ENABLING TRANSPARENCY & TRUST IN FINANCIAL DATA THROUGH SEMANTIC DATA QUALITY RATING SYSTEM

Ashu Bhatnagar
GoodMorningResearch.com
ashu@goXDML.com
Executive Summary/Abstract:

Both, the government regulators and the investors in global capital markets rely heavily on reported financial data. This financial data enters the information supply chain as raw data in a variety of formats and reported by the companies as quarterly and annual filing data. Similarly, stock exchanges provide high volume, real-time market data, including price and trading volumes.

Error-free, high quality data from these sources is critical for compliance and investment management decisions.

At the same time, an intense focus on outlier data is of great interest to both of these groups as it may represent a potential compliance issue or a potential arbitrage opportunity for investment decision, known as an ‘Alpha’ opportunity.
Executive Summary/Abstract: (contd.)

Currently, this screening for errors, for most part, is a manual process, and given the vast amount of data, prone to human errors. Extremely rare, but nonetheless real, so called fat-finger errors continue to cast a large shadow of doubt on overall data integrity and undermine trust in the system.

This score, now metadata, is linked with other contextual and semantic metadata. Aggregating cell level quality score then enables visual display of potential outliers as hot spots on a heat map. A drill-down using linked data increases transparency in data, all the way to the original source enabling trust in the overall data set.
Objectives:

- Contribute to the research on Information Quality in the area of capital markets financial data
- Describe challenges and opportunities
- Describe a practice-oriented solution using semantic tagging of an algorithmically assigned Data Quality Score as linked metadata
- Develop and present insights for potential use cases for enabling Transparency and Trust in financial data
Financial Data Supply Chain

Filers (Companies)

Raw Data (Excel, Text) → XBRL Files 1

Raw Data (Excel, Text) → XBRL Files 2

Raw Data (Excel, Text) → XBRL Files n

Data Aggregation & Distribution

Data (files) Aggregation

Data Extraction/Normalization

SQL Database

Non-XBRL Data

Market Data, Transactions Data
Fixed Income Data
Analysts Forecast & Historical Data
Currency Data

Investment Analytics

Comps
Screens
Valuations
Portfolio Construction
Risk Models
Financial Data is a Multi Dimensional Hypercube

Dimensions:
- Company
- Period
- Metric
- Currency
- As-Reported
- Forecasted
- Scale
- Audited

Financial Information Set

Financial Report

Information Grouping

Measure Relationships

Fact

Measure

Value (of Fact)

Courtesy: Charlie Hoffman
For more info on XBRL please refer to:
http://xdbl.squarespace.com/xbdl-for-dummies/
Data Quality - Challenges
Challenge 1: Data has little meaning without **Context**

Consider the spreadsheet (here). It takes you only a moment to work out that the item surrounded by the blue square is 16,000 in Tangible Assets for the 2002/2003 financial year. You worked that out by synthesizing all of the context that surrounds that blue square.

Now think about the way a computer might digest the contents of the blue square. At best, the computer will know that the Number 16,000 appears at cell reference C8. At worst, it will just know "16,000".

XBRL allows systems to communicate the **entire context that a human needs** to fully understand a concept.

*Source: [http://www.kpmg.com/xbrl/context.asp](http://www.kpmg.com/xbrl/context.asp)*
Challenge 2: Raw Data Quality Forces Manual Scrubbing

Acquire raw data, build models, develop forecasts, Manually Tag and upload financial Models to proprietary databases

Search many databases, co-mingle and massage different data formats and then develop Comps and Screens under severe time pressures

Develop and distribute analytics and Alpha insights to Buy-side clients, Traders and Investment Bankers under business critical time pressures

“biggest frustration is the turnaround time . . .”
Analyst at a top Wall Street firm
Challenge 3: Diversity of Semantics

Strategic View

Data Providers/Sell-side Analysts
- Data Aggregators
- Valuation/Forecast Models
- Ratings
- Stock Recommendations

Buy-side Analysts
- Consensus Estimates
- Screeners
- Comps
- Trading Strategies
- Back-Test
- Portfolio Construction
- Risk Management

Filers/Companies
- SEC Filings (10-K/10Q)
- Earnings / Call Reports
- Annual Reports

Long Tail View

Long Tail of Data Organization

Wiki-Tagged Data Organization Process

- World of Accounting (GAAP)
  - 50 to 100 common Financial Metrics for
    - Income Statement
    - Balance Sheet
    - Cash-flow

- World of Analysts (Industry Specific Taxonomies)
  - Additional 2,000 + unique Financial Metrics for
    - Financial Modeling
    - Valuation
    - Forecasts,
    - Trading Strategies

- World of Investors (Folksonomy)
  - Additional Unlimited, unstructured ‘localized’ synonyms for many Financial Metrics for
    - compliance with country-specific regulators
    - proprietary financial modeling frameworks
    - lack of any appropriate metrics in GAAP or readily available taxonomies, resulting in – Line 6 Items on 10-K, MDAs, and Footnotes
Challenge 4: Quantity & Quality of Data (Cost & Time Issue)

Availability of Data for Institutional Investors

Wall Street Data Types | Quantity | Quality
--- | --- | ---
Reference Data | ★★★★★ | ★★★★★
Company Fundamental Data | ★★★★★ | ★★★★★
Historical Data | ★★★★★ | ★★★★★
Sell-side Research Data | ★★★★★ | ★★★★★
Hold-bin (unpublished) Data | ★★★★★ | ★★★★★
Analyst Forecasted Data | ★★★★★ | ★★★★★
As Reported Data | ★★★★★ | ★★★★★
Audited 10-K Data | ★★★★★ | ★★★★★
UnAudited 10-Q Data | ★★★★★ | ★★★★★
Company Revised Data | ★★★★★ | ★★★★★
Sector Data | ★★★★★ | ★★★★★
Industry Data | ★★★★★ | ★★★★★
Economic Data | ★★★★★ | ★★★★★
Analyst Adjusted Data | ★★★★★ | ★★★★★
Consensus Data | ★★★★★ | ★★★★★
Cross-Asset Class Data | ★★★★★ | ★★★★★
Fixed Income Data | ★★★★★ | ★★★★★
Market/Trading Data | ★★★★★ | ★★★★★

70,000 + Listed Companies on 40+ Exchanges Globally
Challenge 5: Distinguishing Data Error from Alpha Opportunity?
Data Quality - Opportunity

XBRL
XBRL basics

- XBRL is a way of expressing semantics for financial information.

- It allows companies to use nationally and internationally common 'tags' to identify individual reporting concepts that exist in a corporate report. Information that is coded in this way can be instantly and accurately exchanged between systems.

- XBRL allows context to be communicated along with content.

- XBRL is in production use today by FDIC, SEC and regulators around the world.

- XBRL and RDF/OWL are complementary.
Example of how XML, the basic extensible markup language, tags information for a book publisher. Data on the left are the tags for such information as the order number, the name of the person buying the book, the price and the author. Data on the right are the same data but represented in database fields.

Tagging Data using XBRL

A small example of XBRL

This is a small example of XBRL -- intended for reading by computers, not humans. To see what it represents, click here.

<ifrs-gp:AssetsHeldSale contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:AssetsHeldSale>
<ifrs-gp:ConstructionProgressCurrent contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:ConstructionProgressCurrent>
<ifrs-gp:Inventories contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:Inventories>
<ifrs-gp:OtherFinancialAssetsCurrent contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:OtherFinancialAssetsCurrent>
<ifrs-gp:HedgingInstrumentsCurrentAsset contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:HedgingInstrumentsCurrentAsset>
<ifrs-gp:CurrentTaxReceivables contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:CurrentTaxReceivables>
<ifrs-gp:TradeOtherReceivablesNetCurrent contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:TradeOtherReceivablesNetCurrent>
<ifrs-gp:PrepaymentsCurrent contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:PrepaymentsCurrent>
<ifrs-gp:CashCashEquivalents contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:CashCashEquivalents>
<ifrs-gp:OtherAssetsCurrent contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">100000</ifrs-gp:OtherAssetsCurrent>
<ifrs-gp:AssetsCurrentTotal contextRef="Current_AsOf" unitRef="U-Euros" decimals="0">1000000</ifrs-gp:AssetsCurrentTotal>
XBRL Constraint: Document Centricity

Before
(HTML, PDF)

After
(XBRL ZIP)

A set of XML files
- Instance Document
- Taxonomy Schema
- Linkbases
A Practice-oriented Solution using Semantic Tagging of an Algorithmically Assigned Data Quality Score
Author Mode

Member Authentication
Your user ID and password
Help for Login

Username: admin, Password: admin
Domain: goodml

Running ResearchPoint XBRL for the first time?
You must have an active internet connection to authenticate your account information on ResearchPoint-XBRL server.

Login

15th International Conference on Information Quality, 2010

ResearchPoint
XBRL and Semantic Web (RDF) integrated
Windows Desktop Platform

(1) Authoring, Reviewing, Editing Mode

- Select Excel or XBRL file
- Select Taxonomy
- Tag, Review, Comment, Edit,
- Analyze, Chart, Share,
- Save as XBRL-Tagged-RDF in SQL Database
15th International Conference on Information Quality, 2010

Web User Mode

ResearchPoint

XBRL and Semantic Web (RDF) integrated

(2) Web Rendered/User/Analytics Mode

XBRL Rendering
Interactive Charts View

Tabular View

FusionCharts Evaluation - An InfoSoft Global Creation

USD1K

USD600

USD300

USD100

USD0

USD100

USD300

USD500

USD800

USD1K

Purchased in - Amortization of intangibles
Restructuring charges
Equity in earnings of nonconsolidated affiliates
Sundry income - net

Select metrics (maximum up to 6 metrics)
Proof of Concept: Data Quality Heat Map Visualization
Data Quality Rating Implementation

Think: FICO Score for Data
Insights & Potential Use Cases For Enabling Transparency and Trust In Financial Data
Semantic Web & XBRL Architectures

15th International Conference on Information Quality, 2010
Use Case: Outlier Analytics – Single Company, Single Metric, Multi-Year
Use Case: Outlier Analytics – Single Company, Multi-Metric, Multi-Year
Use Case: Forensic Analytics – Alpha Insights
References