UNITED NATIONS
STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEN COMMUNITIES
(EUROSTAT)

Work Session on Electronic Raw Data Reporting
(Geneva, 6-8 November 2006)

Electronic Raw Data Reporting and XBRL

Olivier Servais, XBRL Europe

Electronic Raw Data Reporting and XBRL

Executive Summary

XBRL - the Extensible Business Reporting Language - is an emerging XML-based, freely-licensed format developed by a Consortium (XBRL International, Inc.[1]) made of public authorities, business organizations and commercial companies. Based on the specifications provided by the Consortium, participants in different business reporting supply chains come together to agree on *taxonomies*, hierarchical business dictionaries of the terms, attributes and interrelationships of the concepts in different types of business reporting. One such taxonomy, the standardized Global Ledger (XBRL GL), is particularly focused on raw data. There are many barriers to analyzing information across different formats and representation; agreeing to represent all of the information in one normalized form using *XML* and especially XBRL GL overcomes many of these barriers immediately.

How raw do *you* like your data?

Raw data, also known as *source data* is data that is collected but not yet "processed" or interpreted for use. Raw data is usually collected in a database for later reporting (including sorting, filtering, summarizing and formatting) and analysis. Raw data is that which can not be derived from other data. Statistics are drawn from the raw data.

XBRL - the Extensible Business Reporting Language - is a meta-language developed by a consortium[2] of nearly 500 organizations, including :

- Public authorities, including central banks, banking supervisors, securities regulators, tax authorities, and statistical organizations;
- Business organizations including accounting institutes, banking federations, analysts associations, and investor representatives, and;
- Commercial companies including audit firms, software and service vendors, and companies wishing to publish their own business reporting internally and externally.

These organisations, on a worldwide basis, are collaborating on improving communications throughout the Business Reporting Supply Chain (BRSC) through agreement on technical specifications that leverage W3C[3] Recommendations for XML[4] (Extensible Markup Language). Based on these specifications, participants in different business reporting supply chains come together to agree on *taxonomies*, which are XML-based code books, acting as highly structured and formalized business dictionaries, for the business concepts in their BRSC, including descriptions, attributes and interrelationships between each concept, and create *instances*, which is business data coded with codes from the relevant taxonomy.

Most of the taxonomies are focused on reports, information that has been processed and expressed in agreed-upon sets of terms. This includes financial statements and reports, banking reports, and tax returns. Each taxonomy contains agreement on relevant business reporting concepts, descriptions to better understand each concept - often available in multiple languages, linkages to authoritative reference material or practical guidance, and other tools to help in comparability and consistency of reporting.

---

[1] http://www.xbrl.org
[2] http://www.xbrl.org/AboutTheOrganisation/
[3] The World Wide Web Consortium (W3C) can be found at www.w3c.org
[4] http://www.w3.org/XML/

One taxonomy, developed and promulgated by XBRL International itself, is particularly focused on raw data. That taxonomy is XBRL GL[5], the standardized Global Ledger. This extensible, modular taxonomy has been designed to represent the information on business and trade documents and provide a single view of detailed data found within business and operational systems. A goal of its creators is that XBRL GL will be the drill-down data from any XBRL (reporting) taxonomy, the bridge from other transactional standards to end reporting.

How does raw data enter a system?

Information enters computer systems today through many channels and in many forms. While some applications will require access to the original entry vehicle, most analytical applications will require normalization of the data into one standard view to be able to work with data across the various import forms. XBRL GL has been designed to be that standardized view.

Data is captured and entered into computers in many forms, increasingly in an automated fashion. Where manual entry (and manual re-entry, unfortunately) directly into applications still occurs in many applications, electronic entry has been increasing. Electronic entry takes on many forms. These include:

- Electronic data interchange (EDI)
- XML-based exchange (such as ACORD for insurance transactions)
- Web-based data gathering
- Telephony-based input (voice, touchtone)
- Fax with optical character recognition
- Bar code
- RFID (radio frequency)
- Other forms-based (Adobe PDF, Excel templates) solutions provided to external parties to begin electronic entry

An audit function may require access to the *original* electronic format and related data files which are sent to the applications for import; the original is necessary in order to check for authorization, data integrity and other security-oriented issues.

Most statistical uses, however, will benefit from normalization of the input formats. The barriers to analyzing information across these different forms are many.

- File formats
- Identifying records within the files
- Identifying data fields within the records
- Understanding the content of the data fields, including
  - Disparate date formats
  - Disparate numeric formats
  - Inconsistencies between representation of information in other key fields

Agreeing to represent all of the information in one normalized form using *XML* overcomes some of these barriers immediately. XML would be the single file format; a competitive format, CSV, is not consistent internationally. XML also brings agreement to how to represent dates[6] and numbers[7] internationally, as well as

---

[5] More information on XBRL GL can be found at www.xbrl.org/GLTaxonomy; samples can be found at http://gl.iphix.net with annotated examples for analysis.
[6] Leveraging the ISO date format standard, ISO 8601

providing a standardized convention (lang="xx") for identifying text in multiple languages.

Additional agreement on representing records and agreeing on common data fields across raw data entries requires more than is inherent in XML itself. That is where XBRL GL, the standardized Global Ledger, comes in to play.

About XBRL GL

The Global Ledger has been designed to offer a common way to code the source information that flows into a business.

- It provides standard codes for the attributes of a business document from which that data is collected. This includes information such as document type, document number, and document date.
- It offers a standard set of tags for the parties involved in trade and business transactions - customers, vendors, employees, salespeople, contractors. This would include internal party identifier, external party identifiers (such as governmental assigned IDs), party name, address, and contacts.
- Of course, as part of the world of XBRL, it can represent all of the accounting-oriented information you would expect are necessary for the accounting, finance and tax reporting processes.

One of the most flexible representation capabilities of XBRL GL is that of measures and metrics. You might think of this as the lines on an order, invoice or statistical collection document.

Each measurable item includes

- A measurable *Code*, signifying whether the item is inventory, supplies, services, metrics, a fixed asset or some other qualitative or quantitative statistic
- A measurable *Category*, a free-form field for classification
- A measurable *Identification*, an internal code used to identify this particular item or measure, along with a related *Schema* for Measurable Identification, such as a UN/CEFACT code set.
- Along with a secondary measurable Identifier and Schema
- A measurable *Description,* which can be collected in multiple languages
- A measurable *Quantity* field and related *Unit of Measure*
- An associated *Qualifier* field for non-numeric measures in raw data (e.g., "Grade A", "successfully completed"), which can be used in conjunction with or instead of the Quantity field
- A *per unit cost/price*
- For services, machine time or other time-oriented events, a measurable *Start Time* and *End Time*.
- An Activity tag, to note whether an item is retired, obsolete or otherwise not an *Active* item.

Through the agreed-upon use of the Measurable structure as part of XBRL GL, almost any kind of raw data that a statistical organization would require as part of business reporting can be represented, with references to existing and emerging

---

[7] Should the comma (",") or the period (".") be used as the decimal separator? Is there a thousands separator? Should you use a plus sign ("+")? The answer can be found at http://www.w3.org/TR/xmlschema-2/

UN/CEFACT and external code sets (such as those a statistical organization may define itself) to describe the necessary measures.

For data fields that may be required in addition to those provided by XBRL GL today, XBRL GL is a modular, extensible framework overseen by a working group within XBRL International, Inc., and additional modules can judiciously be created to add necessary representational capabilities. [8]

Ties to end reporting

XBRL GL is the bridge from transactional input to end reporting. As such, it provides the underlying details in a single, standardized (normalized) format for organizations that require raw data in an easily consumed format. In addition, XBRL GL also defines the way to associate lines of raw detail with end reporting concepts in order to more efficiently provide links to the end reports, whether serving as the basis for summarization, or as the means to provide a *Seamless Audit Trail* from end reporting assertions back to the original source document. This understanding of XBRL GL as the glue for business information was recognized by the world's tax administrations in their support for the development of XBRL GL.

One standard for many types of raw data

As we noted previously, there are many barriers to bringing raw data together from disparate systems in order to analyze and reuse that data. XML overcomes many of those, especially through agreement on numeric formats, date formats, and other file level issues. XBRL GL helps overcome understanding the meaning of different fields and brings consistency to much of the content. In conjunction with other agreed-upon code-sets, such as XBRL's end user reporting concepts, exported raw data will be far more consistent regardless of the originating system, facilitating its reuse.

Conclusion

XBRL GL, the standardized **G**lobal **L**edger, is a standard available today to represent the raw data that exists within an organization. It works in cooperation with other standards and code sets and especially with XBRL end reporting. It is useful not only for the exchange of raw data, but for the business environment in which that raw data exists, and is part of a holistic solution, leading to superior audit functionality related to raw data.

Written on 12 October by
- Eric Cohen (PricewaterhouseCoopers Director)
  Eric.E.Cohen@us.pwc.com
- Olivier Servais (XBRL Europe, Director)
  olivier.servais@xbrl-eu.org

---

[8] http://www.oasis-open.org/committees/download.php/14242/OASIS_XML_Position_Paper_for_Tax_Administrations_v2-01.pdf