2015

Taxonomy Development for Complex Emerging Technologies - The Case of Business Intelligence and Analytics on the Cloud

Odette Sangupamba Mwilu
CEDRIC-CNAM, odemwilu@hotmail.com

Nicolas Prat
ESSEC Business School, prat@essec.edu

Isabelle Comyn-Wattiau
CEDRIC-CNAM, isabelle.wattiau@cnam.fr

TAXONOMY DEVELOPMENT FOR COMPLEX EMERGING TECHNOLOGIES – THE CASE OF BUSINESS INTELLIGENCE AND ANALYTICS ON THE CLOUD

Odette Sangupamba Mwilu, CEDRIC-CNAM, Paris, France, odemwilu@hotmail.com

Nicolas Prat, ESSEC Business School, Cergy-Pontoise, France, prat@essec.edu

Isabelle Comyn-Wattiau CEDRIC-CNAM, Paris and ESSEC Business School, Cergy-Pontoise, France, isabelle.wattiau@cnam.fr; wattiau@essec.edu

Abstract

Taxonomies are essential in science. By classifying objects or phenomena, they facilitate understanding and decision making. In this paper, we focus on the development of taxonomies for complex emerging technologies. This development raises specific challenges. More specifically, complex emerging technologies are often at the intersection of several areas, and the conceptual body of knowledge about them is often just emerging, hence the key role of empirical sources of information in taxonomy building. One particular issue is deciding when enough sources have been examined. In this paper, we use Nickerson et al's methodology for taxonomy development. Based on the identified limitations of this method, we extend it for the development of taxonomies for complex emerging technologies. We identify three types of information sources for taxonomies, and present a set of guidelines for selecting the sources, drawing on systematic literature review. The taxonomy development process iteratively examines sources, performing operations on taxonomies (e.g. addition of a dimension, splitting of a dimension…) as required to take new information into account. We characterize operations on taxonomies. We use this characterization, along with the typology of sources, to help decide when the process of source examination may be stopped. We illustrate our extension of Nickerson et al's method to the development of a taxonomy for business intelligence and analytics on the cloud.

Keywords: business intelligence, analytics, cloud computing, taxonomy, systematic literature review, emerging technology.

1 INTRODUCTION

Taxonomies classify objects or phenomena of interest, according to the dimensions that are relevant for characterizing and discriminating between these objects. Through this classification, they facilitate the understanding of complex, multi-faceted objects. This understanding in turn facilitates decision making and, ultimately, action.

Taxonomies are extensively used in information systems. Recent examples include a taxonomy of smart objects (Lopez et al. 2011), a taxonomy of IT solution risks to drive risk management (Herzfeldt et al. 2012), a methodology for developing and validating a cybercrime taxonomy (Land et al. 2013), a taxonomy of political processes in IS development (Sabherwal & Grover 2009), and a taxonomy of information (McKinney et al. 2010).

We focus on the development of taxonomies for complex emerging technologies. An example of this type of technology is business intelligence (BI) and analytics on the cloud (a.k.a. cloud-based BI). BI and analytics on the cloud is a disruptive technology, and major software vendors are positioning themselves on this market ("http://www.forbes.com/sites/groupthink/2014/12/19). Natural-language question answering systems (NLQA) are another example of complex emerging technology. Two characteristics of this type of technologies make the process of taxonomy development challenging. First, they are typically at the intersection of several disciplines (or areas of the same discipline). For example, business intelligence and analytics on the cloud is at the intersection between business intelligence/analytics and cloud computing, and natural-language question answering systems are based on artificial intelligence as well as linguistics. Second, for these technologies, the conceptual body of knowledge is often still limited, hence the key role of empirical information sources in the taxonomy building process. Thus, building taxonomies for emerging technologies is both a complex and a relevant challenge. They are generally the first step before constructing ontologies. The latter are then operationalized as main components of knowledge bases.

In this paper, we use the taxonomy development method of Nickerson et al. (Nickerson et al. 2013). We identify some limitations of this method, requiring its adaptation to the development of taxonomies of complex emerging technologies. We adapt the method and illustrate one application to the case of taxonomy development for BI and analytics on the cloud. The purpose of this taxonomy development effort is to guide the decision of moving BI and analytics to the cloud. The study of the literature (described later in the paper) reveals that there is currently no complete and systematic taxonomy dedicated to BI and analytics on the cloud.

The essential contributions of this paper are: (1) Guidance on the source selection process for building taxonomies of complex emerging technologies. This guidance draws on systematic literature review and is based on the distinction between three types of sources (academic papers on the emerging technology, academic papers specific to each discipline or area implied in the technology, professional articles and case studies on the emerging technology). (2) A characterization of the operations on taxonomies (the examination of a new source may trigger operations on the considered taxonomy, depending on the information in that source). (3) Some preliminary guidance to help decide when the process of source examination may be stopped based on the typology of sources and the operations performed. (4) Application to an initial effort of taxonomy development for BI on the cloud.

The paper is structured as follows. Section 2 summarizes Nickerson et al's method for taxonomy development, and identifies limitations of this method for the development of taxonomies of complex emerging technologies. Section 3 adapts the method. Section 4 applies the adapted method to taxonomy development for BI and analytics on the cloud. Section 5 discusses the issue of deciding when to stop examining information sources (convergence of the taxonomy development process), and concludes the paper.

2 NICKERSON ET AL'S METHOD FOR TAXONOMY DEVELOPEMENT

Even though taxonomies are frequently used in research, the process for building them is often empirical and ad-hoc. A recent paper (Nickerson et al. 2013) proposes a method for taxonomy development, using a design science approach. This detailed method systematically guides the process of taxonomy building and evaluation. The efficacy of the method is illustrated by developing a taxonomy of mobile applications.

The method of Nickerson et al. capitalizes on previous approaches to taxonomy development and is applicable to many domains. Consequently, this is the method we use. However, the method has some limitations when applied to taxonomy development for complex emerging technologies. This section summarizes the method and its limitations. In the next section, based on the identified limitations, we will adapt the method.

2.1 The method

According to Nickerson et al. (p.340) a taxonomy is a set of dimensions with mutually exclusive and collectively exhaustive characteristics such that each object under consideration has one and only one characteristic for each dimension. Taxonomy development is iterative, combining conceptual-to-empirical and empirical-to-conceptual approaches. The conceptual-to-empirical approach is deductive, starting with theoretical identification of dimensions and characteristics, and continuing with application of the taxonomy to the classification of concrete objects, possibly leading to revisions of the taxonomy. The empirical-to-conceptual-approach is inductive, starting from concrete objects to identify dimensions and characteristics (e.g. similarity-based clustering of objects).

To provide a basis for the identification of dimensions and characteristics, taxonomy development requires the choice of a meta-characteristic. The meta-characteristic *"is the most comprehensive characteristic that will serve as the basis for the choice of characteristics in the taxonomy"* (Nickerson et al. 2013). Each characteristic should be a logical consequence of the meta-characteristic. Moreover, the methodology is iterative. Ending conditions are provided to assist the taxonomy builder in deciding when the process may stop. The first and main ending condition is that the taxonomy must meet its definition. In addition, the authors provide *objective* and *subjective* ending conditions. Subjective conditions state that the taxonomy should be concise, robust, comprehensive, extendible and explanatory. They are of particular importance at the very end of the taxonomy development process. In this paper, we focus on objective conditions, which are typically tested at each iteration of the taxonomy development process. These conditions are as follows (Nickerson et al. 2013) (p.344):

1. All objects or a representative sample of objects have been examined.
2. No object was merged with a similar object or split into multiple objects in the last iteration.
3. At least one object is classified under every characteristics of every dimension.
4. No new dimensions or characteristics were added in the last iteration.
5. No dimensions or characteristics were merged or split in the last iteration.
6. Every dimension is unique and not repeated (no dimension duplication).
7. Every characteristic is unique within its dimension (no characteristic duplication within a dimension).
8. Each cell (combination of characteristics) is unique and is not repeated (no cell duplication).

2.2 Need to adapt the method

Although the eight objective ending conditions above provide useful guidance in deciding when taxonomy development may stop, these conditions are often difficult to apply in practice, more specifically in the case of taxonomy development for complex emerging technologies. More specifically;

- In condition (1), what is the basis for deciding that « *all objects or a representative sample of objects have been examined*»? The authors don't specify what they mean by « representative sample of objects ». We believe that application of the principles of systematic literature review may help here.

- If no change has been made in the taxonomy during one iteration (e.g. conditions (2), (4) and (5)), it doesn't necessarily mean that a new iteration would also have found no change to make. In other words, it is important to determine when the taxonomy development process starts to converge (the number of operations performed on the taxonomy at each step stabilizes and approaches zero).
- If some characteristics are such that no object is classified under them (i.e. if condition (3) is not met), does it mean that we must find an object to classify under them or delete these characteristics? Probably not in all cases, especially in the case of emerging technologies, where the number of objects to examine is, by definition, restricted.

In the next section, we adapt the method of Nickerson et al. to the development of taxonomies of complex emerging technologies, based on the previously identified limitations. First, we formally define taxonomies, extending the definition mentioned above. We then define all the possible operations on taxonomies, starting from the initial operations identified by Nickerson et al. Finally, we propose a typology of information sources, distinguishing three types of sources. Drawing on the recommendations of systematic literature review, we provide guidelines for selecting sources for each type of source. The typology of sources and the operations will serve as a basis for deciding when the taxonomy development process may be considered as converging, by studying the evolution of the number of operations performed for the different types of operations and the different types of sources.

3 ADAPTING THE METHOD TO THE DEVELOPMENT OF TAXONOMIES OF COMPLEX EMERGING TECHNOLOGIES

3.1 Formal definition of taxonomies

First, we should define taxonomies (Prat et al. 2014). A taxonomy is a set of dimensions (Nickerson et al. 2013); each dimension consists of a set of two or more characteristics, such that for each object, each dimension has one and exactly one characteristic. This simple definition only allows flat dimensions. We also need hierarchical dimensions, where the characteristics (nodes) are grouped into categories. The highest category (root) comprises all characteristics. The other categories are subsets of the root. Formally, a taxonomy T is defined as:

$$T = \{\mathrm{Dim}_i,\ i=1,\ldots,n \mid \mathrm{Dim}_i = \{\mathrm{Cat}_{ij},\ j=1\ldots,k_i\} \mid \mathrm{Cat}_{i1} = \{\mathrm{Car}_{im},\ \mathrm{m}=1\ldots,p_i;\ p_i \geq 2\} \wedge \forall j \geq 2,\ \mathrm{Cat}_{ij} \subsetneq \mathrm{Cat}_{i1}\}$$

By convention, the first category (Cat_{i1}) is the root. Its name is the name of the dimension Dim_i. For flat dimensions, $k_i=1$.

3.2 Operations on taxonomies

In this section, we present a typology of operations, based on the definition of taxonomies presented above. We list all the possible operations on taxonomies (these operations may be combined to define more complex operations). When building a taxonomy, information sources (e.g. academic papers or case studies) are examined. Each information source may trigger zero, one, or several operations on the taxonomy. Our operations, presented in Table 1, help in defining formally the different steps of taxonomy building.

		Output			
		Nothing	Dimension	Category	Characteristic
Input	Nothing		Addition	Addition	Addition
	Dimension	Deletion	Splitting Merging	Demotion	Demotion
	Category	Deletion	Promotion	Splitting Merging Promotion Demotion	Demotion
	Characteristic	Deletion	Promotion	Promotion	Splitting Merging

Table 1. Operations on taxonomies.

As mentioned above, per our definition, a taxonomy is composed of three concepts: dimension, category and characteristic. The dimension is the first-level concept. Since categories group characteristics together, we consider them as second level concepts. Finally, the characteristic is the lowest level concept. Building a taxonomy iteratively in an emerging field, where the concepts are not well established, implies the ability to transfer a concept from one level to another. Thus, besides the operations of addition, deletion, merging and splitting proposed by Nickerson et al. (Nickerson et al. 2013), we introduce promotion and demotion operations. Promotion takes place when a concept is transferred to a higher level, e.g. a category becomes a dimension. Conversely, demotion refers to the case where a concept must move to a lower level, e.g. a category becomes a characteristic. Thanks to these new operations, we don't use deleting operations, since a new concept can only enrich the existing taxonomy. As shown in Table 1, an operation is characterized by its type (addition, deletion, splitting, merging, promotion, or demotion), the type of its input(s) and the type of its output(s).

Let us illustrate a concrete example of combining these operations, based on the case of taxonomy development for BI and analytics on the cloud (described more thoroughly in Section 4). We assume that the taxonomy contains the dimension "BI application devices", with the characteristics "Mobile" and "Desktop". A new source of information mentions three different mobile types: smartphones, tablets and laptops. To take this new information into account, we perform the following operations:

1. Promotion of the characteristic "Mobile" to category.
2. Addition of the category "Mobile type" (with the characteristics "Smartphone", "Tablet", and "Laptop") as a sub-category of the dimension "BI application devices".
3. Merging of the categories "Mobile" and "Mobile type".

3.3 Sources and selection criteria

Emerging technologies are often located at the intersection of several areas. In order to constitute a taxonomy as a first body of knowledge for these emerging technologies, we recommend a literature review. The objective of comprehensiveness leads to a systematic literature review (SLR). In such a context, a SLR should focus not only on these emerging technologies, but also on each area of the intersection. As an example, BI and analytics on the cloud needs to explore both the BI/analytics domain and the field of cloud computing.

In addition, in order to extend the studies conducted by the authors of academic publications, it is important to know what is happening thanks to these emerging technologies in practice. In really emerging technologies, even recent research papers may not address correctly these issues. Moreover, the empirical experience based on these emerging technologies, if validated by professional experts, may serve as a foundation. Thus, we argue that the SLR should also include professional articles and case studies.
Thus, we have ordered the sources as follows:

1. Academic papers on the emerging technology (sources of type 1),
2. Academic papers on the related domains (sources of type 2),

3. IT magazines (sources of type 3).

This order matters. It helps in improving the quality of the resulting taxonomy. The main concepts are to be found in the academic literature and help in structuring the first levels of the taxonomy. If these concepts are not already established, the second source may be the papers relating the research on the domains on which the emerging technology is based, for example BI/analytics on the one hand and cloud computing on the other hand. Finally, IT magazines usually publish the notable experiments based on emerging technologies.

With the help of a set of keywords, we can query the electronic databases of these sources to collect papers. Of course, not all the results are relevant for the research question. SLR principles include the definition of criteria for making the decision to keep an article in the sample. Chitu Okoli et al. (Okoli & Schabram 2010) propose the following criteria:

- *Content (topics or variables)*: The review must be limited to studies that have bearing on the research question.
- *Publication language*: Reviewers can only review studies written in languages they can read.
- *Journals*: the scope of the review might limit itself to a set of high quality journals, or include only journals in a particular field of study.
- *Authors*: The study might be restricted to works by certain prominent or key authors (potentially including the reviewer).
- *Setting*: Perhaps only studies conducted in certain settings, such as healthcare institutions, or the financial services industry, might be considered.
- *Participants or subjects*: Studies may be restricted to those that study subjects of a certain gender, work situation, age, or other relevant criteria.
- *Program or intervention*: There might be a distinction made based on the nature of the intervention in the study.
- *Research design or sampling methodology*: Studies might be excluded based on the fact that they don't use a particular research design.
- *Date of publication or of data collection, or duration of data collection*: Studies will often be restricted to certain date ranges.
- *Source of financial support*: Studies might be restricted to those receiving non-private funds, in order to avoid bias in the results.

These criteria may be adapted to the selection of sources for developing taxonomies of emerging technologies. More specifically, the following criteria apply:

- *Content*: The topic of the source (journal article, case study…) should be directly related to the emerging technology, or to one of its related domains. This is ensured by the choice of keywords used in searching for the sources.
- *Publication language*: The search may be limited to publications in English.
- *Journals*: To ensure selection of high-quality sources, we propose to select the academic publications from the ERA list of journals and conferences in computer science and information systems ("http://www.arc.gov.au/era/era_2012/archive/era_journal_list.htm). These lists are well accepted and often used in the field. This guarantees quality. However, it may be insufficient in emerging technologies. The potential lack may be compensated by other sources (sources of type 3 as mentioned above).
- *Research design or sampling methodology*: To focus the search on technology-related papers, we propose to limit the scope to design-science papers (Hevner et al. 2004).
- *Date of publication:* Since emerging technologies may quickly become obsolete, we propose to limit the search of academic publications to the last ten years (our experience in taxonomy development for BI and analytics on the cloud even shows that 5 years may be sufficient). For professional publications (sources of type 3), a smaller time span may be used (e.g. one year).

To limit the scope of the study, the persons performing the literature review may also add certain quality criteria. When building a reference taxonomy we have to choose items that have an impact factor greater than a determined threshold, especially for academic papers. Regarding professional articles, the choice can be based on the size of the company and/or on its image.

Based on these criteria, we briefly describe below the process for selecting sources of type 1 (academic papers on the emerging technology), type 2 (academic papers on the related domains), and type 3 (IT magazines). The selection criteria need to be adapted to each type of source.

3.3.1 Selection of academic papers on the emerging technology

Apart from Google Scholar, the main sources of academic papers relating to emerging technologies are ACM Digital Library, IEEExplore, DBLP, ScienceDirect, EBSCOhost, and the electronic library of the AIS (AISeL).

Based on our experience, we suggest limiting the search to the first 50 pages of results. The choice of keywords depends on the emerging technology (the key words used in our case of taxonomy development are detailed in Section 4). The search results should then be integrated (deduplication): several papers are cited in multiple sources (in particular, most papers are cited in Google Scholar, which may be used as the single source if time constraints are high). After this integration, the papers that are not journal of conference papers from the ERA list are eliminated. The abstracts of the papers are also screened to ensure that they are design-science papers. One heuristic for determining if a paper is a design-science paper is to look for the artifact or artifacts contributed by the paper. To this end, a detailed typology of artifacts (Sangupamba Mwilu et al. 2014) may be used.

When building the taxonomy, it makes sense to start with the papers considered as key references in the field. Therefore, the sources are examined according to their decreasing order of citations in Google Scholar.

3.3.2 Selection of academic papers on the related domains

The process for selecting these papers is similar to the process described above for academic papers on the emerging technology. The difference lies in the choice of keywords: the keywords are specific to each of the related domains, and also ensure that only the most fundamental papers for understanding the domains (literature reviews) are selected.

Similarly to sources of type 1, the sources are examined according to their decreasing order of citations in Google Scholar.

3.3.3 Selection of IT magazines

While the processes for selecting sources of type 1 and 2 are quite similar, for sources of type 3 (IT magazines), the process is specific. We need to use specific sources of information and adapt the source selection process.

To determine the list of IT magazines, we propose to search for lists of these magazines on Google, based on specific keywords (e.g. “Best IT magazines”), and integrate the resulting list.

Once the list of IT magazines has been determined, a search is performed on Google for each magazine, with the URL of the Web site of the magazine as domain name. The same set of keywords may be used as for sources of type 1, combined with a keyword to restrict the search to concrete examples or cases (keywords “example” or “case”). To limit the search space, for each combination of keywords and each magazine, only the first page of results is kept. For each magazine, the result pages from the search are then integrated (deduplication), and each result is screened to ensure that it is relevant for the taxonomy development (e.g. sources that are very short and contain no information related to the subject are eliminated).

When building the taxonomy, the sources are examined according to their rank in the Google result page (if two sources have the same rank, they are considered as relevant as each other, and may be examined in a random order or a commonly used order e.g. the alphabetic order of the source name).

For sources of type 3, it is especially difficult to determine a priori the number of sources that will be required for building the taxonomy. We suggest to start with a pilot test (restriction of the search to the first result page, as mentioned above), and then to extend the search based on the results of this pilot test, keeping only the sources (IT magazines) that have contributed to the modification of the taxonomy in the pilot test.

As a conclusion, we propose to define our strategy of literature review by targeting three objectives: comprehensiveness, robustness and feasibility. For comprehensiveness, we ensure 1) the richness of the sources, 2) the variety of sources. For robustness, we order the sources according to their academic reputation. Thus, IT magazines are taken into account at the last step. Finally, for feasibility reasons, we first build on the first source where some structuring results may already be found and, if these results are not sufficient, the second source will help in completing the taxonomy. Our approach is illustrated in the following section.

4 APPLICATION TO THE DEVELOPMENT OF A TAXONOMY OF BUSINESS INTELLIGENCE AND ANALYTICS ON THE CLOUD

In this section, we apply the adapted method for taxonomy development to a taxonomy of BI and analytics on the cloud. As mentioned above, the starting point of taxonomy development is the choice of a meta-characteristic (Nickerson et al. 2013). In our case, the aim of the taxonomy is to support the decision of migrating BI to the cloud. This decision may be made for an organization or a division of an organization. Hence, the taxonomy must contain information relevant to support the decision of migrating to the cloud. This is our meta-characteristic. More specifically, we have two taxonomies: one taxonomy relating to the elements that influence the choice of migrating BI to the cloud, and another one dedicated to the choice itself (i.e. the options available to a decision maker wishing to migrate to the cloud).

We detail below, for each type of source, the process for selecting the sources (application of the generic process detailed in Section 3), and the result of source examination (i.e. for each type of source, the operations performed). To conclude the section, we show the two resulting taxonomies. We should point out that the development of these taxonomies is still a work in progress.

4.1 Academic papers on the emerging technology

BI and analytics in the cloud is an emerging technology, thus our sample of sources of type 1 is constituted of articles at the intersection of BI/analytics and the cloud.

4.1.1 Selection of sources

We have applied the generic principles described in Section 3 for the selection and ordering of examination of sources. The following set of keywords has been used: (« Business Intelligence » OR « Analytics » OR « Data warehouse ») AND (« Cloud » OR « SaaS »). The first set of terms refers to BI and its synonyms, and the second refers to the cloud and its most typical layer (SaaS). For Google Scholar, which does not allow very complex queries, the query had to be decomposed as follows:

- "Cloud" AND ("Business Intelligence" OR "Analytics" OR "Data warehouse")
- "SaaS" AND ("Business Intelligence" OR "Analytics" OR "Data warehouse").

Originally, after querying the databases mentioned above, we got a list of 1030 papers. After integration, deduplication, and elimination of non ERA and non-design science papers, the list was reduced to 46 papers. We examined these papers by decreasing number of Google Scholar citations, performing operations on the taxonomy as needed. Due to space limitations, we only provide the references of the first ten papers (the other references are available upon request). The first ten references are, in decreasing order of Google Scholar citations: (Cao et al. 2011), (Demirkan & Delen 2013), (Chi et al. 2011), (Baars & Kemper 2010), (Mian et al. 2013), (Abelló et al. 2011), (Wang et al. 2010), (Muriithi & Kotzé 2013), (Chang 2014), (Ng et al. 2011).

4.1.2 Operations on the taxonomies

The examination of the 46 sources resulted in a total of 47 operations on the two taxonomies. Figure 1 below illustrates the cumulated number of operations after each iteration (in our approach, an iteration is the examination of a source). In iteration 4, a significant number of operations were performed. Over the last iterations, the curve tends to converge.

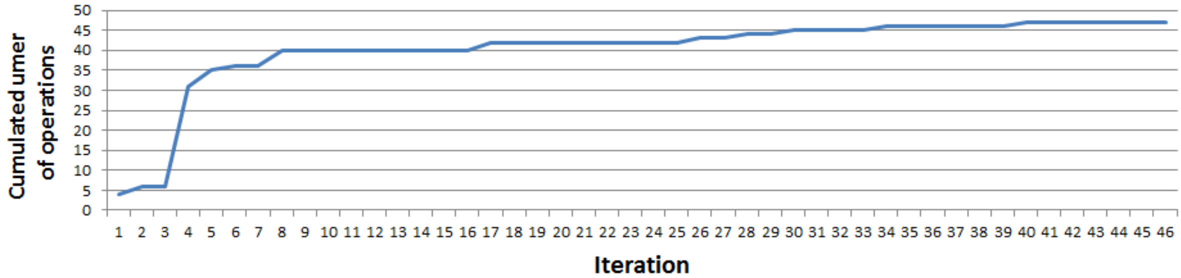

Figure 1. Examination of source 1 - Number of resulting operations

4.2 Academic papers on the domains

4.2.1 Selection of sources

In this stage, we were interested in reference papers in the domains from which BI and analytics on the cloud derive, namely the domains of BI/analytics and cloud computing. Consequently, we performed two queries on the same databases as in step 1, using the following keywords:

- (“Overview” OR “State of the art” OR “Survey” OR “Literature review” OR “Systematic Literature review”) AND (“Business Intelligence” OR “Analytics” OR “Data warehouse”)
- (“Overview” OR “State of the art” OR “Survey” OR “Literature review”) AND (“Cloud” OR “SaaS”).

After applying the same filtering criteria as in step 1, we ended with 19 references. The complete list of references is available upon request. The first four references are, in decreasing order of citations: (Subashini & Kavitha 2011), (Rimal et al. 2009), (Dinh et al. 2013), and (Negash 2004).

4.2.2 Operations on the taxonomies

The examination of the 19 articles did not add much information in our taxonomy because source of type 2 is composed of the literature reviews which have generally been cited in the articles of the first type of source and vice-versa. Most of the articles confirmed the already defined dimensions or categories or characteristics. Figure 2 illustrates that the examination of the 19 papers only resulted in two operations.

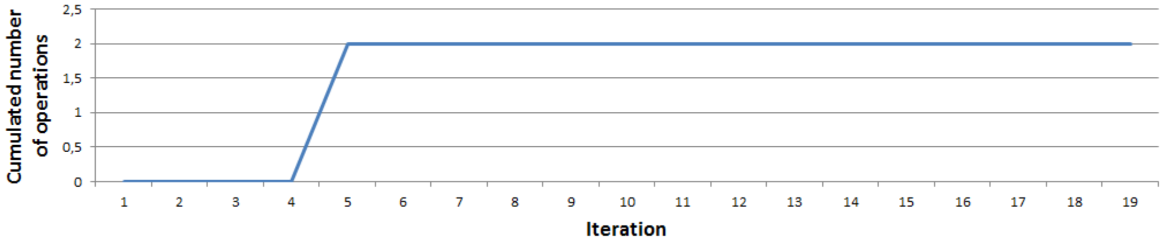

Figure 2. Examination of source 2 - Number of resulting operations

4.3 IT magazines

We are at the intersection of two domains: BI/analytics and cloud computing. For emerging technologies like this, it is especially important to consider knowledge gained from practice and case studies. This is all the more important as the theoretical body of knowledge on emerging technologies is often at its initial stage of development. For emerging technologies, we use IT magazines as a source of knowledge accumulated and formalized by practitioners.

4.3.1 Source selection

According to the general process described in Section 3, we started with the selection of IT magazines. To this end, we performed a Google search using the following keywords: “Best IT Professional magazines” and “Best IT magazines”. We navigated links pertaining to comparisons between the sites to help make our selection. After examination of these links, we kept five sites of magazines: InformationWeek, ComputerWorld, CIO, ZDNet, PCWorld.

As suggested above, we started with a pilot test (search on the web sites of the five magazines, restricting the search to the first page of results). In a second phase, we extended the search by considering the first four pages of results. However, we limited ourselves to the magazines that had provided information resulting in the modification of the taxonomy during the pilot test. This resulted in keeping only the sources InformationWeek and ComputerWorld, since the other three sources did not contribute to the modification of the taxonomy in the pilot test. We chose the number of four pages empirically, in order to have approximately the same number of sources as in the pilot test. Finally, we ended with 34 sources for the first phase (pilot test), and 32 sources for the second phase.

4.3.2 Operations on the taxonomy

As illustrated by Figure 3, the initial phase (pilot test) resulted in 12 operations, thus showing the relevance of considering sources of type 3 for building taxonomies of emerging technologies.

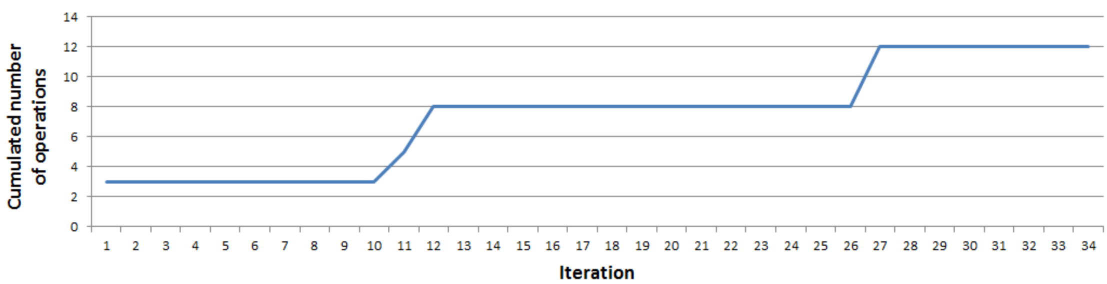

Figure 3. Examination of source 3 – Phase one (pilot test) - Number of resulting operations

Phase 2 (Figure 4) resulted in 9 operations. Based on this result, we conclude that it would be necessary to extend the search again to more fully consider the professional knowledge accumulated on the topic of BI/analytics on the cloud. We leave this for further research.

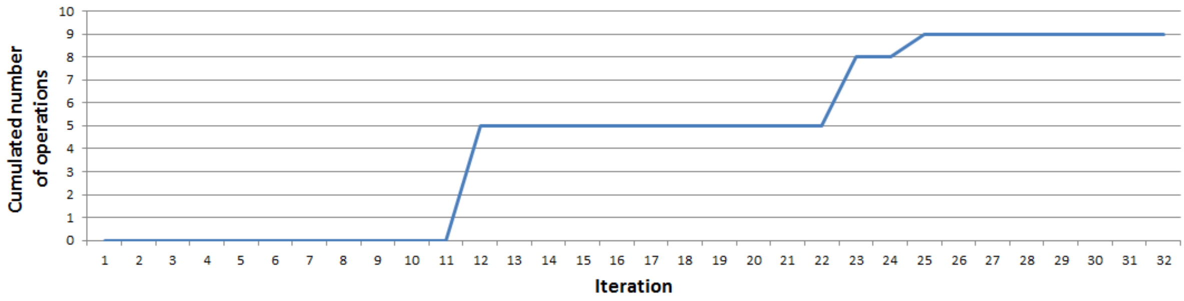

Figure 4. Examination of source 3 – Phase two - Number of resulting operations

4.4 Synthesis by type of operation and type of source

We analyzed the number of operations by type of source (1, 2, or 3) and by characteristics of operations (type of operation, type of input, and type of output). The most noteworthy differences between the three types of sources concern the output of operations. We show in Table 2 the number of operations by type of source and type of output, and comment these results in the discussion and conclusion below.

Number of operations by		Sources				
		1	2	3	Total	
Output	Category	12	0	7	19	27%
	Characteristic	21	1	10	32	46%
	Dimension	14	1	4	19	27%
	Nothing	0	0	0	0	0%
	Total	47	2	21	70	100%
		67%	3%	30%	100%	

Table 2: Number of resulting operations – Global view

4.5 Resulting taxonomies

We illustrate below the taxonomies resulting from the examination of the different sources. As mentioned previously, these taxonomies are still under development. To complete them, we need to extend the search of professional sources (sources of type 3), and to check the subjective ending conditions proposed by Nickerson et al. (Nickerson et al. 2013). The first taxonomy pertains to the context of the decision of moving to the cloud, i.e. to all the elements that influence this decision. It is represented below (Figure 5). While the first taxonomy represents the context of the decision of migrating to the cloud, the second one represents the object of the decision (i.e. the different options available to a company wishing to move to the cloud). This second taxonomy is shown below (Figure 6).

5 DISCUSSION AND CONCLUSION

Taxonomies help structuring a body of knowledge. They often also constitute the first step in ontology building. In this paper, we proposed a set of guidelines to help researchers in elaborating taxonomies for complex emerging technologies. The latter generally encompass two or more domains. We proposed a set of recommendations whose main principles are: 1) adopt systematic literature review guidelines aiming at a comprehensive state-of-the-art on this emerging technology, 2) enlarge the scope by considering references in the two or more background domains, 3) extend the review to IT magazines. The taxonomy is built using an iterative process. The latter stops when a convergence degree is attained, i.e. no more concept may be valuably added. Other guidelines are proposed defining

1) the order to follow when exploring the sources, 2) the number of references to be screened on each source, and 3) how to check the reputation of a given source.

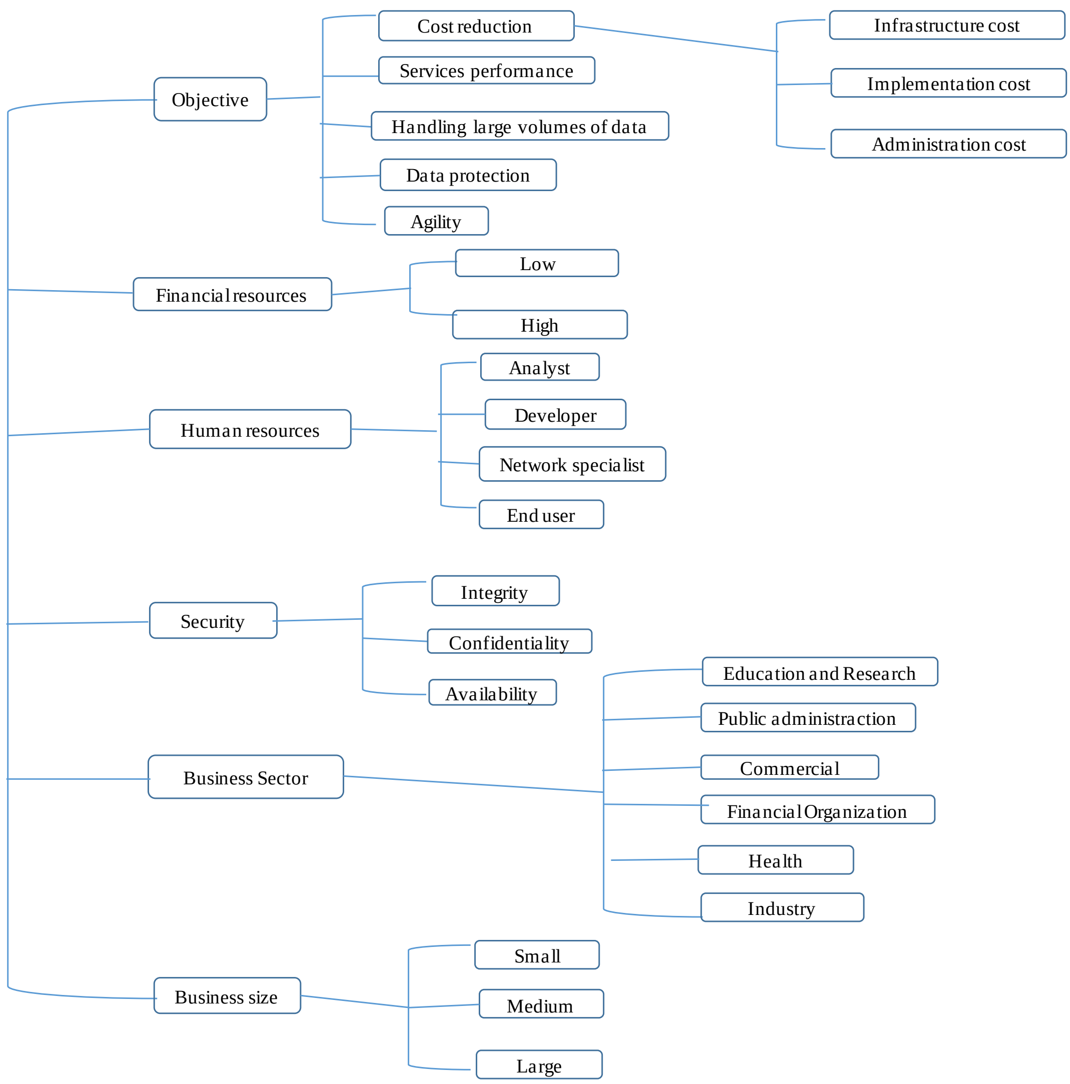

Figure 5. Elements that influence the decision of migrating to the cloud

The two taxonomies described in Section 4 are a first validation of our guidelines. Table 2 shows that main concepts (67%) result from the exploration of source 1. This is an interesting result, since the quality of source 1 is high, being composed of ERA journal and conference papers. The exploration of source 2 only leads to 3% additional concepts, due to the fact that the emerging technology (BI on the cloud) is not completely new. It could probably have led to a very different result if we had chosen a 2015 hot topic. Finally, source 3 allows us mainly to elicit characteristics, which means that we keep the well-structured dimensions and categories of previous sources and only enrich them with empirical results illustrated in IT magazines. We built this set of guidelines starting from Nickerson et al.'s methodology. At each step of the iterative process, Nickerson proposes basic operations on the taxonomy. Another contribution of our paper is the enrichment of this set of operations to provide the researchers with a more formal definition of the taxonomy. Future research will include, inter alia, the definition of an algebra guaranteeing the completeness and the soundness of this set of operations.

Our paper proposes an enrichment of Nickerson et al.'s methodology to the context of emerging technologies. The application of this methodology to the BI on the cloud is a first step of evaluation of our enrichment. This illustrative scenario allowed us to assess the feasibility and the operationality of our proposition. It also shows its utility, which is the first objective of design-science research (Hevner et al. 2004). We could not conduct a relative evaluation of the original methodology compared with our enrichment. However, we developed this enriched method and the associated guidelines since we were unable to obtain a reliable taxonomy just applying Nickerson's recommendations. Moreover we are convinced that combining Nickerson's methodology and SLR are an efficient way to tend to completeness, since systematic review aims at avoiding bias.

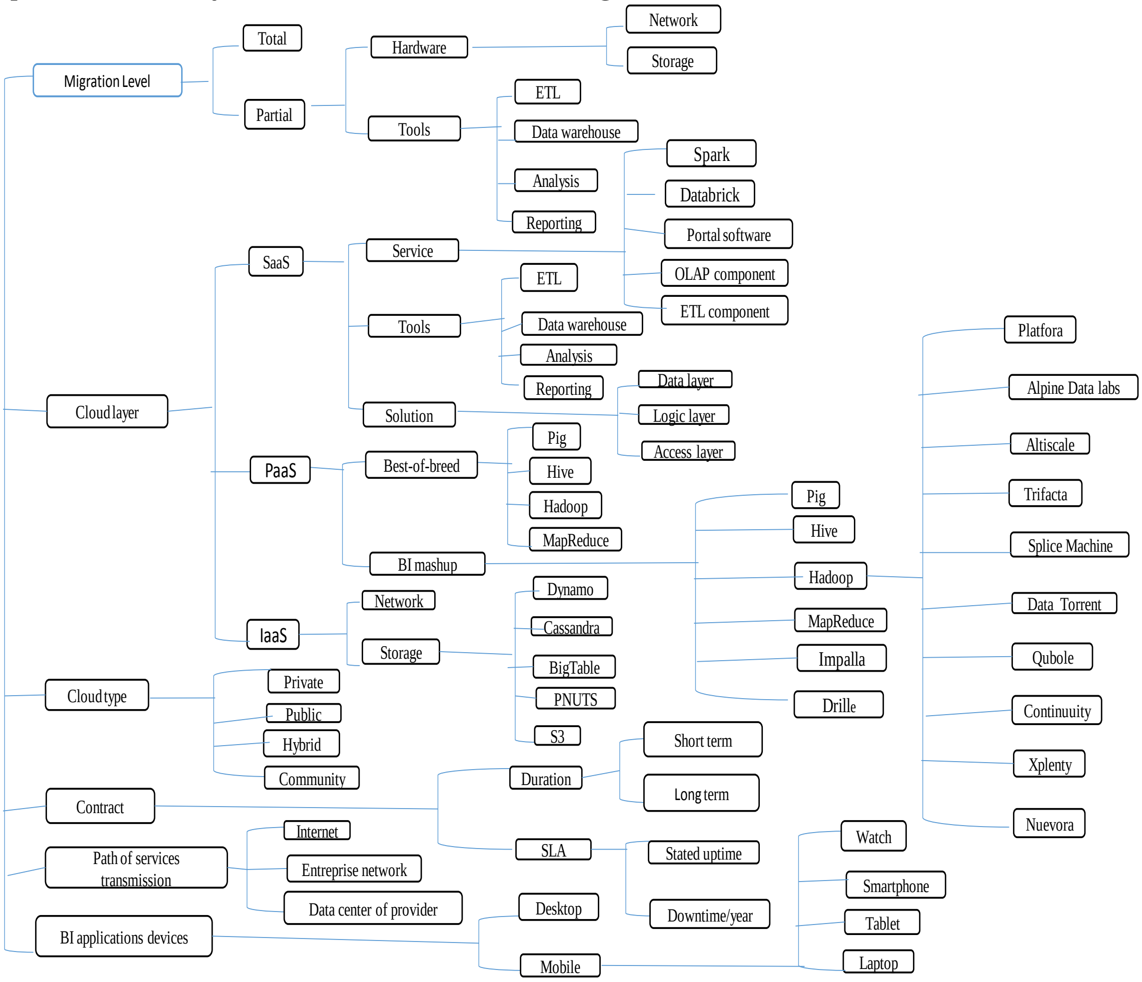

Figure 6. Aspects of BI migration

Finally, this paper proposes two complementary taxonomies for Business Intelligence (BI) on the Cloud. Our main research questions are: how can we help a company deciding to move its BI/analytics on the cloud? What are the questions to be answered? The taxonomy is only the first part of the answer. It should help in defining the factors that should impact the decision (business size, business sector, etc.) and also the different elements characterizing how to migrate on the cloud (cloud type: private, public, hybrid, etc., cloud layer: PaaS, IaaS, SaaS, etc.). The paper described the resulting taxonomies. In their current state, they enable a structured description of the migration problem. For space reasons, we cannot include the valuable information collected through the different papers. This information takes the form of either requirements or guidelines for helping companies deciding if and how they have to move their BI on the cloud. The taxonomies will serve as a reference for organizing the decision process based on these requirements and guidelines. The next step of our research is to complete this taxonomy and use it as a basis for designing a guidance tool for companies that consider moving their BI and analytics to the cloud.

References

Abelló, A., Ferrarons, J., & Romero, O. (2011). Building cubes with MapReduce. In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP* (pp. 17–24). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2064680

Baars, H., & Kemper, H.-G. (2010). Business intelligence in the cloud? Retrieved from http://aisel.aisnet.org/pacis2010/145/

Cao, Y., Chen, C., Guo, F., Jiang, D., Lin, Y., Ooi, B. C., … Xu, Q. (2011). Es 2: A cloud data storage system for supporting both oltp and olap. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on* (pp. 291–302). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5767881

Chang, V. (2014). The business intelligence as a service in the cloud. *Future Generation Computer Systems*, *37*, 512–534. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167739X13002926

Chi, Y., Moon, H. J., Hacigümüş, H., & Tatemura, J. (2011). SLA-tree: a framework for efficiently supporting SLA-based decisions in cloud computing. In *Proceedings of the 14th International Conference on Extending Database Technology* (pp. 129–140). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1951383

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*(1), 412–421. http://doi.org/10.1016/j.dss.2012.05.048

Dinh, H. T., Lee, C., Niyato, D., & Wang, P. (2013). A survey of mobile cloud computing: architecture, applications, and approaches: A survey of mobile cloud computing. *Wireless Communications and Mobile Computing*, *13*(18), 1587–1611. http://doi.org/10.1002/wcm.1203

Herzfeldt, A., Hausen, M., Briggs, R. O., & Krcmar, H. (2012). Developing a Risk Management Process and Risk Taxonomy for Medium-Sized IT Solution Providers. In *ECIS* (p. 165). Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1164&context=ecis2012

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science Research in Information Systems. *MIS Quarterly*, *28*(1), 75–105.

http://www.arc.gov.au/era/era_2012/archive/era_journal_list.htm, http://www.core.edu.au/index.php/conference-rankings. (2013).

http://www.forbes.com/sites/groupthink/2014/12/19/six-business-intelligence-predictions-for-2015/. (2014).

Land, L., Smith, S., Pang, V., & others. (2013). Building a Taxonomy for Cybercrimes. In *PACIS* (p. 109). Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1109&context=pacis2013

López, T. S., Ranasinghe, D. C., Patkai, B., & McFarlane, D. (2011). Taxonomy, technology and applications of smart objects. *Information Systems Frontiers*, *13*(2), 281–300. http://doi.org/10.1007/s10796-009-9218-4

McKinney Jr, E. H., Charles, J., & Yoos, I. I. (2010). Information About Information: A Taxonomy of Views. *MIS Quarterly*, *34*(2), 329–344. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=2908&context=misq

Mian, R., Martin, P., & Vazquez-Poletti, J. L. (2013). Provisioning data analytic workloads in a cloud. *Future Generation Computer Systems*, *29*(6), 1452–1458. http://doi.org/10.1016/j.future.2012.01.008

Muriithi, G. M., & Kotzé, J. E. (2013). A conceptual framework for delivering cost effective business intelligence solutions as a service (p. 96). ACM Press. http://doi.org/10.1145/2513456.2513502

Negash, S. (2004). Business intelligence. *The Communications of the Association for Information Systems*, *13*(1), 54. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais

Ng, W. S., Kirchberg, M., Bressan, S., & Tan, K.-L. (2011). Towards a privacy-aware stream data management system for cloud applications. *International Journal of Web and Grid Services*, *7*(3), 246–267.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, *22*(3), 336–359. Retrieved from http://www.palgrave-journals.com/ejis/journal/v22/n3/abs/ejis201226a.html

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Available at SSRN 1954824*. Retrieved from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1954824

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). *A taxonomy of IS artifact evaluation methods: design science prescriptions versus design research practice. Submitted for publication*.

Rimal, B. P., Choi, E., & Lumb, I. (2009). A Taxonomy and Survey of Cloud Computing Systems (pp. 44–51). IEEE. http://doi.org/10.1109/NCM.2009.218

Sabherwal, R., & Grover, V. (2009). A taxonomy of political processes in systems development: Political processes in systems development. *Information Systems Journal*, *20*(5), 419–447. http://doi.org/10.1111/j.1365-2575.2009.00341.x

Sangupamba Mwilu, O., Prat, N., & Comyn-Wattiau, I. (2014). *Business Intelligence and Big Data in the Cloud: Opportunities for Design-Science Researchers. Advances in Conceptual Modeling - Proceedings of ER 2014 workshops. Atlanta, GA, USA, pp.75-84*.

Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, *34*(1), 1–11. http://doi.org/10.1016/j.jnca.2010.07.006

Wang, Y., Song, A., & Luo, J. (2010). A MapReduceMerge-based Data Cube Construction Method (pp. 1–6). IEEE. http://doi.org/10.1109/GCC.2010.14