Electron Markets (2011) 21:129–139
DOI 10.1007/s12525-011-0060-4

SPECIAL THEME

Quality of data standards: framework and illustration using XBRL taxonomy and instances

Hongwei Zhu · Harris Wu

Received: 26 August 2010 / Accepted: 6 April 2011 / Published online: 27 April 2011

Abstract The primary purpose of data standards is to improve the interoperability of data in an increasingly networked environment. Given the high cost of developing data standards, it is desirable to assess their quality. We develop a set of metrics and a framework for assessing data standard quality. The metrics include completeness, relevancy, and a combined measure. Standard quality can also be indirectly measured by assessing interoperability of data instances. We evaluate the framework on a data standard for financial reporting in United States, the Generally Accepted Accounting Principles (GAAP) Taxonomy encoded in eXtensible Business Reporting Language (XBRL), and the financial statements created using the standard by public companies. The results show that the data standard quality framework is useful and effective. Our analysis also reveals quality issues of the US GAAP XBRL taxonomy and provides useful feedback to taxonomy users. The Securities and Exchange Commission has mandated that all publicly listed companies must submit their filings using XBRL. Our findings are timely and have practical implications that will ultimately help improve the quality of financial data and the efficiency of the data supply chain in a networked business environment.

Keyword Information quality · Data quality · Data standards · XBRL · US GAAP taxonomy

Responsible editor: Boris Otto

H. Zhu (✉) · H. Wu
Department of Information Technology and Decision Sciences,
College of Business and Public Administration,
Old Dominion University,
Norfolk, VA 23529, USA
e-mail: hzhu@odu.edu

JEL Classification M40 - General, Accounting and Auditing

Introduction

Data standards are used to reduce schematic and semantic heterogeneity and to ensure interoperability of data from multiple sources. There have been successful large-scale data standardization efforts such as those within the US Department of Defense (Rosenthal et al. 2004) and across the residential mortgage industry (Markus et al. 2006). As a community-based collaborative approach to information management, data standards have the potential to improve the efficiency of the data supply chain in an increasingly networked business environment.

A data standard is expected to improve the quality, especially the interoperability, of data created by different organizations using the standard. To accomplish this objective, the data standard itself must have high quality and there should be effective methods for assessing its quality. This is important to both developers and users of the standard given that it is often very costly to develop and implement a data standard. To this end, we introduce the notion of data standard quality and develop an assessment framework.

A data standard is metadata that specifies the characteristics of data elements and their relationships. Metadata is also data to those who use the metadata to create data instances. Thus data quality concepts also apply to data standards. Data quality is defined as data's fitness for use (Wang and Strong 1996). Since one of the primary purposes of data standards is to produce highly interoperable data, quality of data standards can be assessed by the interoperability of the resulting data.

We intend to make three contributions in this paper. First, we develop a set of metrics and a framework for assessing data standard quality. Second, we provide an empirical evaluation of the framework using real-world data in eXtensible Business Reporting Language (XBRL) (XBRL International 2006). An XBRL taxonomy is a data standard that defines numerous data elements to be used by companies to report and share their financial statements. The U.S. Securities and Exchange Commission (SEC) has adopted the GAAP (Generally Accepted Accounting Principles) taxonomy and mandated that public companies must submit their financial statements using XBRL and the GAAP taxonomy. The taxonomy was developed by XBRL US and is currently maintained by the Financial Accounting Standards Boards (FASB). A company is allowed to extend the GAAP taxonomy by creating and using its own data elements. We call these elements custom elements. Our empirical analysis is performed using a set of real-world financial statements submitted to the SEC. Preliminary findings of this research have drawn attention from the XBRL community and allowed us to participate in XBRL US committee activities to provide guidance on the use and evolvement of GAAP taxonomy. Thus the third contribution of this paper is that the findings offer timely insights, which will help the XBRL community improve the design and avoid improper uses of the GAAP taxonomy. As a result, the efficiency of the financial data supply chain will be improved to support better decision-making in a networked business environment.

The rest of the paper is organized as follows. After reviewing extant data quality research, we first describe the framework for assessing the quality of data standards and the empirical methods we developed to evaluate the framework. Then we present empirical findings by applying the framework to the XBRL GAAP taxonomy and the corresponding data instances of close to 500 companies. And lastly, we discuss related XBRL research and point out directions for future research.

Quality of data standards

Most data quality research focuses on data, not the standards used to create and organize the data. Data quality is a multi-dimensional concept that goes beyond accuracy. Prior research has identified numerous dimensions (e.g., consistency, interpretability, completeness, relevancy, etc.) of data quality (Gasser and Stvilia 2001; Wang and Strong 1996). Data quality perceived by users of different roles within an organization can be assessed using survey instruments (Lee et al. 2002). Quality of database schemas is discussed in (Redman 1996). Although a database schema is a type of data standard, it is mainly used within a single organization to organize and store data in a database. In contrast, the main objective of many data standards is to allow for meaningful exchange of data among multiple organizations so that the data from different organizations are interoperable. Quality of metadata for digital contents has been investigated in the past (Ochoa and Duval 2009; Park 2009; Bruce and Hillmann 2004), but the quality dimensions are not well defined and their metrics are operationalized differently. For example, the following seven dimensions proposed in (Bruce and Hillmann 2004) are often used in the domain of metadata for digital contents: completeness, accuracy, conformance to expectations, logical consistency and coherence, accessibility, timeliness, and provenance. A careful examination of their definitions shows that completeness and conformance to expectations overlap and both contain composite concepts concerning whether everything needed has been specified and whether everything that has been specified is needed.

Metrics of standard quality

The quality of a data standard is its fitness for multiple users to produce highly interoperable data. Like data quality, data standard quality has multiple dimensions. Further research is needed to determine these dimensions. At the minimum, it has completeness and relevancy dimensions.

For data quality, completeness is defined as “the extent to which data are of sufficient breadth, depth, and scope for the task at hand”, and relevancy is defined as “the extent to which data are applicable and helpful for the task at hand” (Wang and Strong 1996). Schema completeness and pertinence (i.e., relevancy) are defined similarly in (Redman 1996). These definitions need to be adapted for data standard quality. Our definitions are:

- *Completeness* of a data standard is the extent to which the data standard specifies all the data elements needed by users of the standard.
- *Relevancy* of a data standard is the extent to which the data standard specifies only the data elements needed by users of the standard.

Clearly, our definitions explicitly consider users and how users use the standard. This is generally true for all quality dimensions in the “contextual” category (Wang and Strong 1996) since these dimensions depend on the user’s context. The completeness and relevancy of the same data standard can be different to different users. Further, they can be different between an individual user and the user community.

To formalize the metrics, let S be the set of data elements specified in the data standard, U_i be the data elements required by the user i. From the user i's perspective, the metrics can be defined as

$$Completeness_i = \frac{|U_i \bigcap S|}{|U_i|}, \text{ and } Relevancy_i = \frac{|U_i \bigcap S|}{|S|}$$

From the user community's perspective, the metrics can be defined as

$$Completeness_c = \frac{|(\bigcup_i U_i) \bigcap S|}{|\bigcup_i U_i|}, \text{ and } Relevancy_c = \frac{|(\bigcup_i U_i) \bigcap S|}{|S|}$$

A standard can be complete by specifying every possible data elements, but it will suffer from low relevancy because many of the specified data elements may not be needed by most users. Conversely, a standard can be highly relevant by only specifying crucial data elements that are absolutely needed by all users, but it will be incomplete because it does not specify data elements only needed by a few users.

A measure that combines completeness and relevancy is the harmonic mean:

$$\mathrm{F} = 2^{*}\mathrm{completeness}^{*}\mathrm{relevance}/(\mathrm{completenes} + \mathrm{relevancy})$$

The above measure is analogous to the classic F-measure, often used to evaluate the effectiveness of information retrieval. Completeness and relevancy correspond to "recall" and "precision", respectively, in the information retrieval literature (van Rijsbergen 1979).

Methods for assessing data standard quality

Like data quality, data standard quality can be assessed both subjectively and objectively. We discuss various methods for assessing data standard quality below.

Subjective methods

Data quality perceived by various stakeholder such as data users can be assessed using a survey instrument (Lee et al. 2002). Likewise, survey method can be used to assess quality of data standards perceived by different stakeholders, which, at the minimum, should include users and developers of data standards. To develop a robust survey instrument, it is necessary to conduct interviews and case studies to understand how data standards are developed and used in certain application domains. Such studies will help identify quality dimensions most important to stakeholders. Survey method has been used to assess the quality of metadata for digital contents (Zhang and Li 2008; Palavitsinis et al. 2009).

Although it is possible to produce a general-purpose survey instrument, it may be desirable to adapt the instrument for different applications. For example, the stakeholders can change from domain to domain. For XBRL, it may be necessary to distinguish different users, such as technology providers, auditors, filers, analysts, and investors.

A subjective method involves humans in the loop and can identify issues not discoverable using objective methods. Since issues identified using a subject method represent "a human voice", they tend to receive attentions that can lead to actions and solutions. On the other hand, subjective methods typically need to be complemented and validated by objective methods. We will develop subjective instruments in future research.

Objective methods

There are three different objective methods for assessing quality of data standards:

(1) manual inspection of a standard's fitness;
(2) direct measurement of quality metrics; and
(3) indirect assessment by measuring interoperability and other aspects of data instances created using data standards.

Manual inspection This method has been used to examine if an earlier version of XBRL taxonomy meets the reporting needs of various companies (Bovee et al. 2002). The annual financial statements (SEC form 10-K) of 67 companies are examined to see whether the taxonomy contains data elements for all the line items in these statements. Although the method is labor intensive, it is necessary in the development phase of a data standard. By examining data instances produced by many organizations, the standard developer can inductively determine what data should be captured in the data standard. On the other hand, the amount of manual effort required by this method significantly limits its use for quality assessment purposes. It is impossible to use this method for large, complex data standards.

Direct measurement This method relies on a set of well-defined metrics and automated measurement techniques to obtain direct measurement of standard quality. Later in this paper we will show how to measure the two metrics (i.e., completeness and relevancy) by collecting and analyzing a large sample of data instances created using a given standard.

Indirect assessment Standard quality can be assessed indirectly by examining the extent to which it helps to accomplish the objectives of developing the standard. The main objective of most data standards is to enable interoperability of data created by different organizations.

Thus we can also indirectly assess standard quality by examining interoperability of a sample of data instances.

A standard often defines a uniform syntax for data representation. Thus it is relatively trivial to achieve syntactic interoperability. In addition, a standard also defines a set of data elements with their semantics agreed-upon by all users. In certain cases, this can help attain semantic interoperability as well. But in many other cases, such as when the standard contains many elements and users are allowed to extend the standard by redefining elements or introducing new elements, semantic heterogeneity problems can arise to hinder interoperability.

In this paper, we focus on the comparability aspect of semantic interoperability. A set of data instances is interoperable if the instances use the same set of data elements defined in a data standard. Interoperability measures the extent to which the data instances have overlapping data elements defined in a standard. This definition allows us to measure interoperability directly without relying on unreliable semantic matching techniques (E. Rahm and Bernstein 2001; Erhard Rahm et al. 2004). However, it is conservative and may underestimate interoperability.

The interoperability between a pair of data instances is based on the common data elements used. The interoperability between users i and j, $I_{i,j}$, can be defined as

$$I_{i,j} = \frac{|U_i \bigcap U_j|}{\sqrt{|U_i||U_j|}} \tag{1}$$

Apparently, $I_{i,j} = I_{j,i}$. The pair-wise interoperability for all users, I_2, is defined as the arithmetic mean of pair-wise interoperability among all pairs. This definition can be extended to interoperability of any k-tuple (with k>=2) as

$$I_{i_1,\ldots,i_k} = \frac{U_{i_1} \bigcap \ldots \bigcap U_{i_k}}{\sqrt[k]{|U_{i_1}| \ldots |U_{i_k}|}} \tag{2}$$

The k-interoperability of all users, I_k, can be defined as the arithmetic mean of the k-interoperability among all k-tuples. We will limit our discussion to I_2 and I_3 because interoperability calculation is computationally expensive. For I_k, there are $O(n^k)$ k-tuples that need to be computed.

When a user is allowed to extend the standard, U_i can be partitioned into two sets: U_i^s (elements from the standard) and U_i^c (elements custom-made by the user).

One concern about data standard-based approach to semantic interoperability is that if the standard defines too many data elements, the interoperability may decrease due to an oversupply of choices to users of the standard. Thus it is interesting to measure interoperability by just considering standard data elements. For this purpose, we define $I_{i,j}'$ and $I_{i,j,k}'$ by replacing all occurrences of U with U^s in Eqs. 1 and 2.

Methods for empirical evaluation

To assess the framework for data standard quality, we have collected, processed, and analyzed the XBRL GAAP taxonomy and all official XBRL filings submitted to the SEC as of February 26, 2010. We are able to evaluate the framework by applying the direct measurement and indirect assessment methods described in the preceding section. We have developed a set of methods and computer tools to support data acquisition, processing, and analysis of this research. These methods and tools are depicted schematically in Fig. 1.

A data acquisition agent monitors the RSS Feed from the SEC (feed://www.sec.gov/Archives/edgar/usgaap.rss.xml) to obtain company filings submitted to the SEC. The acquisition agent downloads the financial statements and the accompanying taxonomy extensions into a local filing repository. The ETL (Extract/Transform/Load) program parses the files downloaded and loads the extracted data into a relational database. Stored SQL procedures and other programs are used to analyze the data stored in the relational database. While we are continuously downloading new filings as they come in, the results reported here include all XBRL filings to the SEC as of February 26, 2010. The dataset includes 1,231 filings submitted by 481 companies. The GAAP taxonomy is also processed to load the specified elements into the relational database. Out of the 1,231 filings, 266 are annual statements (i.e., 10-K). Several companies have two 10-K's, with one being a revised 10-K overriding the previously submitted 10-K. We have constructed a sample containing only 10-K's by removing the submission being overridden. This limited dataset includes 261 10-K statements, each from a different company.

In XML and XBRL, a data element is identified by its name and name space. When a company extends the standard GAAP taxonomy by introducing new data elements, the elements have a name space unique to the company, typically labeled by the company's stock symbol. Thus even if two companies use the same name for their data elements, the elements are different because they have

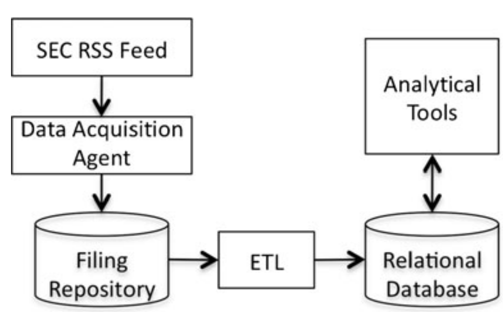

Fig. 1 Methods for data acquisition, processing, and analysis

different name spaces. Their statements may contain elements from namespaces other than GAAP or the company's custom namespace. However, most such elements (such as comments or addresses) are not essential financial information. Therefore we focus on the elements from GAAP and the custom namespaces in this paper.

Results of empirical evaluation

In this section, we first present certain characteristics of the GAAP taxonomy, followed by the main results on completeness, relevancy, and interoperability. Then we present element usage patterns that exhibit a long-tail distribution. In the last subsection, we present observations about potential misuse of the taxonomy such as unnecessary introduction of custom elements.

Characteristics of GAAP taxonomy

The GAAP taxonomy specifies a total of 13,452 data elements, among which 2,653 are abstract and 346 are deprecated (see Table 1). Abstract elements are used for deriving concrete data elements and cannot be used in company filings. All depreciated elements are deprecated on January 31, 2009 and not recommended for use in filings after January 31, 2009. 84 abstract elements are deprecated. Thus the number of concrete elements is 10,799, of which 10,537 are active (i.e., not depreciated).

Using our similarity tool we found that many of the deprecated elements have a corresponding active element. Their names tend to be permutations of the same set of words. For example, the deprecated element Cash Dividends has a corresponding active element Dividends Cash, and the depreciated element CashProvidedBy UsedIn<u>DiscontinuedOperations</u>*FinancingActivities* has a corresponding active element CashProvidedByUsedIn *FinancingActivities*<u>DiscontinuedOperations</u>. As highlighted using underline and italic font the second two elements are different in the sequence of DiscontinuedOperations and FinancingActivities. In total there are five cases where deprecated elements are word-permutations of active elements. Many companies continued to use deprecated standard elements. Overall 40.5% of the companies (195 out of 481) used deprecated elements in 19.82% of the filings (244 out of 1,231). It is not mandatory for companies to void using deprecated elements. However, to maintain longitudinal interoperability of financial statements, companies should not use deprecated elements because these elements are subject to removal from future releases of the taxonomy.

Table 1 Number of elements of different types in GAAP taxonomy

	Concrete	Abstract	Total
Active	10,537	2,569	13,106
Deprecated	262	84	346
Total	10,799	2,653	13,452

Completeness and relevancy of GAAP taxonomy

We can measure the metrics using only 10-K's or all the financial statements in the dataset. In either case, we need to analyze the statistics of standard and custom elements used in the sample. We report the results for both samples.

Let us first look at the 10-K's. All 261 10-K statements used the GAAP taxonomy as the base taxonomy. For each statement, we identified data elements specified in the GAAP taxonomy and those introduced by the filing company. Overall, the statements utilized 2,083 GAAP elements and 4,403 custom elements, among which 1,357 GAAP elements and 351 same-named custom elements were utilized in more than one statement. Generally, more taxonomy elements were used than custom elements in each statement. On average, a statement used 129 elements from the GAAP taxonomy and 20 custom elements. The distributions of number elements used are shown in Table 2.

To compute the metrics, we also need the size of standard, |S|. As discussed earlier, companies are still allowed (albeit discouraged) to use deprecated elements. Thus we use all concrete elements for |S|, which is 10,799. Now, we can compute the metrics for the user community and for the average user using the mean values in Table 2. The results are shown in Table 3. For the average user, the completeness is 0.8678, but the relevancy is only 0.0119 and the F-measure is 0.02. This means that the average user can find most of the elements they need from the taxonomy, but they only need a very small number of the elements and all the other elements in the taxonomy are not relevant. For the user community, the completeness is 0.3212, lower than that for the average user, but the relevancy is 0.1929, significantly higher than that for the average user. The lower completeness is due to the large number of custom elements introduced by the user community. The higher

Table 2 The number of standard and custom elements in all 10-K's

	Standard	Custom	Both
Min	41	0	50
Max	246	82	284
Mean	128.7	19.6	148.3
Median	125	16	142
Stand deviation	23.1	13.9	30.9

Table 3 Completeness and relevancy of GAAP taxonomy from perspectives of average user and user community, measured using 10-Ks

	Completeness	Relevancy	F
Average user	128.7/148.3=0.8678	128.7/10799=0.0119	0.02
User community	2083/(2083+4403)=0.3212	2083/10799=0.1929	0.25

relevancy is due to the fact that the user community uses a larger fraction of the taxonomy elements. The F-measure for the user community is 0.25, much higher than that for the average user. Therefore, the overall quality of the taxonomy for the user community as a whole is higher than that for the average user. We should note that the relatively low relevancy could be due to the fact that many GAAP elements are designed for detailed tagging, which begins after the first year's use of the GAAP taxonomy. With detailed tagging, all numeric values appearing in Notes sections must be tagged using either GAAP elements or custom elements. Our dataset does not contain any financial statements with detailed tagging.

Now let us use all of the 1,231 financial statements in the dataset. All companies together used 2,558 GAAP elements and introduced 10,168 custom elements. The statistics of data elements and the results of the two metrics are presented in Tables 4 and 5, respectively.

Note that the metrics obtained using the two different samples are similar. This indicates that the metrics are scalable, i.e., they do not change much as we use more measurement samples. This property is desirable for robust metrics.

The metrics for the average user is a usually a good approximation of the averages of the metrics measured using all users' financial statements. To illustrate this, we plotted the distribution of the metrics in Fig. 2. The averages are 0.8850, 0.0101, and 0.02 for completeness, relevancy, and the F-measure. The average relevancy and average F-measure are the same as completeness and relevancy, respectively, of the average user. The average completeness is a bit lower mainly because its distribution is not as close to a normal distribution as the other two metrics. The distribution in Fig. 2 also illustrates that the metrics measure contextual quality. Their values depend on the user and vary substantially among users. For example, for certain user contexts, the standard is only 45% complete, while for certain other user contexts, it is 100% complete.

Table 4 The number of standard and custom elements in all statements including 10-Q

	Standard	Custom	Both
Min	17	0	20
Max	246	7	385
Mean	109.5	15.6	125.1
Median	106	12	120
Stand deviation	24.1	13.7	32.3

Interoperability: Indirect measurement of quality of GAAP taxonomy

Out of the 261 annual financial statements, we have computed interoperability of 33,3390 pairs and 2,929,290 triples. The summary statistics of these interoperability scores are reported in Table 6. Note the "Mean" row has values corresponding to I_2, I_2', I_3, and I_3', as previous defined.

The first column of Table 6 shows a summary of the pair-wise interoperability. The highest interoperability score is 0.7646. On average, the interoperability score is only 0.3724. That is, investors can conveniently compare only about 37.24% of the financial information from two companies' statements in XBRL. In the next section we will list the top data elements that are most commonly used by companies, which provide more information on the comparability of the financial statements.

While allowing flexibility, the usage of non-standard elements certainly affects interoperability. Many custom elements extend GAAP elements to allow for more detailed, company-specific reporting. If investors are not going to consider company-specific elements when comparing companies' financial statements, the interoperability can be computed based on GAAP elements only. The results for this scenario are reported in the second column of Table 6. The interoperability score for the dataset is 42.50%.

The interoperability of three documents is expected to be lower than that of two elements (see the third and fourth columns of Table 6). On average, 24.35% of the financial information from three companies' XBRL statements is comparable. If only GAAP elements are considered, 27.81% of the financial information is comparable.

The interoperability results indicate that the standard taxonomy did not make financial statements fully interoperable. When three statements are compared, on average only a quarter of the elements are directly comparable. This indirectly shows that the quality of the standard taxonomy is not very high.

Element usage patterns

In addition to the statistics about the number of elements used, it is also useful to examine use frequency of each

Table 5 Completeness and relevancy of GAAP taxonomy from perspectives of average user and user community, measured using all statements

	Completeness	Relevancy	F
Average user	109.5/125.1=0.8753	109.5/10799=0.0101	0.02
User community	2585/(2558+10168)=0.2027	2585/10799=0.2394	0.22

element. We report the results for the 10-K's only. The trend of all statements is similar to that of the 10-K's.

An element can appear in a company statement more than once. For the usage pattern analysis, we use a binary counting method so that the usage frequency of an element is the number of companies that have used the element. In Fig. 3, the Y-axis is the usage frequency of an element, and X-axis is the frequency rank. For company-introduced elements, we treat the same-named elements used by different companies as the same, disregarding the name space. Both GAAP and custom elements have a long-tail distribution. Some elements are used frequently, but most of the elements are only used in a small number of company statements.

The top 25 most frequently used GAAP elements are listed in Table 7. The "Support" of an element is the fraction of documents in which the element is used. From a financial reporting point view, we expect that certain elements should be frequently used together to disclose information often expected in commonly used statements such as balance sheets, income statements, and cash flow statements. Data elements that are frequently used together can be identified using frequent item set mining, a technique used in association rule (also known as market basket) analysis. A maximal frequent item set for a given support is a frequent item set for which none of its immediate supersets are frequent. We have identified a series of maximal frequent item sets for different set sizes, using the Apriori algorithm in R (Hahsler et al. 2010). Figure 4 lists three maximal frequent item sets and their support (i.e., fraction of documents having these elements). For example, 91.95% of documents had the five elements shown in the upper left corner of the figure under the heading "5-tuple". The support decreases quickly with increase of the set size. Given the computational complexity of the Apriori algorithm, we were only able to compute the maximal frequent item sets for up to 22 elements, for which the support was 41%. In other words, only 41% of the companies can be compared for 22 standard data elements despite the use of the XBRL-based GAAP data standard. This comparability is low for analysts and investors to support decision-making in an increasingly networked businessenvironment.

Custom elements

It is possible that users do not utilize the standard properly even though the standard itself is well designed. When this

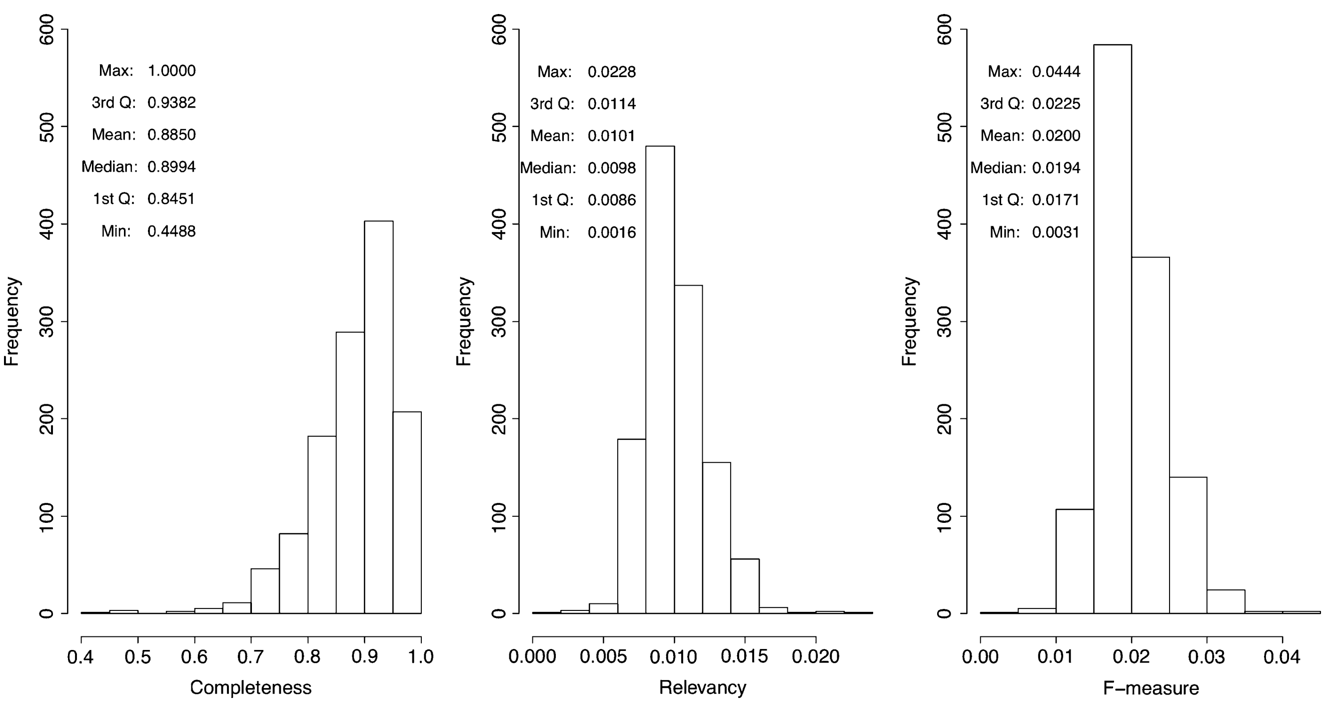

Fig. 2 Distribution of completeness, relevancy, and F-measure

Table 6 Statistics of interoperability among 261 10-K's

	$I_{i,j}$	$I_{i,j}'$	$I_{i,j,k}$	$I_{i,j,k}'$
Min	0.1033	0.1266	0.0300	0.0374
Max	0.7646	0.8655	0.5150	0.5809
Mean	0.3724	0.4250	0.2435	0.2781
Median	0.3798	0.4340	0.2456	0.2811
Standard deviation	0.0837	0.0859	0.0629	0.0668

happens, the assessment methods would underestimate the true quality of the standard. However, with certain standard enforcement in place, abuse of data standards is usually rare. In the case of XBRL, the SEC provides guidelines on how to use the GAAP taxonomy to produce financial statements. Software tools used by companies are also helpful to minimize misuses of the taxonomy. Thus in practice, the influence of users' misuses of a data standard on the quality measurements tends to be negligible.

We attempt to estimate the extent of duplication and misuse in custom definition of elements using a term analysis. For each element, English terms occurring in the element name are identified. Then we use cosine similarity between a custom element and standard element to identify potential duplicates. Out of more than 10,000 custom elements, only 87 have a similarity of 1 with a standard element, i.e., they contain the exact terms as a standard element does. Among them, 49 custom elements have the same name as a standard element, and the rest use a different ordering of the same set of terms of a standard element. Although these 87 custom elements (less than 1%) are potentially duplicates, the rest are quite unlikely. For example, out of the top 50 most used custom element names, 13 of them have a similar element in the GAAP taxonomy with a cosine similarity score greater than 0.8 (see Table 8). We manually inspected the definitions of these elements but could not positively identify any exact match in terms of accounting concept.

We understand the limitation of cosine similarity and other semantic matching algorithms may identify additional semantic matches between custom elements and GAAP elements. However, the number of matches is not expected to be very high. Many custom elements seem to be legitimate. In fact, the FASB has set up a team to analyze the custom elements to identify those that will be included in the next version of the standard.

Related work

XBRL has been used worldwide for various purposes. Thousands of papers have been written on the topic of XBRL (Roohani and Zhao 2009). In an earlier study (Bovee et al. 2002), manual inspection of line items in financial statements of 69 companies was used to examine whether an preliminary taxonomy met reporting needs of most companies. Manual inspection was also used more recently to examine if companies properly used XBRL taxonomies in the Voluntary Filing Program administered by the SEC during 2005 and 2009 before officially adopting XBRL (Bartley et al. 2010; Boritz and No 2008a, b). While this manual approach can reveal deep issues such as violation of reporting convention and inconsistency between XBRL version and non-XBRL version of financial statements, it is labor intensive. Therefore, only a small number of documents can be examined using this manual approach.

In (Chou 2006), approximately 100 XBRL financial statements were examined using commercial software tools. Errors and warning messages were used as indicators of improper use of XBRL technology. Despite the relative large sample examined, it was still a manual approach and

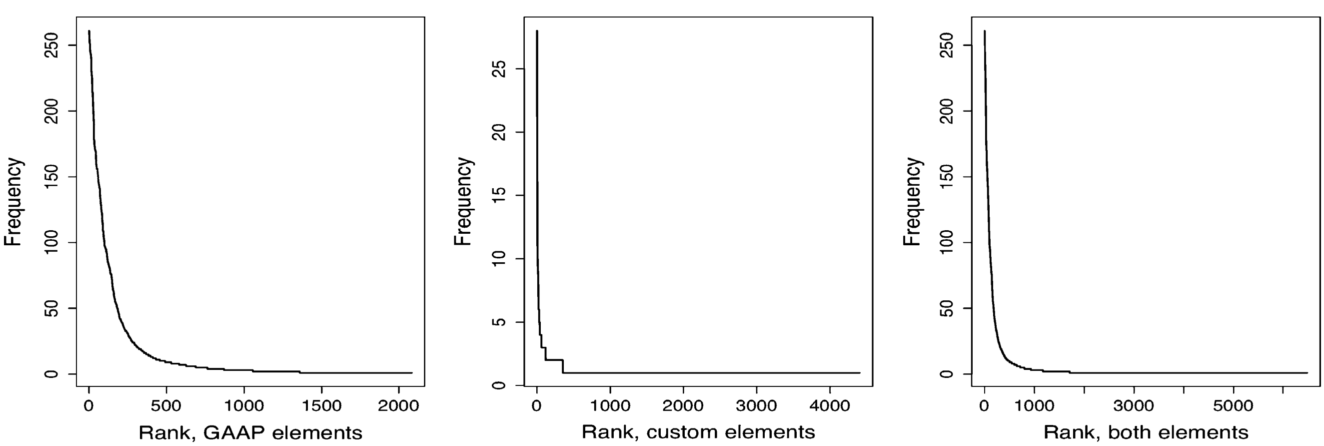

Fig. 3 Use frequencies of GAAP taxonomy elements and custom elements

Table 7 Top 25 most used GAAP elements

	Element	Support
1	Assets	1.0000
2	Liabilities And Stockholders Equity	0.9885
3	CashAndCashEquivalentsPeriodIncreaseDecrease	0.9808
4	Income Tax Expense Benefit	0.9655
5	Cash And Cash Equivalents At Carrying Value	0.9617
6	Income Tax Disclosure Text Block	0.9617
7	Retained Earnings Accumulated Deficit	0.9502
8	Accumulated Other Comprehensive Income Loss Net Of Tax	0.9387
9	Net Cash Provided By Used In Investing Activities	0.9349
10	Earnings Per Share Basic	0.9349
11	Earnings Per Share Diluted	0.9310
12	Net Cash Provided By Used In Operating Activities	0.9310
13	Net Cash Provided By Used In Financing Activities	0.9234
14	Property Plant And Equipment Net	0.9234
15	Common Stock Value	0.8889
16	Net Income Loss	0.8812
17	Segment Reporting Disclosure Text Block	0.8736
18	Assets Current	0.8697
19	Liabilities Current	0.8659
20	Commitments And Contingencies Disclosure Text Block	0.8621
21	StockholdersEquity	0.8352
22	Common Stock Shares Authorized	0.8199
23	Common Stock Par Or Stated Value Per Share	0.8199
24	Other Assets Noncurrent	0.8199
25	Common Stock Shares Issued	0.7931

its results depend on the implementation of the software tools.

To overcome these limitations, automated tools have been suggested to perform large scale analysis (Bovee et al. 2002; Bovee et al. 2005). Using an approach similar to the one used in this paper, financial statements of 140 companies participating in the SEC Voluntary Filing Program were analyzed in (Zhu and Fu 2009), and 478 companies' 1,231 official filings (both 10-Q's and 10-K's) were analyzed in (Zhu and Wu 2010). Both studies examined only relevancy and completeness of a data standard.

Conclusion and future research

Building on extensive work on data quality (Madnick et al. 2009), we have developed a framework for assessing quality of data standard and empirically evaluated this

Fig. 4 Maximal frequent item-sets of 5, 10, and 15 elements

5-tuple: 0.9195402
{Assets,
CashAndCashEquivalentsPeriodIncreaseDecrease,
IncomeTaxDisclosureTextBlock,
IncomeTaxExpenseBenefit,
LiabilitiesAndStockholdersEquity}

10-tuple: 0.7892720
{AccumulatedOtherComprehensiveIncomeLossNetOfTax,
Assets,
CashAndCashEquivalentsAtCarryingValue,
CashAndCashEquivalentsPeriodIncreaseDecrease,
EarningsPerShareBasic,
EarningsPerShareDiluted,
IncomeTaxDisclosureTextBlock,
IncomeTaxExpenseBenefit,
LiabilitiesAndStockholdersEquity,
RetainedEarningsAccumulatedDeficit}

15-tuple: 0.6590038
{Assets,
AssetsCurrent,
CashAndCashEquivalentsAtCarryingValue,
CashAndCashEquivalentsPeriodIncreaseDecrease,
EarningsPerShareBasic,
EarningsPerShareDiluted,
IncomeTaxDisclosureTextBlock,
IncomeTaxExpenseBenefit,
LiabilitiesAndStockholdersEquity,
LiabilitiesCurrent,
NetCashProvidedByUsedInFinancingActivities,
NetCashProvidedByUsedInInvestingActivities,
NetCashProvidedByUsedInOperatingActivities,
PropertyPlantAndEquipmentNet,
RetainedEarningsAccumulatedDeficit}

Table 8 Top custom elements and GAAP elements with similar names

Custom Element	GAAP Element	Sim
Prepaid Expense And Other Assets Current	Prepaid Expense Current	0.92
Interest Expense Net	Interest Income Expense Net	0.92
Interest Expense Net	Interest Expense Other	0.89
Income Loss From Continuing Operations Before Income Taxes And Minority Interest	Income Loss From Continuing Operations Before Income Taxes Minority Interest And Income Loss From Equity Method Investments	0.87
Comprehensive Income Net Of Tax Including Portion Attributable To Noncontrolling Interest	Other Comprehensive Income Loss Net Of Tax Portion Attributable To Noncontrolling Interest	0.87
Earnings Per Share Basic And Diluted	Earnings Per Share Diluted	0.85
Accrued Expenses And Other Current Liabilities	Other Accrued Liabilities Current	0.83
Adjustment Depreciation And Amortization	Depreciation And Amortization	0.83
Proceeds From Repayments Of Commercial Paper	Repayments Of Commercial Paper	0.81
Accrued Expenses And Other Current Liabilities	Accrued Liabilities Current	0.81
Income From Continuing Operations Before Income Taxes	Income Loss From Continuing Operations Before Income Taxes Domestic	0.81
Adjustment Depreciation And Amortization	Other Depreciation And Amortization	0.81
Increase Decrease In Income Taxes Net	Increase Decrease In Income Taxes Receivable	0.80

framework. In this initial step towards quality of data standards or metadata schemas, we have defined two separate metrics, completeness and relevancy. The metrics are applied to a real-world data standard. The results show that the metrics provide a useful measurement of the quality of data standards. Furthermore, our analysis of XBRL GAAP taxonomy and XBRL data has important and timely implications.

Our empirical analysis shows that, on average, approximately 37% of data elements between a pair of financial statements were common elements from the GAAP taxonomy. That is, out of the 148 elements, 129 were from the data standard. Out of the 129 standard elements, 55 of them were the same and could be directly compared between a pair of financial statements. From a standard developer's point view, this value may seem low because it would be nice if all elements were directly comparable. But with the size of the GAAP taxonomy, this value is not too bad. Given a data standard with 10,000 defined elements, if all elements have the same probability of being used, the expected interoperability of selecting 129 elements from the standard is only 0.01290 for pairs and 0.0001664 for triples. Because of the long-tail distribution, the observed interoperability is significantly higher than the case with a uniform distribution.

For future research, we will develop a comprehensive framework for data standard quality. The framework will consist of all important quality dimensions, metrics, and corresponding assessment methods. In addition, the dataset we collected will allow us to further validate the existing framework. We will analyze interoperability of companies from the same industry because they may have similar information to report. Due to limited computing resource, we did not perform any k-interoperability analysis for $k>3$ in this study. We will analyze interoperability of 4, 5 or even more companies, to see how much the interoperability decreases as a greater number of companies are compared. We will also closely examine the financial statements trying to gain deeper insights on why companies adopt different sets of elements from the GAAP taxonomy, and whether the adoption of elements was affected by different software tools they use.

We plan to conduct interviews with the reporting companies and accounting firms to further understand how interoperability of XBRL statements can be improved. For example, perhaps the software tools can make context-sensitive suggestions of elements to use, or prompt users to choose existing elements from the US GAAP Taxonomy when they try to define custom elements. Or, perhaps the companies can make coordinated efforts, such as through use of the same accounting firms, to converge in their choice of elements to use in financial reports. The GAAP taxonomy is fairly new to filing companies. Through learning, the companies may begin to use more elements from the taxonomy. Thus we will continue to gather and analyze new filings as they come in to see how companies learn and how the measurements of the metrics evolve over time. Further, company-introduced elements may converge over time. If this trend emerges, the GAAP taxonomy may consider adopting these converging elements introduced by the companies. Allowing companies to directly contribute to the standardization effort, such as by a wiki system, may help improve the next release of the US GAAP XBRL taxonomy.

In summary, we think we have opened an exciting area of information quality research. The work reported here represents an important step towards quality assessment of data standards. In addition to the financial domain, we will evaluate the framework in other domains with large data standards such as the HL7 standard in healthcare.

Acknowledgement This research is supported in part by Office of Research, Old Dominion University and US National Science Foundation Award #0713290.

References

Bartley, J. W., Chen, Y. A., & Taylor, E. Z. (2010). A comparison of XBRL filings to corporate 10-ks - evidence from the voluntary filing program. SSRN.

Boritz, E. J., & No, W. G. (2008a). Auditing an XBRL instance document: The case of united technologies corporation. University of Waterloo.

Boritz, E. J., & No, W. G. (2008b). SEC's XBRL voluntary program on edgar: The case for quality assurance SSRN: http://ssrn.com/abstract=1163254.

Bovee, M., Ettredge, M. L., Srivastava, R. P., & Vasarhelyi, M. A. (2002). Does the year 2000 XBRL taxonomy accommodate current business financial-reporting practice? *Journal of Information Systems, 16*(2), 165–182.

Bovee, M., Kogan, A., Nelson, K., Srivastava, R. P., & Vasarhelyi, M. A. (2005). Financial reporting and auditing agent with net knowledge (FRAANK) and extensible business reporting language (XBRL). *Journal of Information Systems, 19*(1), 19–41.

Bruce, T. R., & Hillmann, D. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D. Hillmann & E. L. Westbrooks (Eds.), *Metadata in practice* (pp. 238–256). Chicago: American Library Association.

Chou, K. H. (2006). How valid are they? An examination of XBRL voluntary filing documents with the SEC Edgar system. Paper presented at the 14th International XBRL Conference, Philadelphia, USA.

Gasser, L., & Stvilia, B. (2001). *A new framework for information quality*. Urbana Champaign: University of Illinois at Urbana Champaign.

Hahsler, M., Grun, B., Hornik, K., & Buchta, C. (2010). Introduction to arules—a computational environment for mining association rules and frequent item sets.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information and Management, 30*(2), 133–146.

Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and framework for data and information quality research. ACM Journal of Data and Information Quality, 1(1), Article #2.

Markus, M. L., Steinfield, C. W., Wigand, R. T., & Minton, G. (2006). Industry-wide information systems standardization as collective action: The case of the U.S. Residential mortage industry. MIS Quarterly, 30(Special Issue), 439–465.

Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International Journal on Digital Libraries, 10*(2/3), 67–91.

Palavitsinis, N., Manouselis, N., & Alonso, S. S. (2009). Evaluation of a metadata application profile for learning resources on organic agriculture. In F. Sartori, M. A. Sicilia, & N. Manouselis (Eds.), *MSTR 2009, CCIS 46* (pp. 270–281). Berlin: Springer.

Park, J.-R. (2009). Metadata quality in digital repositories: a survey of the current state of the art. *Cataloging & Classification Quarterly, 47*(3/4), 213–228.

Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *VLDB Journal, 10*(4), 334–350.

Rahm, E., Do, H.-H., & Maßmann, S. (2004). Matching large xml schemas. *ACM SIGMOD Record, 33*(4), 26–31.

Redman, T. C. (1996). *Data quality for the information age*. Boston: Artech House.

Roohani, S., & Zhao, X. (2009). XBRL citation analysis: A decade of progress and puzzle. Paper presented at the International Conference on XBRL, Lawrence, Kansas, USA, April 24–25, 2009.

Rosenthal, A., Seligman, L., & Renner, S. (2004). From semantic integration to semantics management: case studies and a way forward. *ACM SIGMOD Record, 33*(4), 44–50.

van Rijsbergen, C. V. (1979). *Information retrieval* (2nd ed.). London: Butterworth.

Wang, R., & Strong, D. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.

XBRL International (2006). Extensible business reporting language (XBRL) 2.1. XBRL International.

Zhang, Y., & Li, Y. (2008). A user-centered functional metadata evaluation of moving image collections. *Journal of the American Society for Information Science and Technology, 59*(8), 1331–1346.

Zhu, H., & Fu, L. (2009). Towards quality of data standards: Empirical findings from XBRL. Paper presented at the The 30th International Conference on Information Systems (ICIS'09), Phoenix, AZ, USA, December 15–19.

Zhu, H., & Wu, H. (2010). Quality of XBRL US GAAP taxonomy: Empirical evaluation using SEC filings. In 16th Americas Conference on Information Systems, Lima, Peru, August 12–15 2010.