15th International Conference on Information Quality, 2010

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.

A data error or poor quality data is also more than likely to appear as a perfectly valid data but an outlier.

Distinguishing between a bad data or valid outlier data requires further scrutiny, transparency and analysis.

Executive Summary/Abstract: (contd.)

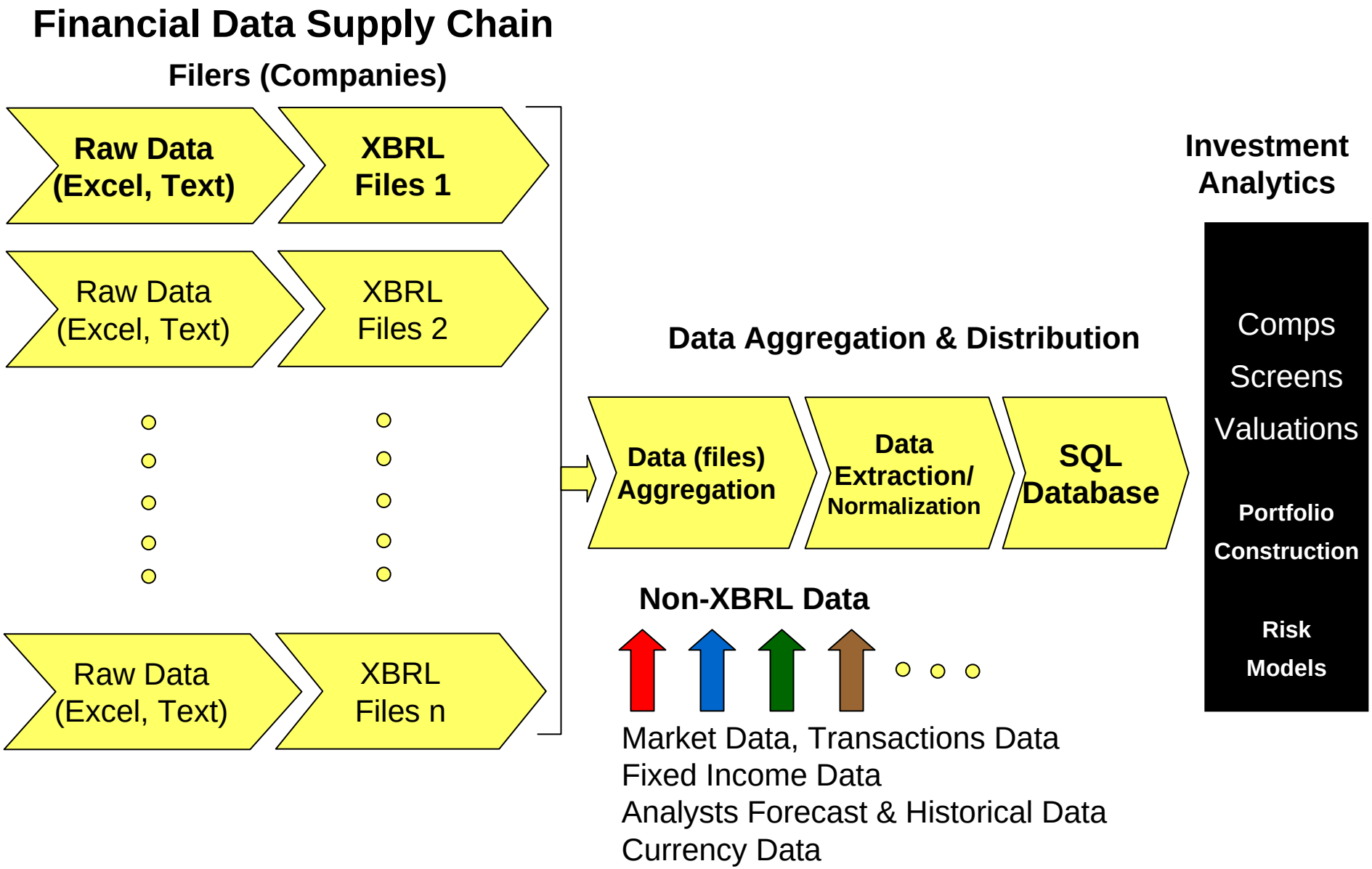
Currently, this screening for errors, for most part, is a manual process, and given the vast amount of data, prone to human errors. Extremely rare, but nonetheless real, so called fat-finger errors continue to cast a large shadow of doubt on overall data integrity and undermine trust in the system.

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.

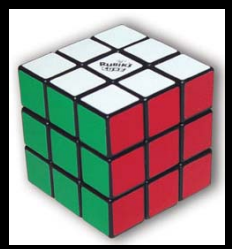
This paper presents a case study, where we applied business logic driven data quality rating system, and automatically tag a semantic data quality rank or score to each data element.

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

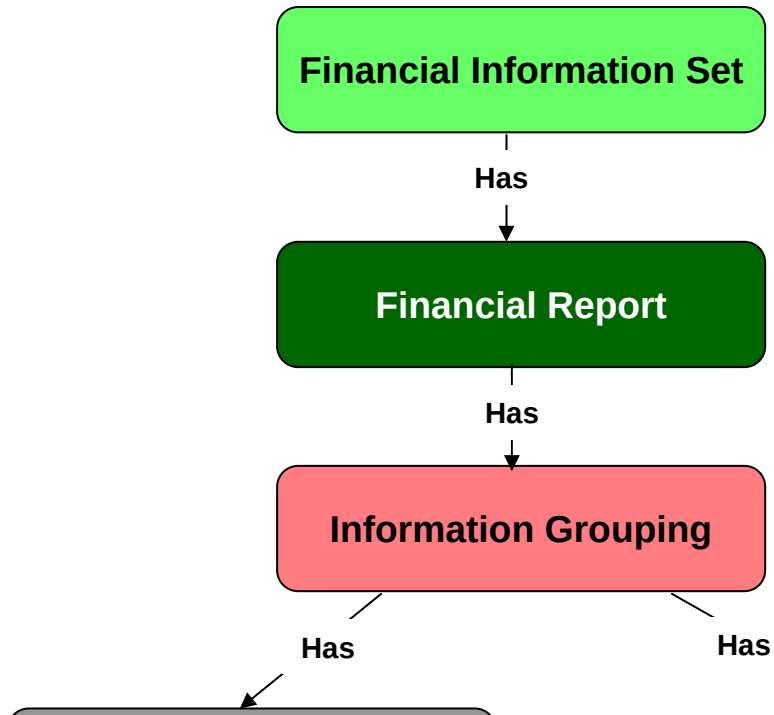
Financial Data is a Multi Dimensional Hypercube



Dimensions:

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

. . .



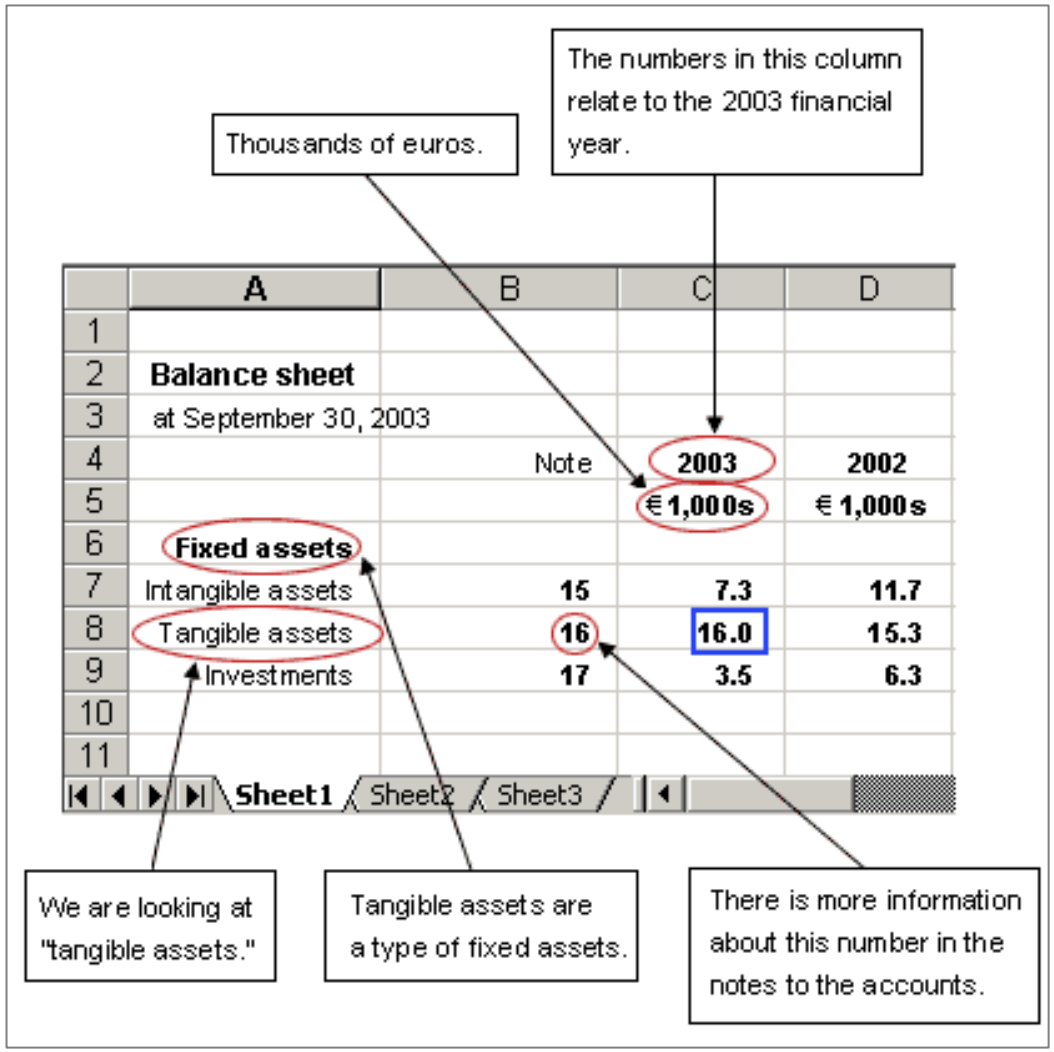
Courtesy: Charlie Hoffman
For more info on XBRL please refer

http://xbrl.squarespace.com/xbrl-for-dummies/

Data Quality - Challenges

Challenge 1: Data has little meaning without Context

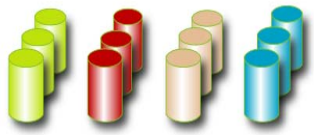
Consider the spreadsheet (here). It takes you only a moment to work out that the item surrounded by the blue square is 16,000 in Tangible Assets for the 2002/2003 financial year. You worked that out by synthesizing all of the context that surrounds that blue square. Now think about the way a computer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

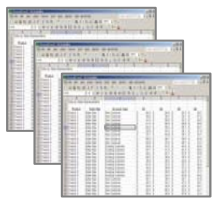


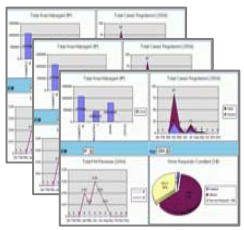
Source: http://www.kpmg.com/xbrl/context.asp

Challenge 2: Raw Data Quality Forces Manual Scrubbing

Research Analysts Work-flow







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

Manual Process

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

Manual Process

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

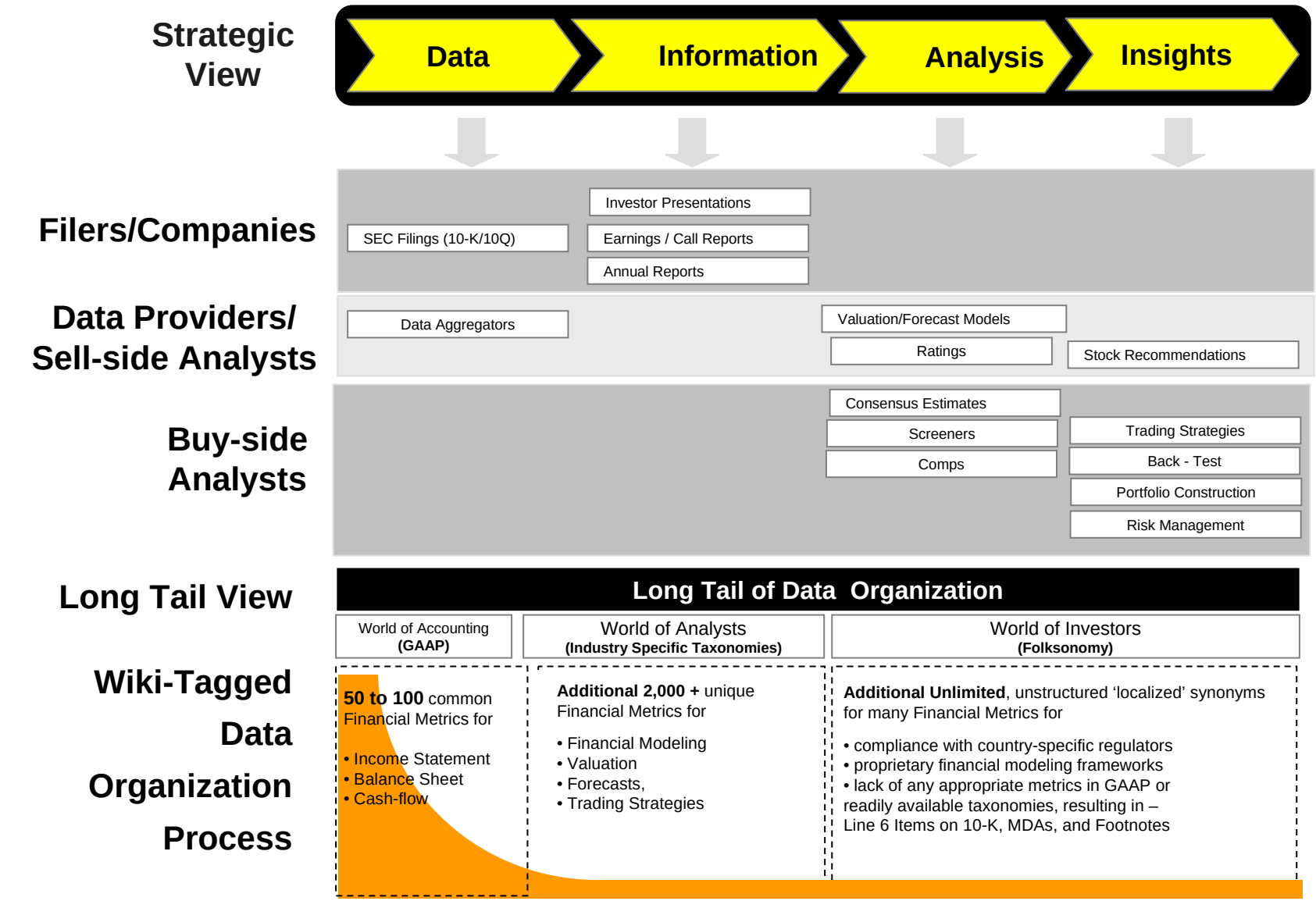
Manual Process

"biggest frustration is the turnaround time . . ."

Analyst at a top Wall Street firm

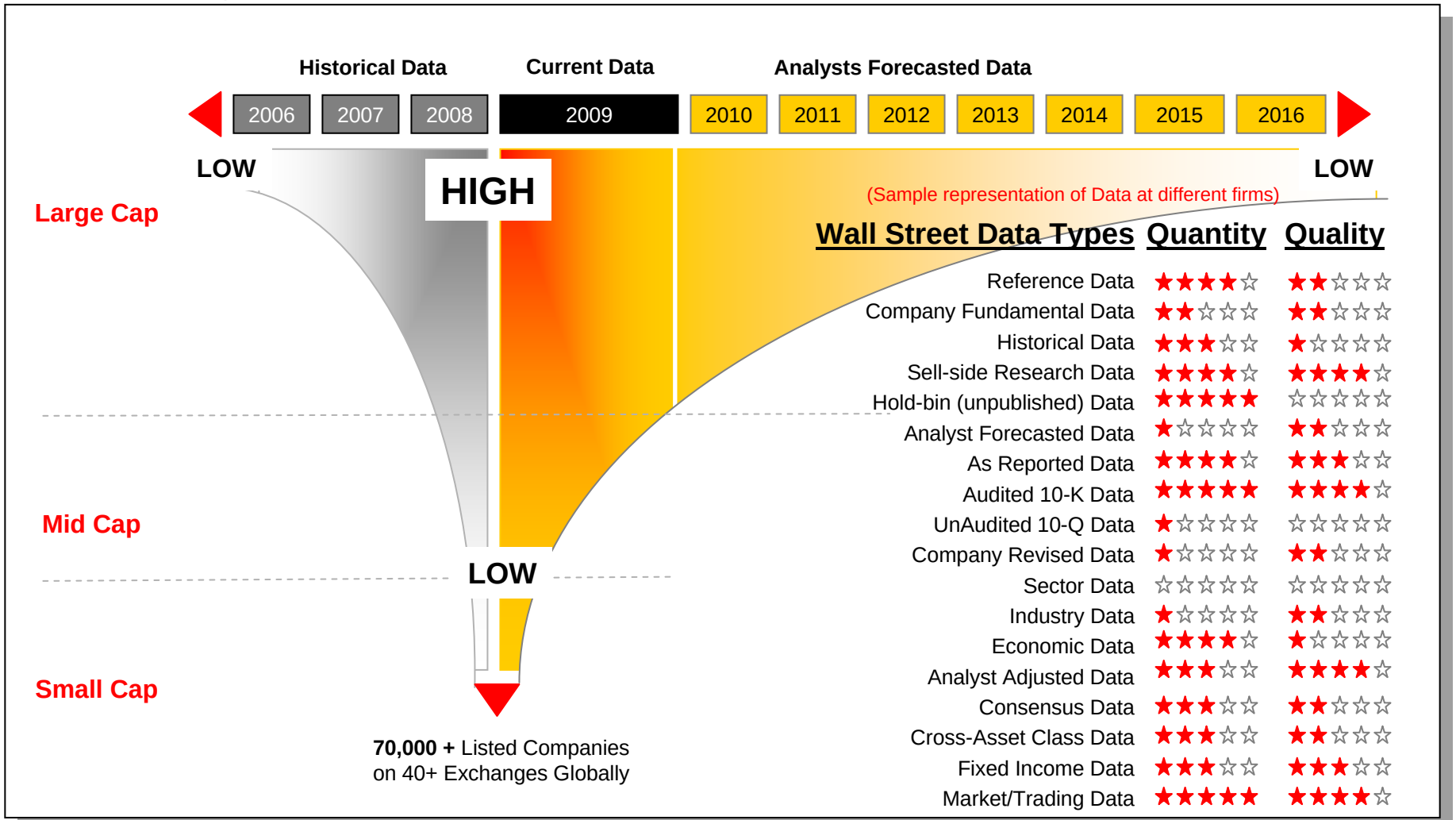
Help!

Challenge 3: Diversity of Semantics

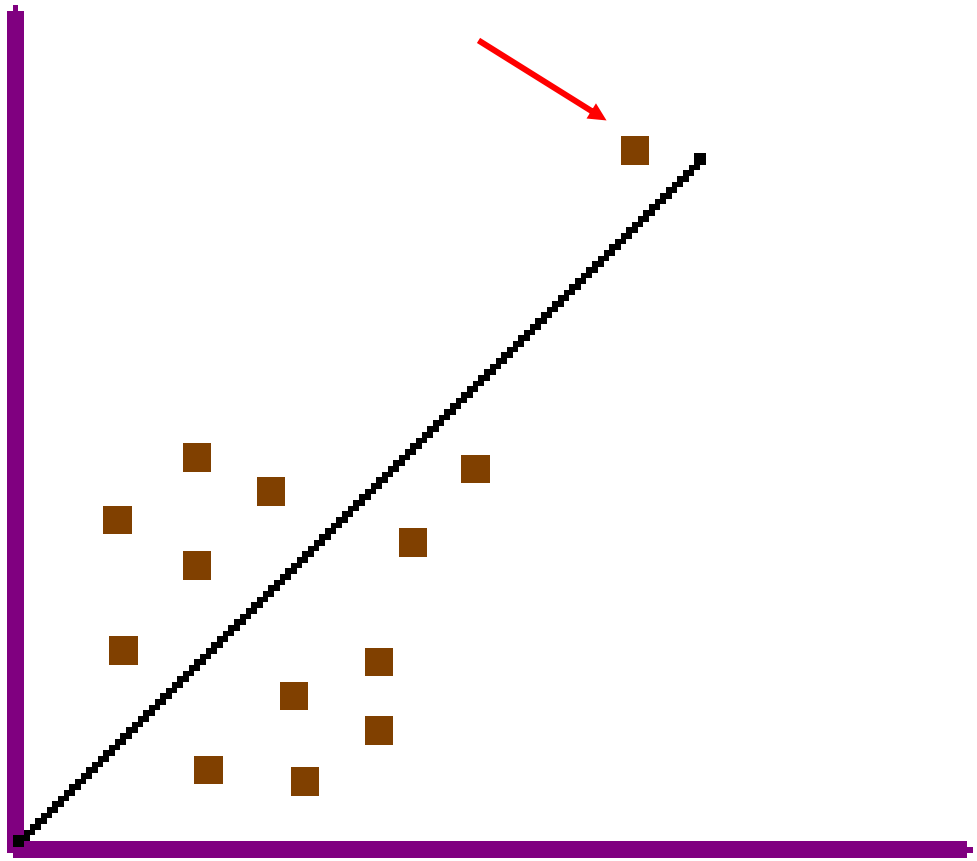


Challenge 4: Quantity & Quality of Data (Cost & Time Issue)

Availability of Data for Institutional Investors



Challenge 5: Distinguishing Data Error from Alpha Opportunity?



Data Quality - Opportunity

XBRL

XBRL basics

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

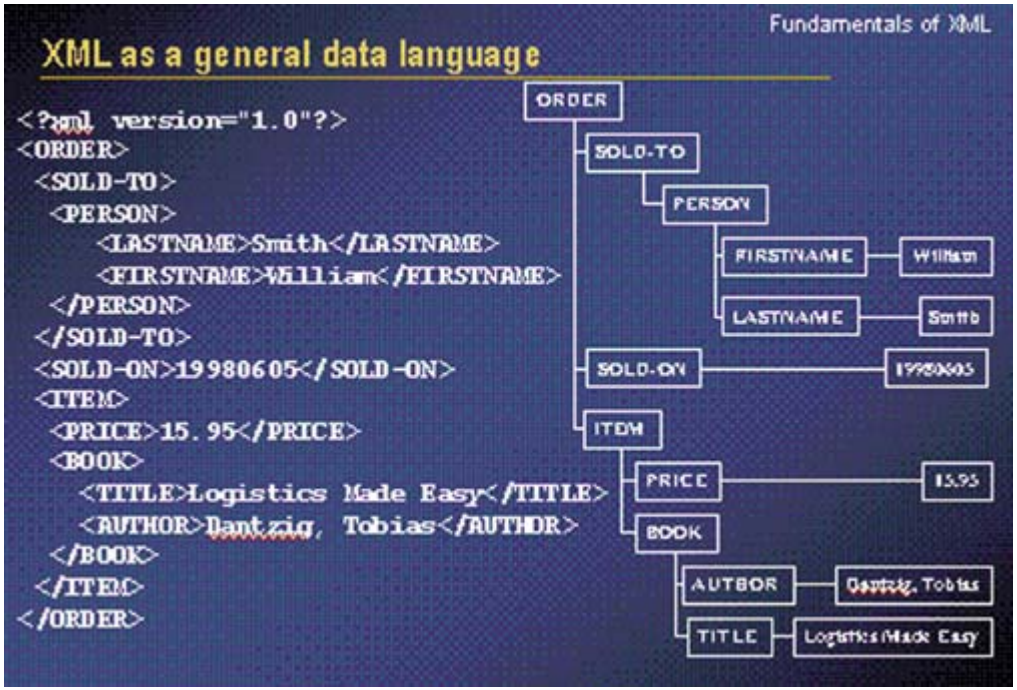
- It allows companies to use nationally and internationally common 'tags' to identify individual reporting concepts that exist in a corporate report. Information that is coded in this way can be instantly and accurately exchanged between systems.

- 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

Tagging of Data Using XML



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.

Source: http://www.aicpa.org/pubs/jofa/aug2000/zaro_ex3.htm

Microsoft Excel - XML as aGeneral data language

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U

H18

	A	B	C	D	E	F	G	H	I	J	K
1	XML Version 1.0										
2			Book			Customer					
3	Book Orders	Sale Date	Title	Author	Price	First Name	Last name				
4		5-Jun-98	Logistics Made Easy	Dantzig, Tobias	15.95	William	Smith				

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

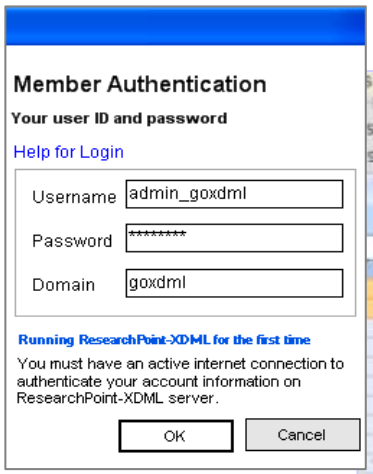
15th International Conference on Information Quality, 2010

A Practice-oriented Solution using Semantic Tagging of an Algorithmically Assigned Data Quality Score

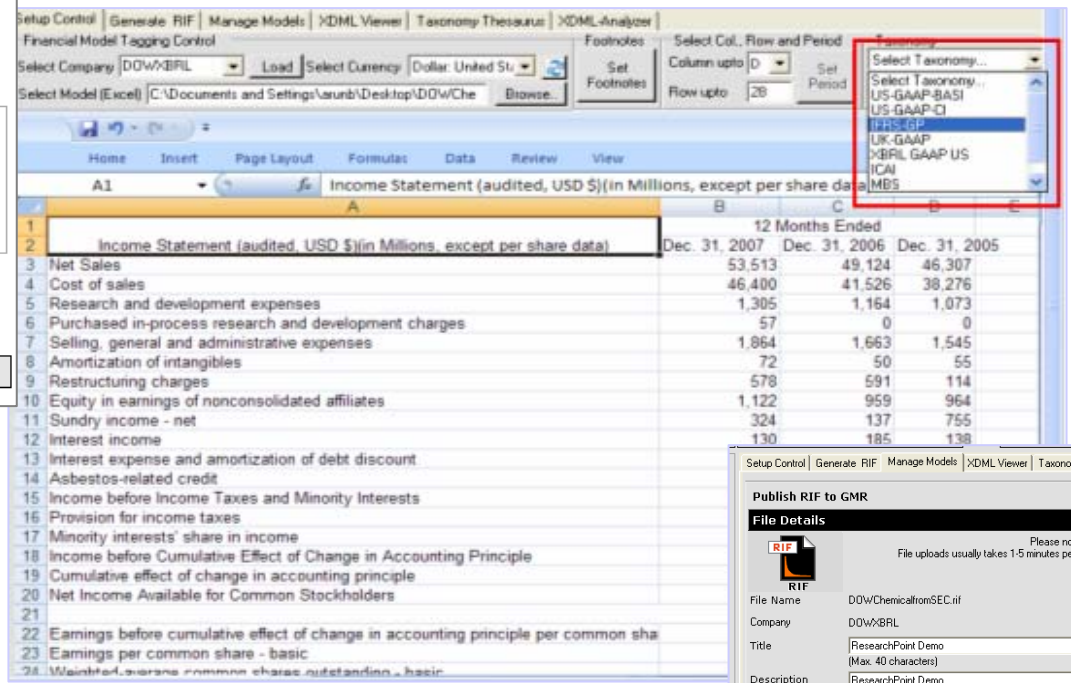
ResearchPoint
XBRL and Semantic Web (RDF) integrated
Windows Desktop Platform

Author Mode

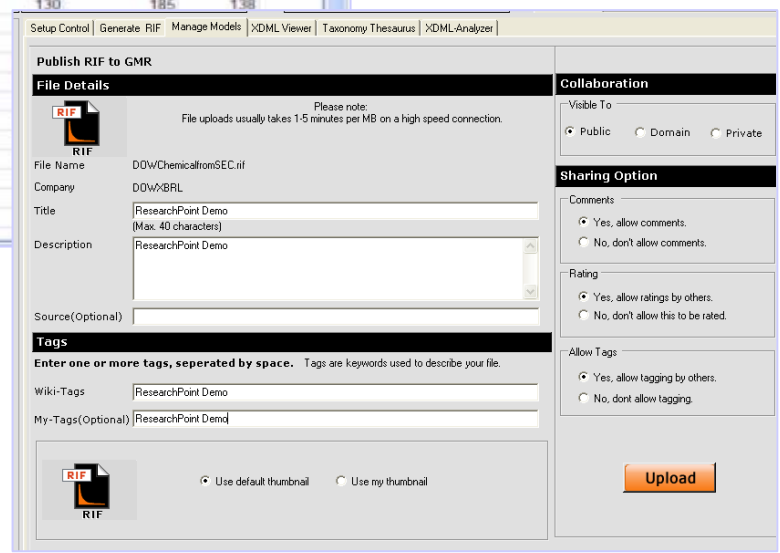
(1) Authoring, Reviewing, Editing Mode



Login



Upload to Web



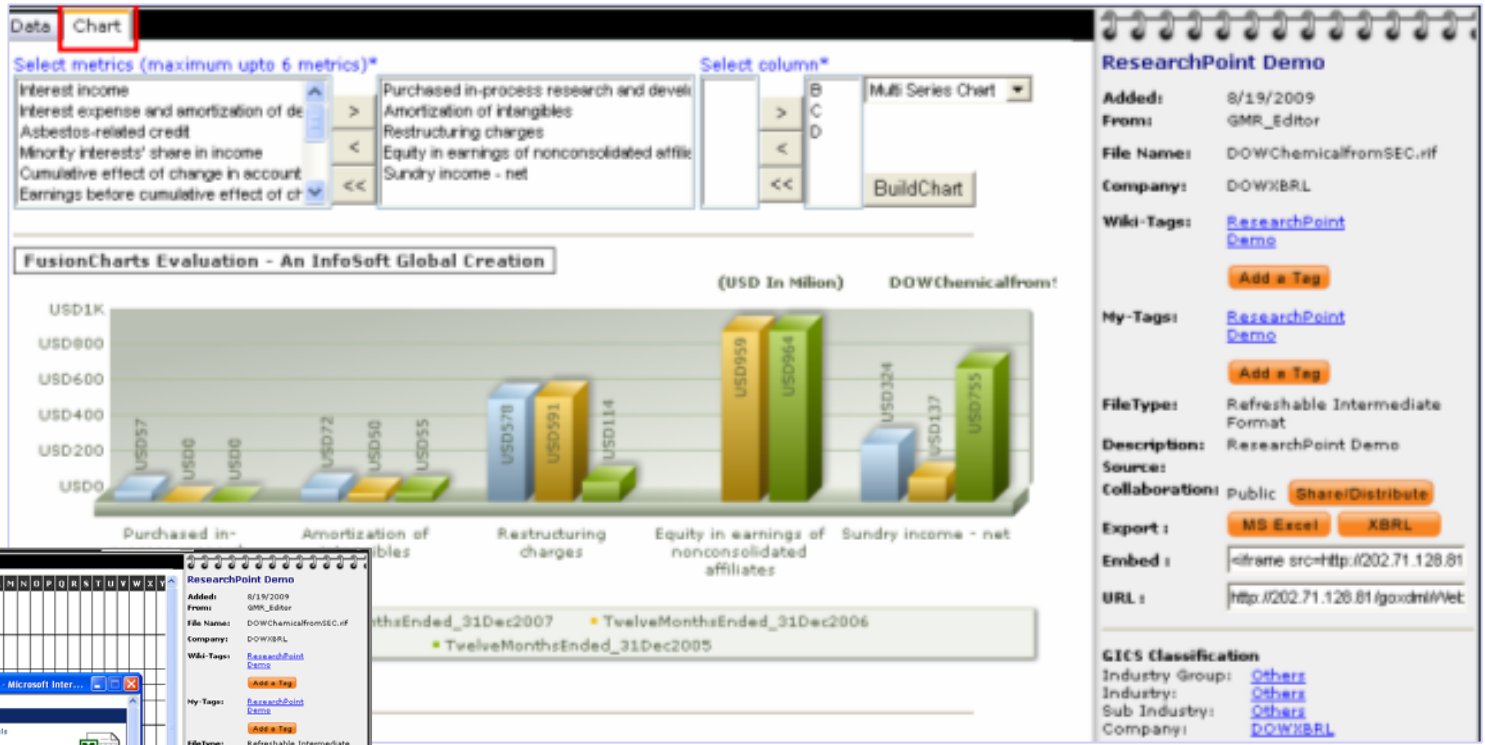
- **Select Excel or XBRL file**
- **Select Taxonomy**
- **Tag, Review, Comment, Edit,**
- **Analyze, Chart, Share,**
- **Save as XBRL-Tagged-RDF in SQL Database**

Web User Mode

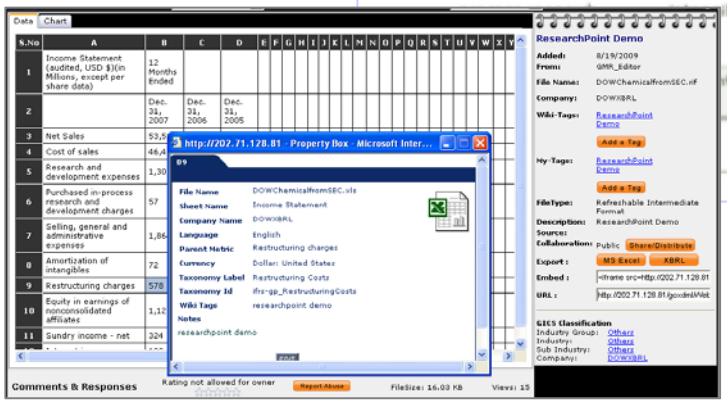
ResearchPoint
XBRL and Semantic Web (RDF) integrated

(2) Web Rendered/User/Analytics Mode

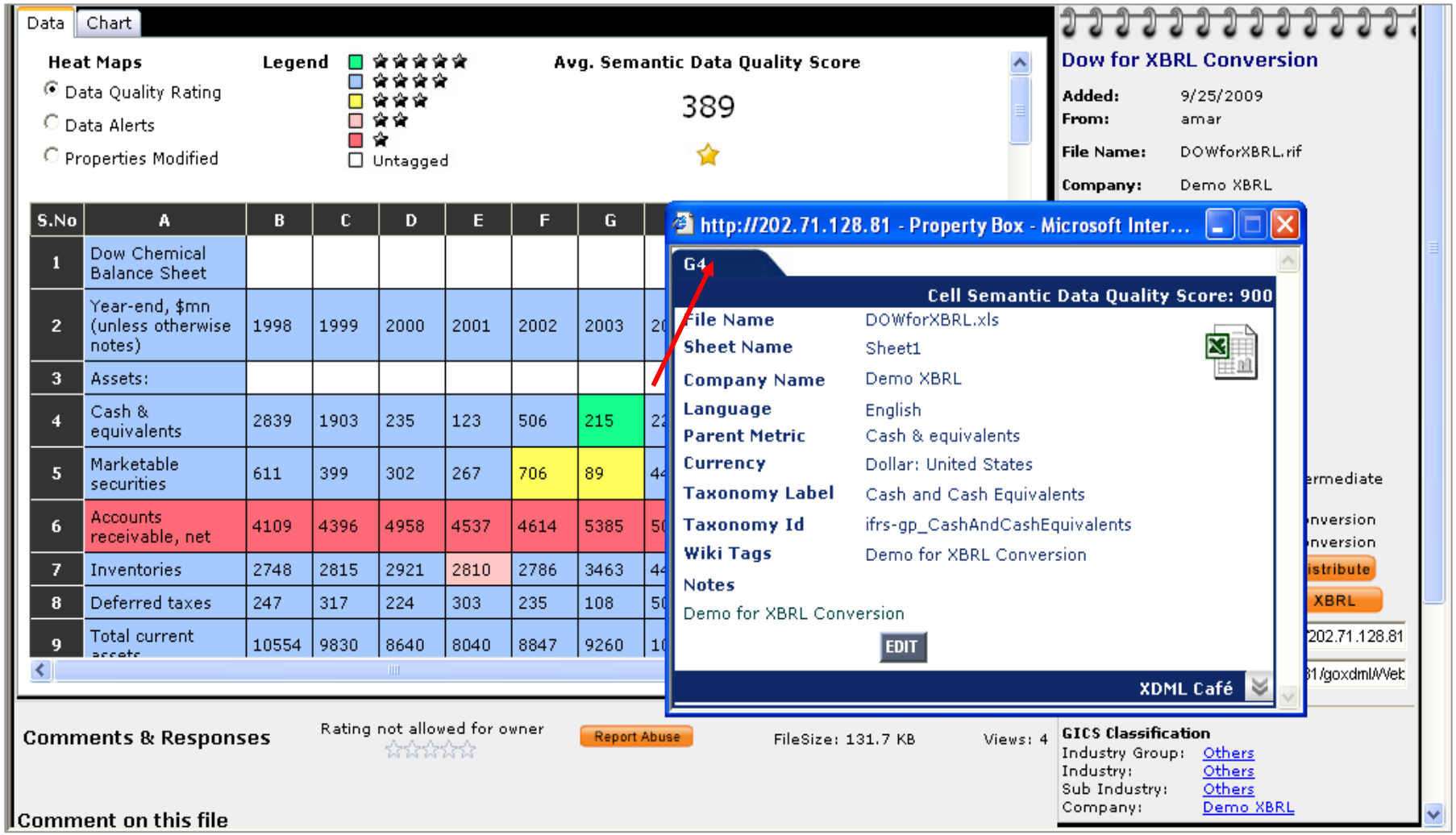
XBRL Rendering Interactive Charts View



XBRL Rendering Tabular View

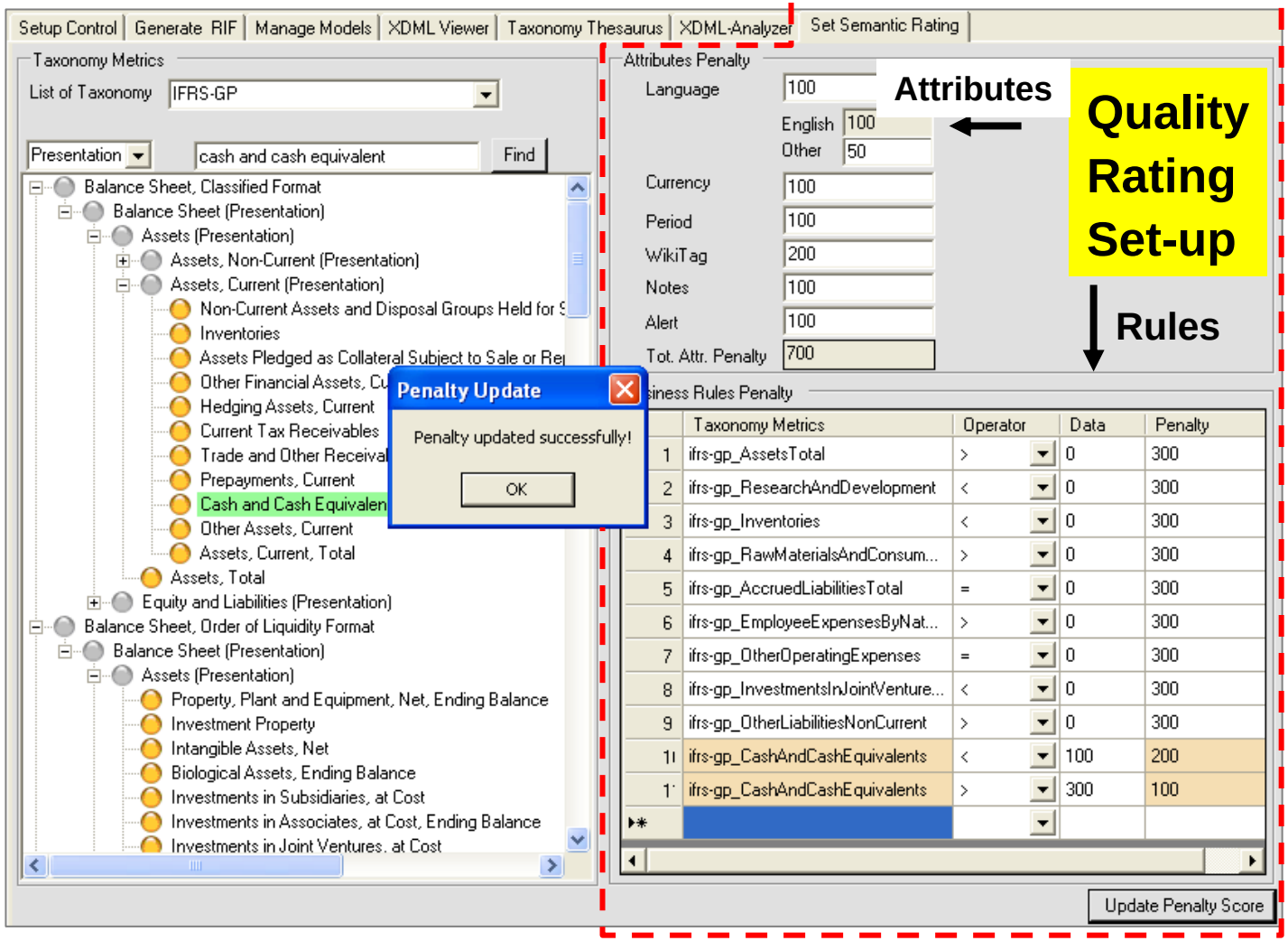


Proof of Concept: Data Quality Heat Map Visualization



Data Quality Rating Implementation

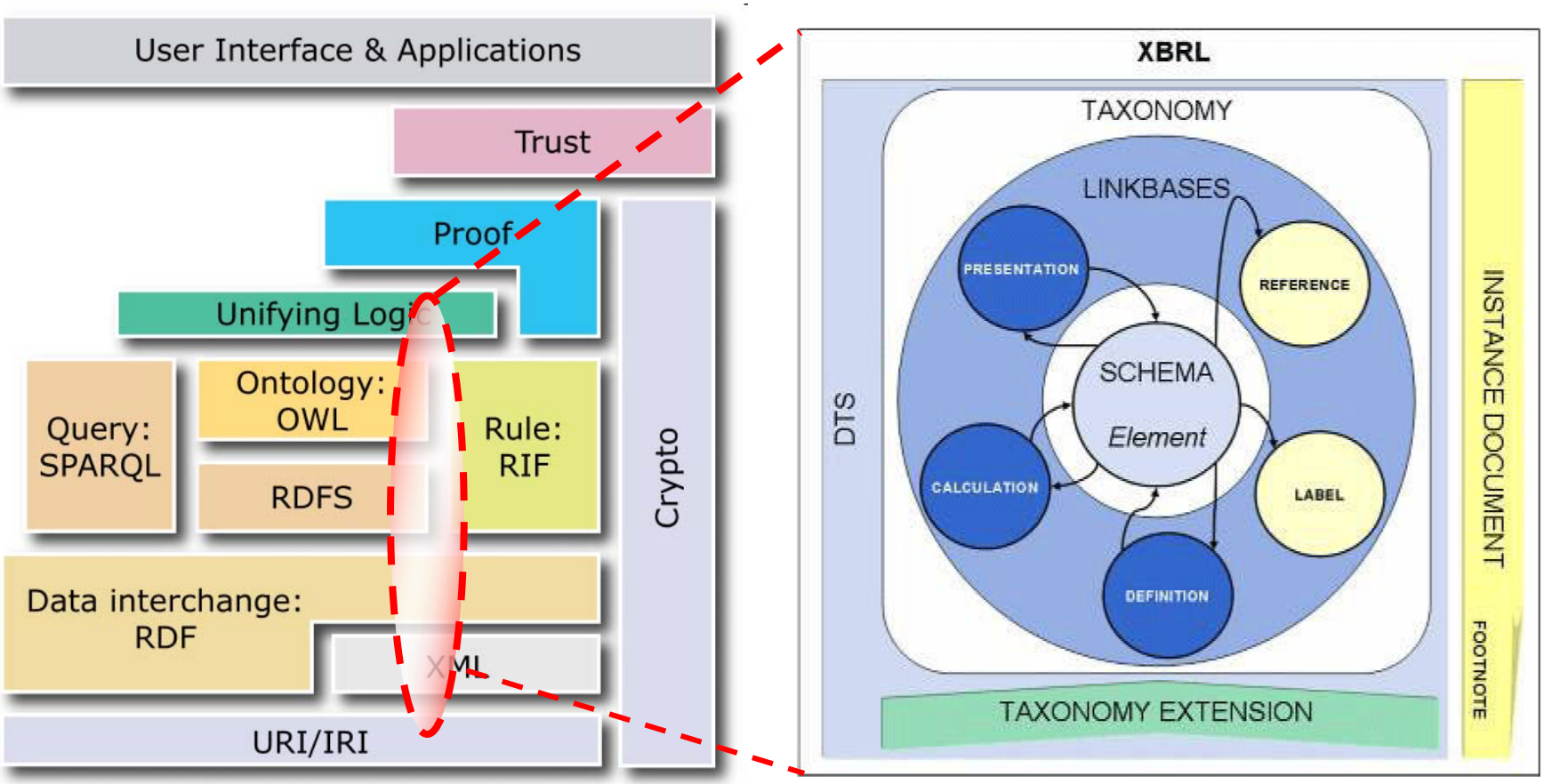
Standard Taxonomy →



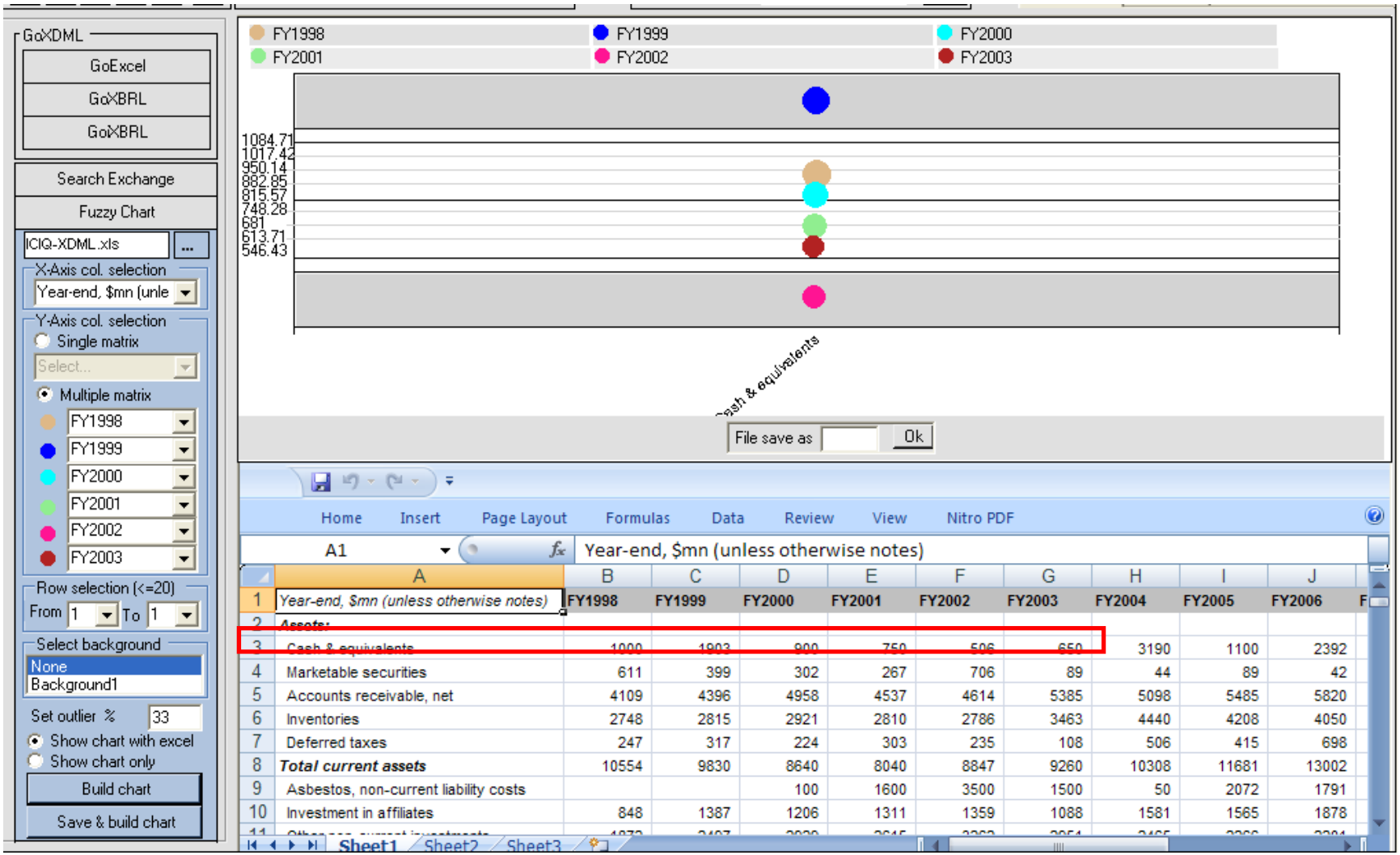
Think: FICO Score for Data

Insights & Potential Use Cases For Enabling Transparency and Trust In Financial Data

Semantic Web & XBRL Architectures

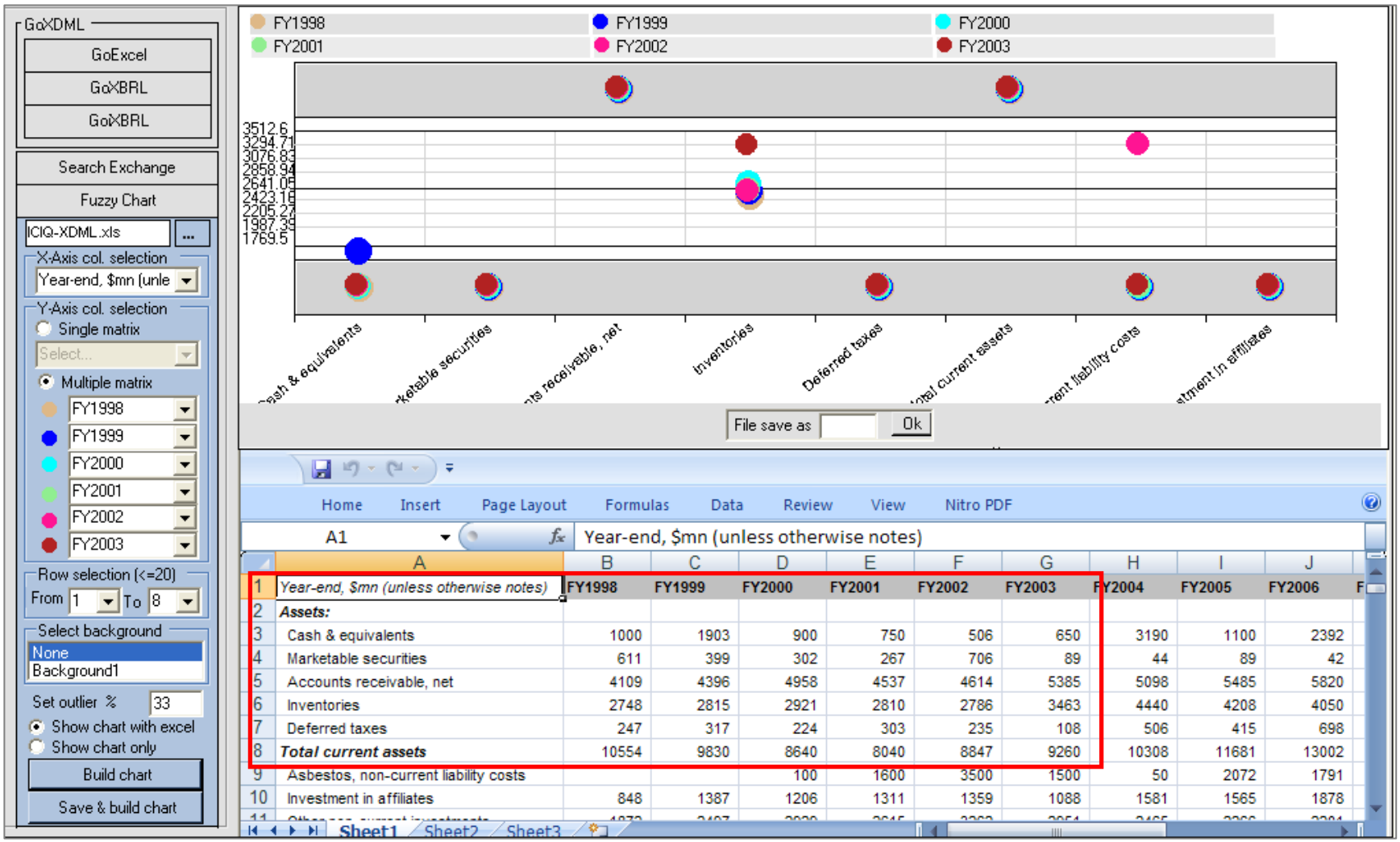


Use Case: Outlier Analytics – Single Company, Single Metric, Multi-Year

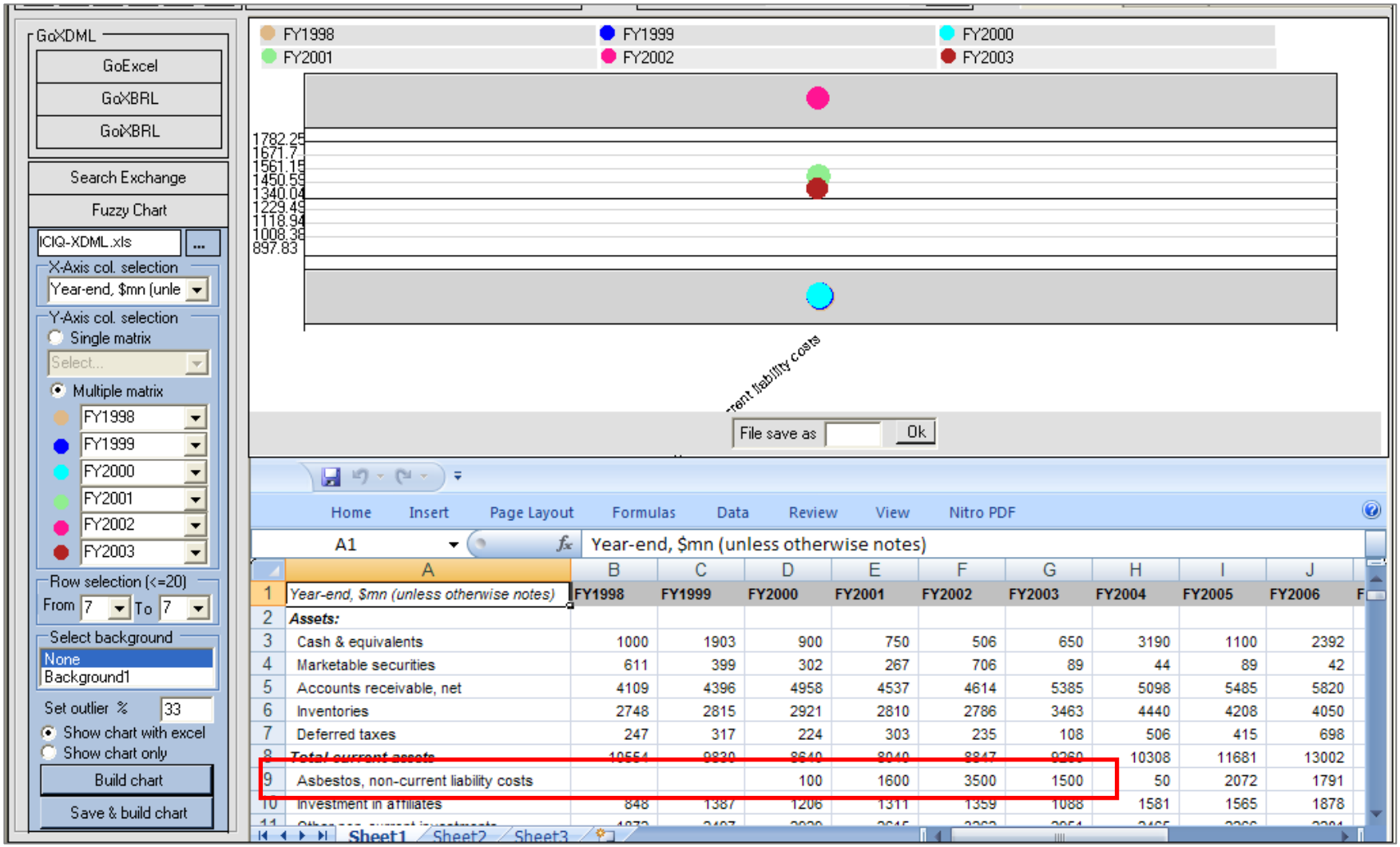


	A	B	C	D	E	F	G	H	I	J
1	Year-end, $mn (unless otherwise notes)	FY1998	FY1999	FY2000	FY2001	FY2002	FY2003	FY2004	FY2005	FY2006
2	***Assets:***									
3	Cash & equivalents	1000	1903	900	750	506	650	3190	1100	2392
4	Marketable securities	611	399	302	267	706	89	44	89	42
5	Accounts receivable, net	4109	4396	4958	4537	4614	5385	5098	5485	5820
6	Inventories	2748	2815	2921	2810	2786	3463	4440	4208	4050
7	Deferred taxes	247	317	224	303	235	108	506	415	698
8	***Total current assets***	10554	9830	8640	8040	8847	9260	10308	11681	13002
9	Asbestos, non-current liability costs			100	1600	3500	1500	50	2072	1791
10	Investment in affiliates	848	1387	1206	1311	1359	1088	1581	1565	1878

Use Case: Outlier Analytics – Single Company, Multi-Metric, Multi-Year



Use Case: Forensic Analytics – Alpha Insights



Year-end, $mn (unless otherwise notes)	FY1998	FY1999	FY2000	FY2001	FY2002	FY2003	FY2004	FY2005	FY2006
Assets:									
Cash & equivalents	1000	1903	900	750	506	650	3190	1100	2392
Marketable securities	611	399	302	267	706	89	44	89	42
Accounts receivable, net	4109	4396	4958	4537	4614	5385	5098	5485	5820
Inventories	2748	2815	2921	2810	2786	3463	4440	4208	4050
Deferred taxes	247	317	224	303	235	108	506	415	698
Total current assets	10554	9830	8640	8040	8847	9260	10308	11681	13002
Asbestos, non-current liability costs			100	1600	3500	1500	50	2072	1791
Investment in affiliates	848	1387	1206	1311	1359	1088	1581	1565	1878

References

- **[1] Bhatnagar, A. XBRL Taxonomy Extension Comparability Issues And Potential Semantic Web Solutions. *Workshop on Improving Access - Financial Data on the Web*, October 2009.**