structural complexity of the PWC and XYZ taxonomy structures is substantially different. Firm and industry differences likely account for that substantial increase by XYZ over PWC. For example, XYZ dropped category 6 entirely and greatly expanded category 4 (marketing) and category 9 (information systems). PWC decreased the number of categories.

DOI: 10.1002/isaf

Table VII. Matches by category

Category	Identical	Very similar	Similar	Total matches[a]	Total
Panel A: APQC versus PWC					
1	7	3	2	12	12
2	10	3	0	13	19
3	15	1	1	17	22
4	4	0	3	7	15
5	13	2	4	20	22
6	11	0	1	12	13
7	9	0	1	10	10
8	10	1	5	16	21
9	0	2	1	3	12
10	21	1	2	24	31
11	0	0	1	1	15
12	7	1	0	8	9
13	14	0	2	16	16
	121	14	24	159	217
Panel B: APQC versus XYZ					
1	5	2	5	12	12
2	10	3	2	15	25
3	13	4	4	21	37
4	3	0	9	12	75
5	10	4	5	19	49
6	0	0	0	0	0
7	8	0	3	11	14
8	19	1	6	26	48
9	1	0	5	6	137
10	20	0	4	24	50
11	0	0	1	1	27
12	4	0	4	8	32
13	13	1	0	14	14
	106	15	48	169	520

[a] Correlation of 'total matches' Panel A : Panel B, 0.790, *F*-test significant at 0.0013.

However, in spite of these individual and industry differences, the total numbers of new items added by PWC and by XYZ were compared and I found a correlation of 0.548, which was significant at the 0.0262 level. Finally, the total number of changes by PWC and XYZ had a correlation coefficient of 0.7908, significant at the 0.0064 level.

Further, the changes by PWC and XYZ did vary by the type of process, in particular, operating processes, as opposed to management processes. In PWC's framework, there were 217 items, with 113 items in the 'operating processes' (categories 1–7), and 114 items in 'management and support' processes (categories 8–13). However, roughly 30 % of the change occurred in the seven operating processes, whereas roughly 70 % of the change occurred in the six management processes. Similarly, in XYZ's taxonomy, roughly 33 % of the change occurred in the seven operating processes, whereas roughly 67 % of the change occurred in the six management processes (categories 8–13). Based on these results, it appears that knowledge about a firm's operating processes have been more stable than management and support processes. For example, changes have had a major impact on information technology since 1996. Technology, including the Internet, e-business and e-commerce, has become routine and integrated into many leading businesses. As a result, it is probably not surprising that for both PWC and XYZ the biggest change has occurred in category 9, 'Manage information'.

DOI: 10.1002/isaf

Table VIII. Changes by category

Category	Total new[a]	Total old[b]	Total changes[c]
Panel A: APQC versus PWC			
1	0	4	4
2	6	3	9
3	5	7	12
4	8	6	14
5	2	5	7
6	1	2	3
7	0	1	1
8	5	29	34
9	9	32	41
10	5	6	11
11	14	10	24
12	1	0	1
13	0	8	8
	56	113	169
Panel B: APQC versus XYZ			
1	0	4	4
2	10	1	11
3	16	3	19
4	63	1	64
5	30	5	35
6	0	14	14
7	3	0	3
8	22	20	42
9	131	29	160
10	26	5	31
11	26	10	36
12	24	0	24
13	0	10	10
	351	102	453

[a] Correlation 'total new' Panel A : Panel B, 0.548, *t*-test significant at 0.0262.
[b] Correlation 'total old' Panel A : Panel B, 0.9004, *t*-test significant at 0.001.
[c] Correlation 'total changes' Panel A : Panel B, 0.7908, *t*-test significant at 0.0064.

One explanation for these results is that knowledge is efficient and that both organizations knew about changes that were occurring and wanted to embed them in their taxonomies. As a result, the taxonomies reflect the current state of knowledge about best practices, consistent with the notion of knowledge efficiency.

6.3. Assertion 3: Knowledge that stays the Same is Correlated with the Parts of the Taxonomies Populated with Chunks of Knowledge

Within the business process taxonomies that I have analysed, taxonomy categories can be populated with chunks of knowledge, such as measures, diagrams and case studies. If a category is populated with materials, then that provides one measure of the extent to which that category is understood or fleshed out, when compared with a category that is not populated. Further, category population might be seen as a measure of the extent of 'commitment' to the category, since developing, populating and indexing artefacts requires cost and effort. Firms are not likely to heavily populate a chunk of knowledge that is in transition, because it would be a wasted cost. As a result, if cost is a factor, then those that are populated are also likely to be continued over time and, thus, be relatively stable.

DOI: 10.1002/isaf

Table IX. Number of populated categories in original version versus number of matches[a]

Category	APQC no. of measures	Total matches	
		PWC	XYZ
1	16	12	12
2	4	13	15
3	61	17	21
4	39	7	12
5	150	20	19
6	0	12	0
7	28	10	11
8	91	16	26
9	54	3	6
10	154	24	24
11	61	1	1
12	0	8	8
13	5	16	14
	663	159	169

[a] Correlation APQC number of measures and PWC total matches is 0.525, significance of *t* is 0.032. Correlation APQC number of measures and XYZ total matches is 0.594, significance of *t* is 0.016.

As seen in Table IX, in the APQC framework there were 663 measures provided across the 13 major, top-level categories. In order to substantiate assertion 3, we would expect that the APQC measures would be correlated with both the number of matches between the APQC and PWC taxonomies and the APQC and XYZ taxonomies. In particular, I found that the APQC measures and PWC measures had a correlation of 0.525, which was significant at the 0.032 level. Further, I found the APQC measures and XYZ matches had a correlation of 0.594, with a significance of 0.016. Thus, the results are consistent with assertion 3.

Thus, for both taxonomies the knowledge that stays the same is correlated with the populated knowledge, the knowledge to which both firms are committed, the knowledge that appears the most stable. This is consistent with the notion of efficient knowledge artefacts, since both firms have committed to correlated knowledge.

6.4. Assertion 4: Changes in General Knowledge of Business Processes is Correlated with Changes in the Best-Practices Taxonomies

If we had a metric of total change in best-practices knowledge, since the APQC framework was developed, then the amount of that change should be related to changes in the frameworks to the PWC and XYZ frameworks. Accordingly, in order to investigate assertion 4, I will analyse the extent to which the changes to the APQC framework in the PWC and XYZ taxonomies are correlated with a measure in the change of knowledge about the best practices.

Increasingly, the World Wide Web is becoming the source of information about knowledge. Perhaps the best-known search engine is Google.com. An important feature of Google.com is that it indicates

DOI: 10.1002/isaf

Table X. Number of changes versus change in knowledge

Category	Google pages measure[a]		Changes	
	PWC	XYZ	PWC	XYZ
1	29,600,000	27,900,000	4	4
2	5,640,000	5540,000	9	11
3	15,900,000	16,000,000	12	19
4	16,900,000	10,900,000	14	64
5	513,000	248,0000	7	35
6	396,000		3	14
7	361,000	360,000	1	3
8	45,800,000	45,500,000	34	42
9	26,800,000	28,800,000	41	160
10	3,440,000	23,500,000	11	31
11	38,000,000	11,800,000	24	36
12	5,500,000	5,490,000	1	24
13	5,500,000	5,500,000	8	10

[a] Google pages measure gathered April 2008. Correlation between Google pages and changes for PWC is 0.741 significant at the .002 level, and correlation between Google pages measure and changes for XYZ is 0.372, significant at better than 0.105.

how many Web pages it finds with information relating to each specific query. The number of Web pages provides a metric as to how much knowledge there is on the Web regarding the query.

If we assume that the number of Web pages in 2008 was substantially larger than the number of Web pages in 1996, then the total number in 2008 approximates the number of new Web pages since 1996. According to Gray (1996), there were approximately 100,000 Web sites in 1996. As of May 2009, a Google search of ‘html’ alone yielded over 16 billion Web pages. As a result, the total change since 1996 is close to the current number of pages.

Accordingly, I typed each of the top-level titles for the PWC and XYZ versions in the taxonomy into the Google search engine as key words in order to generate a measure of Google pages and an estimate of the total knowledge regarding those concepts. The results of those searches are summarized in Table X.

The correlations between the Google page measure for both the PWC changes and the XYZ page changes were 0.741 and 0.372. As a result, these findings provide strong evidence for assertion 4. Thus, change in both taxonomies is correlated with change in the number of Google pages at the top level of the taxonomy. One explanation for this finding is that change in general knowledge about business processes is reflected as change in the organizations’ taxonomies about best practices, which is consistent with the notion of efficient knowledge artefacts.

7. SUMMARY AND EXTENSIONS

This paper has analysed a leading taxonomy for best practices: the evolution of the APQC process classification framework from its original published version to two different versions, one by PWC and one by XYZ. I employed an empirical approach to begin to understand how such taxonomies evolve over time and to begin to understand how organizations impact those taxonomies. In order to explain the findings, the ‘efficient knowledge artefacts hypothesis’ was developed and analysed.

DOI: 10.1002/isaf

7.1. What Were the Sources of Change?

The primary sources of change identified in discussions with APQC were the expectation that industry- and firm-specific changes could force changes to the taxonomies of individual firms. We saw that to a certain extent, particularly with the size and complexity category changes in the structures associated with the PWC and XYZ versions. For example, XYZ dropped one of the major categories because they were a manufacturing organization and not a service organization. Further, XYZ added a large number of categories and complexity across multiple categories (marketing and information systems), suggesting that individual agent differences would not account for the change. Also, PWC reduced the size of the framework, as discussed above.

The discussion with PWC suggested that change in knowledge is an important factor driving taxonomy change. The number of matches between APQC and PWC and APQC and XYZ were found to be highly correlated, suggesting that both PWC and XYZ kept much of the same knowledge in their evolved taxonomies. The number of categories removed from the APQC framework by PWC and XYZ was also highly correlated. Similarly, the number of categories added to the PWC and XYZ frameworks was also correlated, suggesting that the organizations changed the same knowledge in their taxonomies. I also found a relationship between the number of measures used to populate the APQC version and the number of matches between the APQC and PWC frameworks and between the APQC and XYZ frameworks. These relationships suggest a kind of 'momentum' between different versions, with populated portions providing knowledge that is apparently more stable. Finally, I found a relationship between the number of changes in the original framework in both the PWC frameworks and the XYZ frameworks with the number of pages found in Google.com, based on queries from the original APQC taxonomy. This suggests that the changes in the framework mirror changes in knowledge about best practices.

7.2. Limitations

There are at least two limitations of this research. First, this research has focused on the evolution of a single taxonomy in two firms. As a result, it provides a limited sample in a case-based analysis. Second, the research is correlational, so true causation is not proved. However, for each of the assertions listed above, there was strong statistical evidence to support the assertions.

7.3. Contributions

This is the first study to propose 'efficient knowledge artefacts'. This is also one of the first studies to analyse an existing taxonomy/knowledge base empirically from the perspective of how it has evolved in multiple organizations, generating a comparative analysis of that change. That analysis has resulted in a number of findings on the 'behaviour' of knowledge that are summarized above.

In addition, those findings provide the basis to facilitate determination of when a taxonomy should be changed. For example, if investigation of a number of organizations' taxonomies yields a shrinking base of knowledge used across applications or a consistent base of changes, perhaps it is time to change the overall taxonomy so that individual organizations do not have to make the changes themselves. For example, in the APQC taxonomy, 159 and 169 categories out of 271 categories were being used (58 % to 62 % of the taxonomy). Organizations can then analyse the respective costs and benefits of changing the taxonomy to meet their user's needs as the usability of the taxonomy decreases.

DOI: 10.1002/isaf

Further, the types of change that organizations are making should influence the decision for when a taxonomy should be changed. If the changes are made because of individual agents or industry-specific changes or firm-specific changes, then change may not be appropriate, because of the individual nature of the impact. However, if it is found that knowledge changes are driving the change and those knowledge changes are substantial, then change may be appropriate.

Finally, although the resulting taxonomies were so different in complexity, the statistical analysis uncovered many similarities, apparently based on response to changes in knowledge about best practices. Apparently, knowledge changes can drive similar taxonomy evolution in multiple organizations.

7.4. Extensions

This research can be extended to include other organization's taxonomies or evolutions of the process classification framework. In addition, this research and the approach used here can be used to investigate the evolution of other taxonomies. Further, this research may be used as a basis to analyse the extent to which an organization evolves and innovates its taxonomies and how that change relates to other organizations' taxonomies.

ACKNOWLEDGEMENTS

I would like to acknowledge the comments of the anonymous referees.

REFERENCES

Alberdi E, Sleeman D. 1997. RETAX: a step in the automation of taxonomic revision. *Artificial Intelligence Journal* **91**(2): 257–279.

APQC. 1996. American Productivity & Quality Center and Arthur Andersen & Co. *Process classification framework*. http://www.bin.co.uk/pcf.pdf (accessed 22 May 2009).

Bailey K. 1994. *Typologies and Taxonomies*. Sage Publications: Thousand Oaks, CA.

Chandrasekaran B. 1983. Towards a taxonomy of problem solving types. *AI Magazine* **4**(1): 9–17.

Dixon W, Massey F. 1969. *Introduction to Statistical Analysis*. McGraw-Hill: New York, NY.

Fama E. 1965. Random walks in stock market prices. *Journal of Business* **21**(5): 55–59.

Feiler PH. 1993. Reengineering: an engineering problem. Technical report CMU/SEI-93-SR-5, Software Engineering Institute, Carnegie Mellon University.

Gray M. 1996. *Web growth summary*. http://www.mit.edu/people/mkgray/net/web-growth-summary.html (accessed 22 May 2009).

Hogg RV, Craig AT. 1995. *Introduction to Mathematical Statistics*, 5th edition. Macmillan: New York, NY.

O'Leary DE. 2007. Empirical analysis of the evolution of a taxonomy for best practices. *Decision Support Systems* **43**: 1650–1663.

Porter M. 1985. *Competitive Advantage*. Free Press: New York, NY.

Price Waterhouse. 1998. *International business language*. http://www.pw.com/kv/ibltxt.htm (link no longer functional since merger of Price Waterhouse and Coopers & Lybrand).

PWC. 2008. *Process classification framework*. http://www.globalbestpractices.com/Home/ProcessFrameworks.aspx?FW=Process+classification+framework (accessed 22 May 2009).

Sherif K, Hoffman J, Thomas B. 2006. Can technology build organizational social capital? The case of a global IT consulting firm. *Information and Management* **43**(7): 795–804.

Simon HA. 1985. *The Sciences of the Artificial*. MIT Press: Cambridge, MA.

Van Deursen A, Visser E. 2002. The reengineering Wiki. In *Sixth European Conference on Software Maintenance and Reengineering*, Budapest.

DOI: 10.1002/isaf