+ MODEL

ARTICLE IN PRESS

DECSUP-11300; No of Pages 14

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Decision Support Systems xx (2006) xxx – xxx

Decision Support Systems

www.elsevier.com/locate/dsw

Empirical analysis of the evolution of a taxonomy for best practices

Daniel E. O'Leary *

University of Southern California, Los Angeles, CA 90089-0441, United States

Abstract

Taxonomies play an increasingly role in knowledge management, providing the basis on which to find and communicate knowledge, information and metrics. However, knowledge continues to evolve over time. As a result, taxonomies also need to continue to evolve.

Two different evolved versions of a taxonomy for best practices, each based on the same original taxonomy were analyzed. This research investigated empirical approaches to trace the changes in the original taxonomy. In so doing, an approach using empirical findings to monitoring and anticipating taxonomy change is initiated. There were a number of findings, including a tendency to evolve to greater complexity.

Keywords: Taxonomy; Evolution; Knowledge management; Complexity

1. Introduction

Best practices are ways that firms perform particular processes. Increasingly, firms are developing "best practices" knowledge bases as part of their knowledge management systems. Best practices (or leading practices) knowledge bases provide access to knowledge about enterprise processes that appear to define the best ways of doing things.

Firms capture and index those best practices in knowledge bases using "taxonomies." Since best practices change, it is likely that any taxonomy for best practices also would need to evolve over time, if it is to be used over time. However, evolution has been identified as one of the primary impediments to the development and use of taxonomies. Unfortunately, development of different taxonomies can limit the ability of organizations to fully leverage taxonomy capabilities.

1.1. Purpose

The purpose of this research is two-fold. First, this research builds an empirical approach to analyzing taxonomy evolution. It does that by investigating "what changes over time" and "what stays the same." The resulting data are analyzed using a number of approaches and statistical analyzes. For example, how many categories change and how many stay the same, and how those changes are related statistically between an original version and a later version?

Second, because of the relationships it finds, this research begins to build a descriptive model of how best practice taxonomy knowledge changes. For example, how are the number of new categories added and old categories dropped related and does knowledge become more complex as the taxonomies evolve?

* Tel.: +1 213 740 4856; fax: +1 213 747 2815.
E-mail address: oleary@usc.edu.

0167-9236/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2006.09.004

As a result, the questions that this research examines are

- How can one analyze evolution of a knowledge-based taxonomy?
- How can one categorize how the taxonomy has changed?

This research does not investigate the issues as to whether or not the taxonomy is the best framework for categorizing best practices, or whether some other framework would be better. Instead, it is an empirical analysis of the evolution of one of the best-known business knowledge management frameworks in two different organizations.

1.2. Importance of the research

Each question is important for different reasons. First, consider the question, "how can one analyze evolution of a knowledge-based taxonomy?" There are few guidelines to facilitate an understanding of taxonomy or knowledge base evolution. This research provides the beginnings of a road map.

Second, consider the question, "how can one categorize how the taxonomy has changed?" Since knowledge bases need to evolve over time, organizations need to understand, a priori, how those knowledge bases will evolve and how to evolve them. By looking at how knowledge bases change, one can begin to answer this question. Using knowledge garnered from these and other questions, ultimately the goal would be to build a system that could anticipate and facilitate taxonomy evolution.

1.3. Plan of the paper

Section 1 has provided an introduction. Section 2 examines taxonomies for best practices, providing some background, uses and some discussion of a particular framework, the process classification framework. Section 3 analyzes the particular taxonomy investigated in this research. Section 4 lays out the approach and summarizes the available data. Section 5 presents the analysis of the taxonomy data. Section 6 summarizes the primary findings. Section 7 provides a summary of the contributions and discusses some extensions.

2. Taxonomies and taxonomies for best practices

Taxonomies are specifications of terminology in an application or set of applications. Historically, taxonomies have been designed to categorize, in a hierarchical manner, a set of terms or concepts. Taxonomies can facilitate classification, communication and search through a knowledge base.

Best practices are those known ways to perform a business process that are among the best ways to perform that process. Best practices knowledge bases are used by a number of firms, including the professional services firms, as a means for capturing best practices for reuse. Best practices information can include descriptions about the best practices, graphic illustration of the best practice (such as a flowchart) and case studies.

Taxonomies have a long history in knowledge management and artificial intelligence (Chandrasekaran [3]). In addition, taxonomies have played an important role in research on best practices, by providing a basis on which to categorize best practices.

There are pressures for both change and stability in best practice taxonomies. On the one hand, if taxonomies are stable then firms can use the taxonomies to facilitate benchmarking. 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 may change.

3. Process classification framework

There are a number of versions of a taxonomy for best practices analyzed in this research, including American Productivity and Product Center, PriceWaterhouseCoopers and another corporate version.

3.1. American productivity and product center

Arthur Andersen and the American Productivity and Product Center (APQC [1]) developed the "Process Classification Framework" (PCF) starting in 1992, culminating with a published version in 1996 (Fig. 1). 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)

The process classification framework differentiates between *operating* processes (market and sell), and *management* and *support* processes (develop and manage human resources).

The taxonomy has multiple levels of concepts. In most of the original taxonomy there are three levels,

Please cite this article as: Daniel E. O'Leary, Empirical analysis of the evolution of a taxonomy for best practices, Decision Support Systems (2006), doi:10.1016/j.dss.2006.09.004

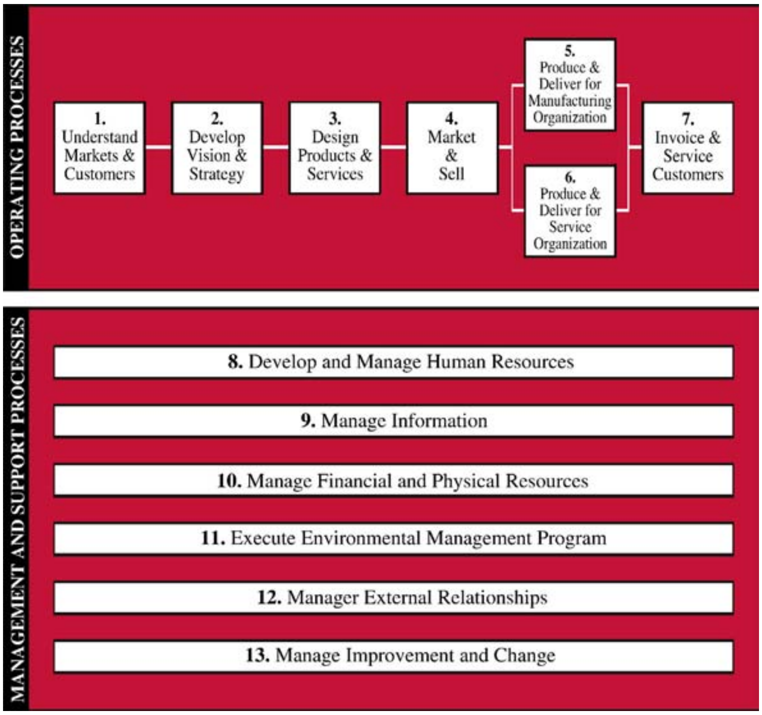

Fig. 1. APQC and Arthur Andersen process classification framework (source: APQC and Arthur Andersen [1]).

starting with broad concepts and drilling down to greater detail, for example,

- 1.0 understand markets and customers
  - 1.1 determine customer needs
    - 1.1.1 conduct qualitative assessments

Best practices, case studies and metrics can be indexed at each level of the framework.

3.2. 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. The PCF was maintained to provide their clients with information about best practices. Clients would pay 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 (Fig. 2).

When the 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 Arthur Andersen [1], p. 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)

3.3. Corporate versions

Apparently a number of corporations have used the APQC version as a jumping off point for generating their own version of a best practices taxonomy. XYZ is a pseudonym for the real company's name for which access was obtained. XYZ is a Fortune 100 manufacturing firm. They have developed their own version of the PCF to meet their particular needs. One of the key decisions made by XYZ was to drop an entire category, "6. Produce and Deliver for a Service Organization," all together, because XYZ sees themselves as a manufacturing organization, not a service organization.

3.4. Selection of these taxonomies

The evolution of this particular taxonomy was selected to research for a number of reasons. First, this taxonomy

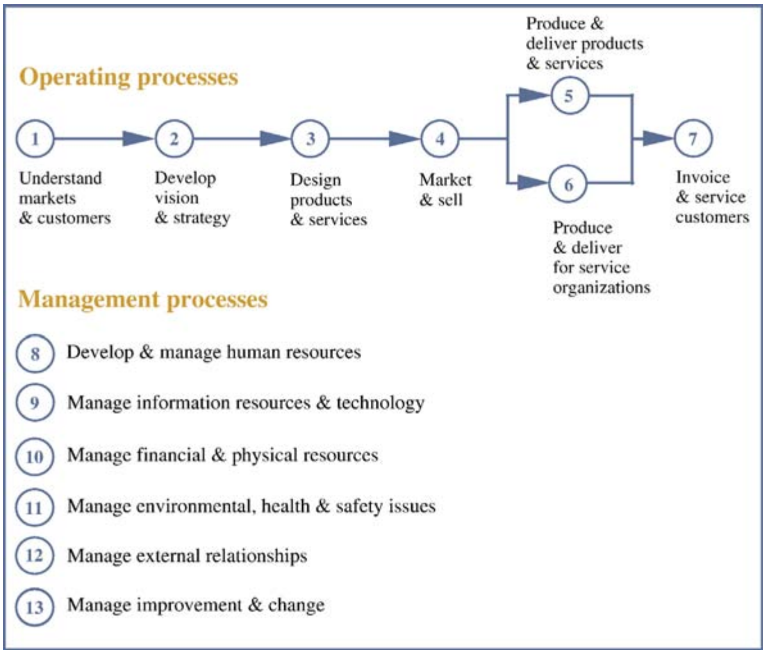

Fig. 2. PWC process classification framework (source PWC [9]).

allows for analysis of the evolution of a taxonomy over time and against other versions, since they are all based on the same original taxonomy. Second, the taxonomy has been reported on as being a "success" and having "longevity" by Gartner (Caldwell [2]). That same report referred to the taxonomy as a "public standard … for organizing and sharing information." Third, the taxonomy provides a basis for best practice sharing by a wide number of major firms including PriceWaterhouseCoopers, Ford Motor Company and SchlumbergerSema (Caldwell [2]). Finally, three versions of the taxonomy were available, including information such as measures associated with parts of the taxonomies used in this comparison between the taxonomies.

4. Data and approach

The process classification frameworks from APQC [1], PWC [9] and XYZ's versions to provide the data used in this analysis.

- In 1996, the APQC taxonomy had 271 categories and subcategories. There were 13 top-level categories, 71 second level concepts, 184 third level concepts and 3 fourth level concepts.
- In 2003, the PWC taxonomy had 217 categories and subcategories. There were 13 top-level, 60 second-level and 144 third-level categories.
- In 2003, the XYZ taxonomy 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.

4.1. Pair wise comparison

The research involved pair wise comparison of the three pairs of taxonomies. APQC was considered the original when compared to PWC and XYZ. PWC was considered the original when compared to the XYZ framework.

4.2. Approach

The categories of APQC and PWC, APQC and XYZ and PWC and XYZ were matched. For each category and subcategory of one framework there was an attempt to match that category with another in the other framework. In a very few cases this meant mixing second level in one version and third level with another version or other levels. 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, such as, "Assess technology innovations" vs. "Assess *new* technology innovations"

Similar Basic concept is roughly the same, for example, "Monitor competitive offerings," versus, "Determine customer reactions to competitive offerings."

Any of those three categories ultimately was considered a "match." However, some categories did not match others. Those categories were characterized using two other groupings, and were characterized as changes:

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 dis-aggregations. 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" in PWC.

Two categories of non-matches were generated. Those concepts that were only in the original APQC framework were captured as "old only." Those concepts that were only in the new taxonomies (PWC or XYZ) were captured as "new only." Redundancies (aggregations and dis-aggregations) from old to new and new to old were also captured "More than one New," and "More than one Old."

4.3. "Level changes"

"Level changes" in a taxonomy occur when a concept "moves" from one level to another, say level 2, to level 3 or level 2 to level 3. Level changes can occur as part of an aggregation, or dis-aggregation or as part of regrouping concepts. Unfortunately, the research found few "level changes." For example, the APQC and the PWC taxonomies had three level changes, two from level 2 to 3 and one from level 3 to 2. As an example, one had identical items, but at different levels, e.g., "4.2 Sell Products and Services" and "4.1.7 Sell Products and Services."

As a result, level changes were not included in this analysis, for a number of reasons. First, this paper is primarily concerned with semantic evolution of the taxonomies. Level changes are not semantic changes, per se. With a level change, the same descriptor is still in both versions of the taxonomy. Second, very few level changes were found in these particular data sets so that there was no real basis of comparison in these data sets. Three changes across hundreds of items limit the comparability and the usefulness for these particular data sets. Third, meaningful representation of the change can even further dilute the extent of change. With level changes, investigation can focus on the extent of level changes forced down and the extent forced up. We probably also should focus on which levels items were forced up and which levels items were forced down. For example, assuming three levels in a taxonomy, change can go from level 1 to level 2 or level 3, or level 3 to level 2 or level 1, etc., with each of these potentially requiring representation. In the case of APQC and PWC, there were two changes from level 2 to level 3 and one from level 3 to level 2. Unfortunately, such categorizations takes a small amount of data (3 changes) and makes it even smaller (0 from level 1 to 2, 0 from level 1 to 3, 2 from level 2 to 3, etc.).

However, this is not to say that level changes are unimportant in the future for this data set or for other data sets. Perhaps as these particular taxonomies continue to evolve over time it could be important to watch the role that "level changes" play. For example, the level changes may all move in a single direction, such as from level 2 to level 3, indicating concepts are being pushed down to lower levels. Further, level changes may be more important in other taxonomies where such evolutionary changes could occur more frequently, making them more influential and more important to monitor beyond the few in these data sets.

5. Analysis

This section analyzes various aspects of the original and revised frameworks, which allows us to begin to understand some of the factors involved in the evolution of a taxonomy, from one version to another.

5.1. Matches by taxonomy level

The data were analyzed for number of identical, very similar and similar "matches" and "changes." The results are summarized by the different levels of taxonomy concepts in Table 1.

Table 1
Quantitative analysis by taxonomy level

	Top level concept	2nd-level concept	3rd-level concept
Panel A: APQC vs. PWC			
Identical	10	40	71
Very similar	2	3	9
Similar	1	7	16
No match (PWC only)	0	10	46
No match (APQC only)	0	21	87
More than one New	0	0	3
More than one Old	0	0	27
Percent identical	77%	67%	50%
Percent match	100%	83%	68%

	Top level concept	2nd-level concept	3rd-level concept	4th-level concept	5th-level concept
Panel B: APQC vs. XYZ					
Identical	9	20	77	0	0
Very similar	0	3	12	0	0
Similar	3	22	17	5	1
No match (XYZ only)	0	22	112	205	12
No match (APQC only)	1	28	79	0	0
More than one New	0	0	3	0	0
More than one Old	0	2	6	2	0
Percent identical	69.23%	20.43%	22.95%	0.00%	0.00%
Percent match	92.31%	44.09%	31.51%	0.00%	0.00%
Panel C: PWC vs. XYZ					
Identical	9	29	84	0	0
Very similar	0	3	0	0	0
Similar	3	14	31	0	0
No match (XYZ only)	0	16	108	210	13
No match (PWC only)	1	12	31	0	0
More than one New	0	0	0	0	0
More than one Old	0	0	0	0	0
Percent identical	69.23%	39.19%	33.07%	0.00%	0.00%
Percent match	92.31%	62.16%	45.28%	0.00%	0.00%

Notes:
"Level" refers to the level in the taxonomy.
"Identical" means wording between corresponding items in two taxonomies was identical.
"Very similar" means wording between corresponding items in two taxonomies was very similar.
"Similar" means wording between corresponding items in two taxonomies was similar.

5.2. APQC vs. PWC

There was little change at the top level (Level 1) concepts, with 100% matches and 77% were identical. However, as the concept level decreased from top level to second level to the third level, the percent of identical matches went from 77% to 67% to 50%. In addition, as the concept level went from top level to second level to third level, the percent of matches went from 100% to 83% to 68%.

5.3. APQC vs. XYZ

As with APQC and PWC, the top level had the least amount of change, with 12 out of 13 matches and 69% identical. One top-level concept was dropped by XYZ from the APQC taxonomy to reflect XYZ's manufacturing business. The percent of matches decreased as the taxonomy became more detailed, going from 92% to 44% to 31%. There were 3 fourth level concepts in the APQC version, but 225 fourth and fifth level concepts in the XYZ version.

5.4. PWC vs. XYZ

As with the other two comparisons, the least amount of change was at the top level. One of the top level concepts was in PWC but not XYZ. There were more identical matches with PWC and XYZ than with PWC and APQC. However, there were fewer matches overall because of the lower number of items in the PWC taxonomy.

5.5. Discussion

For each the taxonomies, there is greater similarity at the higher levels, and less similarity at lower levels. Accordingly, the results suggest that the more detailed level of the concept, the fewer the percentage of matches and the greater likelihood of a change, as compared to higher-level concepts.

5.6. Category matches

The matches by category are summarized in Table 2. Matches provide a measure of knowledge "stability" from version to version: The greater the change in a taxonomy, the fewer the matches.

Knowledge stability appears to vary by category of knowledge within the taxonomy. In general, there were more matches and more average matches per category in the operating processes, than in the "management support" processes.

Anticipating taxonomy evolution requires anticipating what knowledge will stay the same and match up over time. As seen here, knowledge stability varies from category to category. However, there was substantial stability. For example, XYZ employed almost 80% of the categories from PWC and approximately 62% of the categories from APQC. The greatest number of matches was between the largest and the smallest taxonomies, PWC and XYZ. However, for each of the pairs of taxonomies, the number of matches by categories was highly correlated with each other. Further, the number of matches varied only slightly, ranging from 159 to 173.

There were two publicly available taxonomies, APQC and PWC to the developers of XYZ. It appears that in construction of the XYZ taxonomy both the APQC and the PWC taxonomies were used: Some portions were more similar to APQC and other portions were more similar to PWC. For example, XYZ had more matches with APQC in categories 3, 4, 7, 8, and 13. However, XYZ had more matches with PWC in categories 2, 5, 10, and 11. The greatest difference was with category 11, where XYZ seems to have largely adopted the PWC taxonomy. This is not surprising since ideally taxonomy builders would use all of the information available to them. However, this also suggests that development of taxonomies becomes a more complex process over time since previous versions of taxonomies can be examined to determine what has been changed out or to determine what "fits" best with the developers' needs.

Table 2

Panel A: total number of items by category

Category	APQC	PWC	XYZ
1	16	12	12
2	16	19	25
3	24	22	37
4	13	15	75
5	24	22	49
6	14	13	0
7	11	10	14
8	45	21	48
9	35	12	137
10	30	31	50
11	11	15	27
12	8	9	32
13	24	16	14
Total	271	217	520

Panel B: matches by category for each pair of taxonomies

Category	APQC vs. PWC	APQC vs. XYZ	PWC vs. XYZ
1	12	12	12
2	13	15	17
3	17	21	20
4	7	12	8
5	20	19	22
6	12	0	0
7	10	11	10
8	16	26	16
9	3	6	7
10	24	24	27
11	1	1	14
12	8	8	7
13	16	14	13
Total	159	169	173

Panel C: correlations

	APQC vs. XYZ	XYZ vs. PWC
APQC vs. PWC	0.790***	0.685***
APQC vs. XYZ		0.785***

Notes:
M vs. *N* — number of matches in between the sets *M* and *N*.
*** Significant at the .01 level or better.

5.7. Category changes

The category changes are presented in Table 3. The changes are not spread uniformly across the 13 categories. Instead, change in categories occurs unevenly, with most of the change centered in a few categories of knowledge. In general, the categories with the greatest change were 4, 8, 9 and 11 were among the categories with the greatest change for each of the three

Please cite this article as: Daniel E. O'Leary, Empirical analysis of the evolution of a taxonomy for best practices, Decision Support Systems (2006), doi:10.1016/j.dss.2006.09.004

Table 3
Percentage of change by category

Category	Total new	Total old	Total changes	Percentage of change (%)
Panel A: APQC vs. PWC				
1	0	4	4	2.37
2	6	3	9	5.33
3	5	7	12	7.10
4	8	6	14	8.28
5	2	5	7	4.14
6	1	2	3	1.78
7	0	1	1	0.59
8	5	29	34	20.12
9	9	32	41	24.26
10	5	6	11	6.51
11	14	10	24	14.20
12	1	0	1	0.59
13	0	8	8	4.73
	56	113	169	100.00
Panel B: APQC vs. XYZ				
1	0	4	4	0.88
2	10	1	11	2.43
3	16	3	19	4.19
4	63	1	64	14.13
5	30	5	35	7.73
6	0	14	14	3.09
7	3	0	3	0.66
8	22	20	42	9.27
9	131	29	160	35.32
10	26	5	31	6.84
11	26	10	36	7.95
12	24	0	24	5.30
13	0	10	10	2.21
	351	102	453	100.00
Panel C: PWC vs. XYZ				
1	0	0	0	0.00
2	8	2	10	2.58
3	17	2	19	4.91
4	67	7	74	19.12
5	27	0	27	6.98
6	0	13	13	3.36
7	4	0	4	1.03
8	32	5	37	9.56
9	130	5	135	34.88
10	23	4	27	6.98
11	13	1	14	3.62
12	25	2	27	6.98
13	1	3	4	1.03
	343	44	387	100.00

Notes:
"Total new" refers to the total number of new categories added to the taxonomy.
"Total old" refers to the total number of the old categories not used in the taxonomy.
"Total changes" refers to the total number of changes from one version to another.
Percentage of change — portion of the change, out of 100%, due to that category.

Table 4
Correlation coefficients

	Total matches	Total new	Total old	Total changes
Panel A: APQC vs. PWC				
Total matches		−0.495 *	− 0.200	−0.328
Total new			0.445	0.695 ***
Total old				0.953 ***
Total changes				
Panel B: APQC vs. XYZ				
Total matches		−0.179	−0.528 *	−0.226
Total new			0.200	0.999 ***
Total old				0.293
Total changes				
Panel C: PWC vs. XYZ				
Total matches		−.149	−0.152	−0.161
Total new			0.573 **	0.985 ***
Total old				0.706 ***
Total changes				

Notes:
"Total matches" refers to the number of matches in each of the 13 categories.
"Total new" refers to the number of new categories in each of the 13 categories.
"Total old" refers to the number of old categories at levels in the original version in the 13 categories.
"Total changes" is the sum of the "Total old" and "Total new."
* Significant at .1 level.
** Significant at .05 level.
*** Significant at .01 level.

pairs of taxonomies. Accordingly, three of the categories with the greatest change were from the "management support" grouping.

As seen in Table 4, matches were generally negatively correlated with the changes in the taxonomies. The correlation between the matches and new items added in the comparison between APQC and PWC, and the number of matches and the number of old items between APQC and PWCC were statistically significant. Further, the number of new items and the number of old items, by category between PWC and XYZ also was significant. As a result, proactive evolution of a taxonomy could leverage relationships between matches and changes.

5.8. Increased complexity

There is evidence that the evolved taxonomies have greater "complexity." One measure of complexity is "entropy." Shannon [10] proposed the use of entropy using the following formula

$$H(S) = -E(\log S),$$

Table 5
Entropy analysis

Panel A: APQC framework — number of categories by level

Category	No. level 2	No. level 3	No. level 4	Total	Entropy (13)	Entropy (71)	Entropy (184)	Entropy (271)
1	3	9	3	16	0.1973	0.1337	0.1476	0.1671
2	4	11	0	16	0.1973	0.1620	0.1684	0.1671
3	6	17	0	24	0.1973	0.2088	0.2201	0.2147
4	2	10	0	13	0.1973	0.1006	0.1583	0.1457
5	4	19	0	24	0.1973	0.1620	0.2345	0.2147
6	4	9	0	14	0.1973	0.1620	0.1476	0.1531
7	3	7	0	11	0.1973	0.1337	0.1244	0.1301
8	9	35	0	45	0.1973	0.2618	0.3157	0.2981
9	8	26	0	35	0.1973	0.2460	0.2765	0.2643
10	6	23	0	30	0.1973	0.2088	0.2599	0.2436
11	10	0	0	11	0.1973	0.2761		0.1301
12	7	0	0	8	0.1973	0.2284		0.1040
13	5	18	0	24	0.1973	0.1868	0.2274	0.2147
	71	184	3	271	2.5649	2.4709	2.2803	2.4472

Panel B: PWC framework — number of categories by level

Category	No. level 2	No. level 3	Total	Entropy (13)	Entropy (60)	Entropy (144)	Entropy (217)
1	3	8	12	0.1973	0.1498	0.1606	0.1601
2	4	14	19	0.1973	0.1805	0.2266	0.2132
3	7	14	22	0.1973	0.2507	0.2266	0.2320
4	3	11	15	0.1973	0.1498	0.1965	0.1847
5	4	17	22	0.1973	0.1805	0.2522	0.2320
6	4	8	13	0.1973	0.1805	0.1606	0.1686
7	3	6	10	0.1973	0.1498	0.1324	0.1418
8	7	13	21	0.1973	0.2507	0.2171	0.2260
9	2	9	12	0.1973	0.1134	0.1733	0.1601
10	9	21	31	0.1973	0.2846	0.2808	0.2780
11	3	11	15	0.1973	0.1498	0.1965	0.1847
12	7	1	9	0.1973	0.2507	0.0345	0.1320
13	4	11	16	0.1973	0.1805	0.1965	0.1922
	60	144	217	2.5649	2.4712	2.4541	2.5056

Panel C — APQC without category 6

Category	Level 1	Level 2	Level 3	Level 4	Total	Entropy 12	Entropy 67	Entropy 178	Entropy 259
1	1	3	9	2	15	0.2071	0.1391	0.1509	0.1650
2	1	4	11	0	16	0.2071	0.1683	0.1720	0.1720
3	1	6	17	0	24	0.2071	0.2161	0.2243	0.2204
4	1	2	10	0	13	0.2071	0.1048	0.1618	0.1502
5	1	4	22	0	27	0.2071	0.1683	0.2584	0.2357
7	1	3	7	0	11	0.2071	0.1391	0.1273	0.1342
8	1	9	35	0	45	0.2071	0.2697	0.3198	0.3041
9	1	8	26	0	35	0.2071	0.2538	0.2810	0.2705
10	1	6	23	0	30	0.2071	0.2161	0.2644	0.2497
11	1	10	0	0	11	0.2071	0.2839		0.1342
12	1	7	0	0	8	0.2071	0.2360		0.1074
13	1	5	18	0	24	0.2071	0.1937	0.2317	0.2204
	12	67	178	2	259	2.0708	1.9590	1.9599	2.0359

Panel D: XYZ framework — number of categories by level

Category	Level 1	Level 2	Level 3	Level 4	Level 5	Total	Entropy (12)	Entropy (67)	Entropy (218)	Entropy (520)
1	1	3	8	0	0	12	0.2071	0.1391	0.1213	0.0870
2	1	4	15	5	0	25	0.2071	0.1683	0.1842	0.1459

(continued on next page)

Table 5 (*continued*)

Panel D: XYZ framework — number of categories by level

Category	Level 1	Level 2	Level 3	Level 4	Level 5	Total	Entropy (12)	Entropy (67)	Entropy (218)	Entropy (520)
3	1	7	18	11	0	37	0.2071	0.2360	0.2059	0.1881
4	1	5	15	41	13	75	0.2071	0.1937	0.1842	0.2793
5	1	8	27	13	0	49	0.2071	0.2538	0.2587	0.2226
7	1	3	6	4	0	14	0.2071	0.1391	0.0989	0.0973
8	1	6	24	17	0	48	0.2071	0.2161	0.2429	0.2199
9	1	10	40	86	0	137	0.2071	0.2839	0.3111	0.3514
10	1	7	23	19	0	50	0.2071	0.2360	0.2373	0.2252
11	1	3	15	8	0	27	0.2071	0.1391	0.1842	0.1536
12	1	8	17	6	0	32	0.2071	0.2538	0.1990	0.1716
13	1	3	10	0	0	14	0.2071	0.1391	0.1414	0.0973
	12	67	218	216	13	520	2.0708	2.0049	2.0286	2.2391

Notes:

"No." stands for "Number." For example, "No. level 2" means the number at the second level in the taxonomy.

"Entropy (*j*)" means the amount of entropy in *j* items in the taxonomy. Different measures are provided for different levels in the taxonomy.

where E is the expectation operator. That same formula was then proposed for use in a number of applications (Lev [5] and Theil [11]).

Entropy is higher ("more complex") when probability is equally distributed and lower when it is not equally distributed. Distributing the probability equally across n categories maximizes entropy for n categories. Any shifting of probability within those n categories will drive the entropy down. For example, assuming that the probability is equally distributed among the thirteen top-level categories, the original and revised frameworks each have entropy of 2.5649, the maximum entropy measure for 13 categories.

Entropy can be used to analyze the data across different levels of the taxonomy, for example levels 1, 2 and 3 (Table 5) for "complexity." Alternatively, entropy can be used to analyze particular categories, such as "understand markets and customers" (Table 6) across all levels.

5.9. APQC

Assume that probability is distributed based on number of items, for example, number of items in the taxonomy at level 2, level 3 and in total. For the original framework, entropy for level 2 items (2.4709) and total items (2.4745) is very similar. However, for the original framework the level of entropy at level 3 (2.2791) is lower, indicating less complexity at level 3.

5.10. APQC vs. PWC

In the PWC framework, when compared to APQC's version, entropy is higher at each of level 2 (2.4712), level 3 (2.4541) and for the total set of categories (2.5056), respectively, all suggesting a greater complexity within each of the 13 categories.

Further, comparing the entropy measures for each respective category of Tables 6 finds that the original only has greater entropy than the PWC framework, in category 2, suggesting that the revised version is also more complex by category, not just level.

5.11. APQC vs. XYZ

Comparing the entropy measures for XYZ to APQC indicates that across all categories (Table 5), XYZ has greater entropy. Similarly, by category (Table 6), for 9 of the 12 categories, XYZ has greater entropy than APQC.

5.12. Discussion

Despite the fact that PWC shrank the number of items to 217 and XYZ increased the number of items to 520, both of the new taxonomies generated increased entropy, suggesting an increase in complexity of the number of categories in each of the respective taxonomies. Future research can investigate the extent to which other taxonomies have a tendency to become more complex as they evolve.

5.13. Populated categories

The three frameworks were "populated" with different information in different categories of the taxonomies in order to provide best practices information. Depending on the taxonomy, populated information could include measures, diagrams or case studies, designed to provide additional information about the best practice. As seen in Table 7, in the APQC framework there were 663 measures provided across the 13 top level categories.

Table 6
Entropy by category

1st level	2nd level	3rd level	Total	Entropy
Panel A: APQC version by category				
1	3	9	13	0.7903
1	4	11	16	0.7775*&
1	6	17	24	0.7233
1	2	10	13	0.6871&
1	4	19	24	0.6160
1	4	9	14	0.8305
1	3	7	11	0.8600
1	9	35	45	0.6019
1	8	26	35	0.6597&
1	6	23	30	0.6390
1	10		11	0.3046
1	7		8	0.3768
1	5	18	24	0.6750
Panel B: PWC version by category				
1	3	8	12	0.8240*
1	4	14	19	0.7080
1	7	14	22	0.7925*
1	3	11	15	0.7299*
1	4	17	22	0.6497*
1	4	8	13	0.8587*
1	3	6	10	0.8979*
1	7	13	21	0.8081*
1	2	9	12	0.7215*
1	9	21	31	0.7337*
1	3	11	15	0.7299*
1	7	1	9	0.6837*
1	4	11	16	0.7775*

Panel C: XYZ version by category

Category	Level 1	Level 2	Level 3	Level 4	Total	Levels 1–3	Levels 1–4
1	1	3	8	0	12	0.8240&	N/A
2	1	4	15	5	25	0.7285	1.0504
3	1	7	18	11	37	0.7631&	1.1238
4	1	5	15	41	62	0.6129	0.8864
5	1	8	27	13	49	0.7037&	1.0557
7	1	3	6	4	14	0.8817&	1.2397
8	1	6	24	17	48	0.6872&	1.0548
9	1	10	40	86	137	0.5864	0.8787
10	1	7	23	19	50	0.7107&	1.0784
11	1	3	15	8	27	0.6928&	1.0532
12	1	8	17	6	32	0.7909&	1.1048
13	1	3	10	6	20	0.7809&	1.1421

Notes:
* Larger than corresponding Table 6 Panel B item & Larger than corresponding Table 6 Panel C item.
* Larger than corresponding Table 6 Panel A item.
& Larger than corresponding Table 6 Panel A item (APQC had negligible items in level 4).
Level *j* — level in the taxonomy.

The PWC version has the entire taxonomy on its web site of best practices. However, it only has a portion of the taxonomy actually "populated" with measures and best practices. In particular, there are 142 categories populated with best practices, as seen in Table 7. Unfortunately, access was not available to the number of measures available in the PWC version. The XYZ taxonomy was the least populated, with fewer measures than APQC (405) and fewer populated categories than PWC (30).

Table 7
Number of measures in original vs. number of populated categories in evolved versions

Category	APQC number of measures	PWC sections populated	XYZ number of measures	XYZ sections populated
1	16	7	16	1
2	4	13	4	1
3	61	13	4	1
4	39	13	38	3
5	150	14	120	6
6	0	4	0	0
7	28	7	65	11
8	91	19	0	0
9	54	11	0	0
10	154	25	158	7
11	61	4	0	0
12	0	7	0	0
13	5	5	0	0
	663	142	405	30

	PWC populated	XYZ measures	XYZ populated
APQC measures	0.750***	0.768***	0.426
PWC populated		0.609**	0.306
XYZ measures			0.791***
XYZ populated			

Notes:
APQC measures — number of measures in each category of the APQC taxonomy.
PWC sections populated — number of sections in each category that are populated in the PWC taxonomy.
XYZ measures — number of measures in each section of the XYZ taxonomy.
XYZ sections populated — number of sections in each category that are populated with measures.

5.14. APQC vs. PWC

The results indicate that the APQC number of measures by top-level category, is highly correlated with the number of sections summarized by top level category, that are populated in the PWC version (.750), that is significant at the .01 level. Accordingly, the numbers of measures in the different categories in the APQC version are associated with the number of populated categories in the PWC version.

5.15. APQC vs. XYZ

XYZ has populated their taxonomy with limited amounts of data. However, the number of APQC measures in the top-level categories of the original version is statistically significantly correlated with the number of measures in the XYZ version (.768) that is significant at the .01 level.

5.16. PWC and XYZ

The numbers of sections of the PWC taxonomy is correlated with the number of measures in the XYZ taxonomy, and is statistically significant at the .05 level or better.

5.17. Discussion and comparison

The existence of measures in categories works as if it has some "momentum" to keep the measures in the taxonomy. Categories with measures continue on in different versions. Further, measures developed for one version appear to carry over into other versions. Populated category information damps out slowly.

Table 8
Non-populated: new or matching categories

Panel A: PWC

Category	Non-populated	Non-populated that match	Non-populated that are new only
1	5	5	0
2	6	6	0
3	9	6	3
4	2	0	2
5	8	7	1
6	9	9	0
7	3	3	0
8	2	2	0
9	1	1	0
10	6	5	1
11	11	1	10
12	2	2	0
13	11	11	0
	75	58	17

Panel B: correlation coefficients

	Match	New only
Non-populated	0.679***	0.475
Match		−0.322
New only		

Notes:
Non-populated — number of items included in each category that were non-populated.
Match — number of items in each category that match in the comparison between APQC and PWC.
New only — non-populated categories that are in the new version but not the old.

Please cite this article as: Daniel E. O'Leary, Empirical analysis of the evolution of a taxonomy for best practices, Decision Support Systems (2006), doi:10.1016/j.dss.2006.09.004

5.18. Non-populated categories

Although many of the categories are populated with measures, some categories are not. The available information limits analysis, since there was no access available to those APQC categories that were not populated. Further, a large number of XYZ categories are not populated. As a result, this discussion is really focused primarily on PWC's version as it relates to the APQC version.

The non-populated categories in the PWC taxonomy were broken into two categories, those with matches and those with no match. The results are summarized in Table 8.

Out of 75 non-populated subcategories, only 17 were new categories that were added, and 10 were in a single category, number 11. Excluding category 11, using a test "difference in proportions" (Dixon and Massey [4]), the proportion of the items that were matched in both the old and new framework, compared to the rest of the sample, is significantly different at the .0025 level. This suggests that change in categories is made for purposes of population, i.e., change is made for a "reason." As a result, outside of category 11, only in 7 subcategories were new categories added, but not populated. At PWC, the addition of a category to the revised framework has meant that it would be populated.

Non-populated categories are correlated with the number of matches. One explanation for this is that non-populated categories may go "under the radar" of taxonomy change. That is, there is not as much concern for them, so they are not populated.

6. General discussion

A number of changes in the process analysis framework, going from the original APQC version to two different evolved versions have been documented. The two evolved versions (PWC and XYZ) were substantially different in their size, with the PWC version having 217 items and the XYZ version having 520 items. However, observation and statistical analysis yielded a number of findings.

- Concepts at more detailed levels were found more likely to be changed. Concepts at higher levels were found more likely to stay the same.
- Knowledge change is not uniform. A number of categories stayed the same in evolved versions, while others changed substantially. Some categories of content were subject to greater change, such as information technology.
- The greatest amount of change occurred to processes in the "Management and Support Processes."
- The number of new items was linearly related to both the total number of matches and the number of items removed. When knowledge is removed other knowledge takes its place. New knowledge also is related to existing knowledge.
- The structure of the taxonomy became more complex both taxonomies. Both the PWC and XYZ versions had greater entropy compared to APQC.
- There was a relationship between the number of measures used to populate the APQC version and the number of categories and subcategories that were populated in the PWC version. In addition, the number of measures in the APQC version was correlated with the number of measures in the XYZ version. These relationships suggest a kind of "momentum" between different versions, with populated portions providing "stability."
- Finally,there appears to be relationship between categories that are non-populated and matches between those categories.

7. Contributions and extensions

The evolution of a taxonomy for best practices was analyzed in order to begin to understand the evolution of knowledge bases. The APQC process classification framework evolved from its original published version to two organization's versions (PWC and XYZ). An empirical approach was used to analyze taxonomy changes from the original APQC version to the PWC version and to the XYZ version.

7.1. Contributions

This research is one of the first studies to empirically analyze an existing knowledge base from the perspective of how it has evolved. Accordingly, it provides a basis to set expectations as to how taxonomies are likely to change from version to version. Further, this article provides the beginnings of a "methodology" to study those evolutions. Finally, the research generated here could form the basis of a system designed to anticipate and facilitate taxonomy evolution.

7.2. Extensions

This research can be extended to include other versions 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, other areas of artificial intelligence and knowledge management could be used to extend these results. For example, fuzzy sets might be used to analyze the concepts of "identical," "very similar," and "similar."

There are a number of alternatives to the entropy measure. Future research could employ the Gini coefficient, lift measures (including L-Qual), and other measures based on Receiver Operating Characteristic analysis, such as AUC (*a*rea *u*nder the *c*urve). See Provost and Fawcett [7,8] and Piatetsky-Shapiro [6]. (I would like to thank one of the anonymous referees for bringing these to my attention.)

Acknowledgement

I would like to thank the anonymous referees from the conference and the special issue for their many comments on earlier versions of this paper. The research was supported in part by a grant from the MKIDS program.

References

[1] APQC's International Benchmarking Clearinghouse and Arthur Andersen and Co., Process Classification Framework, http://old.apqc.org/free/framework.cfm1996.
[2] F. Caldwell, Taxonomy for Business Best Practices Thrives, Gartner Report, , September 10, 2003.
[3] B. Chandrasekaran, Towards a taxonomy of problem solving types, AI Magazine 4 (1) (1983) 9–17.
[4] W. Dixon, F. Massey, Introduction to Statistical Analysis, McGraw – Hill, New York, 1969.
[5] B. Lev, Accounting and Information Theory, American Accounting Association, Illinois, 1969.
[6] G. Piatetsky-Shapiro, Discovery, Analysis and Presentation of Strong Rules, in Knowledge Discovery in Databases, American Association for Artificial Intelligence, Menlo Park, 1991, pp. 228–248.
[7] F. Provost, T. Fawcett, Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions, Proceedings of the Third International Conference in Knowledge Discovery and Data Mining (KDD-97), Huntington Beach, CA, American Association for Artificial Intelligence, Menlo Park, 1997, pp. 43–48.
[8] F. Provost, T. Fawcett, Robust classification systems for imprecise environments, Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), Madison, Wisconsin, American Association for Artificial Intelligence, Menlo Park, 1998, pp. 706–713.
[9] PWC (PriceWaterhouseCoopers), Process Classification Framework, http://www.globalbestpractices.com/Best_practices/2003.
[10] C.E. Shannon, A mathematical theory of communication, Bell System Technical Journal XXVII (1948) 379–423.
[11] H. Thiel, Economics and Information Theory, Rand McNally and North Holland Publishing Co., Chicago and Amsterdam, 1967.

Daniel O'Leary is a Professor in the Marshall School of Business at the University Of Southern California (USC), focusing on information systems, including enterprise resource planning systems and knowledge management systems. Dan is a full professor and has been at USC since 1985. Dan received his PhD from Case Western Reserve University. He received his MBA from the University of Michigan and his bachelor's degree from Bowling Green University (Ohio).

He is the former editor of *IEEE Intelligent Systems* and John Wiley's *International Journal of Intelligent Systems in Accounting, Finance and Management*. His book, *Enterprise Resource Planning Systems*, published by Cambridge University Press, has been translated into both Chinese and Russian.

Please cite this article as: Daniel E. O'Leary, Empirical analysis of the evolution of a taxonomy for best practices, Decision Support Systems (2006), doi:10.1016/j.dss.2006.09.004