

From Spreadsheet Trash Talking to Data Recycling Nirvana

BY GIANLUCA GARBELLOTO

Companies are increasingly being evaluated by their commitment to the environment and society. While the bits and bytes associated with business data don't fill up landfills per se, both archived data made obsolete by updated and replaced systems and exports and spreadsheets of data that are later manually reentered are wasteful and symbolic of unsustainable systems. In this

column I will discuss the value of data reusability and describe the stages of transforming your data from an untapped potential resource into a true contributor to the lasting competitiveness of your organization.

It may seem a bit of a stretch to compare promoting standardized data with recycling. It's often said that recycling is good business: It's good for the environment, it conserves natural resources, it reduces the need for landfills, and it involves changes in business processing and, perhaps, some compromises.

Moving toward standardized data in the business reporting supply chain is also good business. Standardized

data is important for establishing sustainable processes that can be carried out over and over without negative "environmental effects" or impossibly high costs. With our ongoing goal that a piece of business data, once entered into a computer, never needs to be retyped, we strive

for systems that are increasingly interchangeable and reusable—the stuff that Web services and service-oriented architecture (SOA) are all about.

Just as embracing sustainability is a process, so is embracing standardized data. Small efforts can bring tremendous benefit.

Costs and struggles without data sustainability

There's no doubt that retyping data as it flows across various information system components or across different systems is a major implicit cost that most entity or corporate environments often ignore or underestimate.

Traditional solutions to this problem are application centric. Enterprise resource planning systems claim to integrate all information system needs in one application, eliminating the need of "migrating" data, but they often fail to meet this goal. ETL (Extract, Transform, and Load) applications, recognizing the substantial failure of the ERP promise, claim to facilitate universal interoperability within all components of the information system. But



both solutions have shown their limitations (see the August and September 2006 XBRL columns for more information).

A data-centric view

Relying on the format in which data is represented, rather than on applications that “translate” that data into different formats, would solve the problem at its roots. A different approach to data reusability, based on the success of standards-based technologies, is indeed possible. This approach can be defined as *data centric*. Technology—or, more appropriately, an agreement involving the use of technology—is key to the data-centric approach. The terms of the agreement, as well as embedding domain knowledge into that agreement, however, are vital. XML (Extensible Markup Language) and XBRL (eXtensible Business Reporting Language) represent foundational technology agreements. XBRL GL, the standardized Global Ledger, is the key that embeds domain knowledge into these agreements.

XML has been incredibly successful as a standard format to represent data, making it easier to share the data across different applications. With XML, certain barriers to sharing data, such as agreements on the representation of numbers and dates, are lowered. XBRL is an agreement on how to use XML (and related technologies, such as XLink) for the specific purpose of representing business, financial, and accounting data. Additional benefits from XBRL include formalizing how to provide human-readable labels and definitions in numerous languages.

Through XBRL GL, XBRL can also be used to represent granular data at the document and transaction level. The working group devel-

oping XBRL GL can be likened to promoters of data conservation—working for more efficient technologies and to reduce wasteful practices.

Standards are about agreement. Embracing XBRL GL means embracing two parallel agreements. On one side is the XBRL Specification, which includes the rules on the structure and form of taxonomies and instances. On the other side is a standardized way to represent concepts, interrelationships, and enumerations (fixed choices for entries to facilitate understanding and data interchange) as found in source applications and ERP systems. This side has nothing to do with the technical format in which the data is expressed.

Even from this very high-level discussion, it should be evident that there are different ways to implement a data-centric approach to data reusability, just as there are different levels of data reusability that can be achieved. Can XML solve all data interoperability problems on its own? Is XBRL really more effective for the purpose? What is that specific purpose anyway?

From data trash to data sustainability

XML by itself does bring some benefit. Without broader agreement, however, it simply turns old legacy data into “new” legacy data. XBRL brings additional overhead along with the additional benefit of greater agreement. As for the purpose—the more reusable the data gets, the more purposes it can fulfill.

To describe the level of reusability of business data, here are the “Nine Degrees of Data Reusability.” They begin with the sheer wastefulness of data that is created for one-time use, manually reentered in whole or in

part in other systems, and then forgotten. In recycling terms, electrons may not hit the trash pile, but the paper reports that are produced so people can retype the data will most likely end up in a landfill somewhere. The Nine Degrees then move through intermediary steps that can bring some quick benefit. Finally, they end with the promise of full data sustainability, where standardized data not only can be reused for every conceivable purpose, but it can be evaluated more easily as input into new and more powerful processes. Instead of being expended, the data brings vast new potential benefits.

The “Nine Degrees of Data Reusability” are a type of measurement tool for gauging the effectiveness of different data-centric approaches. The tool provides a context to evaluate if and to what extent a certain solution is really effective and how the solution compares to others located up or down the same scale. Where is your business in making your data recyclable? Where are your software developers in building data conservation into their systems?

1. No reusability. Everybody retypes everything.

Even though this is obviously a dramatization, most entities usually are closer to this situation than they think. An ERP system fails to integrate one or more specialized components of the information system, spreadsheets start to appear to provide a certain representation of data for specific purposes, and data needs to be shared with external parties that can’t “consume” it in their proprietary format. This scenario should sound familiar to many.

2. Format reusability. All data represented with XML.

As a standard format to represent data, XML is obviously an important step forward. Other formats aren't as inherently reusable: Fixed-length ASCII doesn't carry meaning; CSV files fail when exchanging them between the U.S. and Europe; EDI isn't extensible. XML brings agreement on date and numeric formats and has validation capabilities and many other benefits for human and machine exchange.

But having data in the same format, even standardized, doesn't automatically mean being able to give the data the same meaning. Data represented with XML isn't "understandable" by two different systems unless both systems agree on how to use XML to represent the data. Each entity typically develops its own proprietary XML schema with no easy way to automatically map between the different schemas. So this degree really means only technical reusability.

3. Specification reusability. All data represented with XBRL.

With XBRL, the business rules that give meaning to the data are packaged with the data itself and are defined in a standard way. XBRL brings an agreement on formally assigning labels, definitions, linkages to authoritative and practical guidance, and the expression of certain expressions of formulas. In this respect, it's better than XML for the purpose of reusing business data, but it still poses a similar issue: Data isn't really reusable unless the same data dictionary—in this case, a taxonomy—is shared.

4. Structural reusability—common vocabulary. All data represented with XBRL GL.

XBRL GL is the XBRL taxonomy that represents a worldwide agreement on how to standardize business

and financial information as it can be found in any ERP system from the moment in which it is first entered into a system up to the end reporting in XBRL or other standard or proprietary XML schemas. As mentioned above, embracing XBRL GL involves two parallel things: the semantic agreement on concepts, relationships, and enumerations that describe business and financial data along with the technical agreement on how to represent this within the XBRL specification.

In this respect, the XBRL part of XBRL GL acts as an enabler that allows businesses (and the software developers who support them) to implement the semantic agreement (the GL part of XBRL GL) with a technology that is universal and very effective.

Someone could make an argument that the semantic agreement of the Global Ledger would have the same value if it were represented through CSV or text files. An export from an ERP system in a CSV format where the first line, or headers, of the CSV file included the associated element names from XBRL GL would be very helpful, but it would still be subject to the lack of precision of CSV files. Likewise, a monolithic XML schema that uses XBRL GL's element names, hierarchical structures, and enumerations to drive document creation should require only mechanical transformation to turn into XBRL GL representations. These benefits may be important if the technologies are a better short-term fit for the organization implementing them instead of XBRL GL.

5. Structural reusability—common grammar. Representing XBRL GL instances the same way.

In some human languages, there

are differences in word endings for the subject and direct object of a sentence. In others, placement in the sentence (subject before the verb; direct object afterwards) helps the listener understand. Those of us who still cringe at having to diagram sentences in school can appreciate the idea that reducing the number of different ways the same concepts can be stated makes it easier to understand the expression of those concepts.

XBRL GL has been designed to be able to semantically represent the same information according to the specific features of the source data and the purpose of the representation. In particular, some people use general ledger accounts to store certain information, while others use special codes. One example is the broad European use of general ledger accounts for customers and vendors compared to the separate account listings common in the U.S. The ability for XBRL GL to act as a true audit file is a good thing in many respects, but it obviously can pose a reusability problem if different approaches can be used to represent the same type of data; XBRL GL also strives to be the ideal data exchange format. This degree of reusability—the "common grammar"—is about agreeing on how to use XBRL GL's representational power in a consistent way, building best practice profiles and templates that add an additional layer of standardization.

6. Mapping reusability. Company code sets are mapped externally.

Once data is represented with XBRL GL, it's far easier to move from one system to another. XBRL GL brings with it standardization at the database-field level and key fields at the content level. Within an orga-

nization, reusing information may require crossing from an account used at a division to the consolidating account at headquarters or bringing together vendor data where a different vendor code is used in different subsystems.

An organization can begin to inventory and match these code sets across systems and make them available for applications to query in some manner of repository. This external mapping can then be used to support the consolidation process and aid analysis. An application would find the account number or customer number of an inventory part in the data file, look up the consolidating number in the external mapping, and then add that number to the file (or replace the existing number—XBRL GL generally supports both approaches) to aid consumption and analysis.

7. Semantic reusability. Establishment and use of XBRL reporting taxonomies with XBRL GL.

While common internal code sets are helpful within an organization, true data reusability will have to span organizational boundaries. That's one reason XBRL has been uniquely designed to properly associate business reporting data with the increasingly visible XBRL financial reporting (and similar externally focused) taxonomies; you can associate accounts and data-entry lines with the standardized code lists of, for instance, the U.S. GAAP taxonomy. No matter what your internal numbering sequence and descriptions are, an external party would know you were referring to the agreed-upon concept that represents cash, sales, or cost of goods sold within that taxonomy.

XBRL GL allows the referencing of multiple taxonomies so consum-

ers can better understand the meaning under different end-reporting scenarios. While XBRL reporting taxonomies may provide the best shared understanding, XBRL GL also lets this detail be associated with other standard and proprietary schema formats, such as the IRS tax forms.

8. Code reusability. Common code sets used across instances everywhere.

The next logical step is for the continued development of common code sets that can be used in the many data fields that are exchanged and then immediately in the data files, reducing or eliminating the need for external maps.

In the U.S. government, for example, there's an agreement on accounts (the U.S. Standardized General Ledger) and on transaction sets based on those accounts and transaction identifiers. The use of these accounts and entry classifications, entered in a standardized fashion and represented with XBRL GL, will go a long way toward the reusability of data in government systems.

As the market embraces and reuses other existing code sets, or as they are developed for many interesting uses (which will be discussed in future columns), data will become increasingly more independent of the system producing it and, consequently, more readily interpretable by consuming systems. For certain code sets that are unique by their very nature (e.g., my warehouse locations will be different from yours), other standards such as geospatial coordinates may come into play. As our systems become more interdependent, this will become a requirement for efficient operations.

9. Ultimate reusability. Data is fully independent of systems.

With full data sustainability, it's impossible and unnecessary to figure out the data's source. The output from any system is theoretically exactly the same as from any other system. At this point, data flowing through your information system is like blood in your body's circulatory system: Wherever you put a needle, the same blood comes out.

Moving forward

There's obviously no all-purpose solution to full data sustainability today, which makes the above spectrum a hypothetical experiment at the present time. While total reusability of data sounds like a good thing, and "no reusability" sounds like a bad thing, both are extremes. Each degree of reusability between these extremes can have advantages and disadvantages in different situations (e.g., cost of implementation, new or heightened security issues to overcome), but being able to identify the options and position one entity's particular situation in a clear context can help build an effective data-centric strategy tailored to each entity's unique needs.

Is your business reaping the benefits of data sustainability? Are the steps you are taking to improve your business reporting limited to short-term gains in order to deal with the urgent need, or are they establishing new processes that will be true contributors to the lasting competitiveness of your organization? ■

Gianluca Garbellotto is an internationally known expert on both the business and the technical aspects of XBRL and XBRL Global Ledger and a member of the XBRL Standards Board. Gianluca can be contacted at gg@iphix.net.