

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.

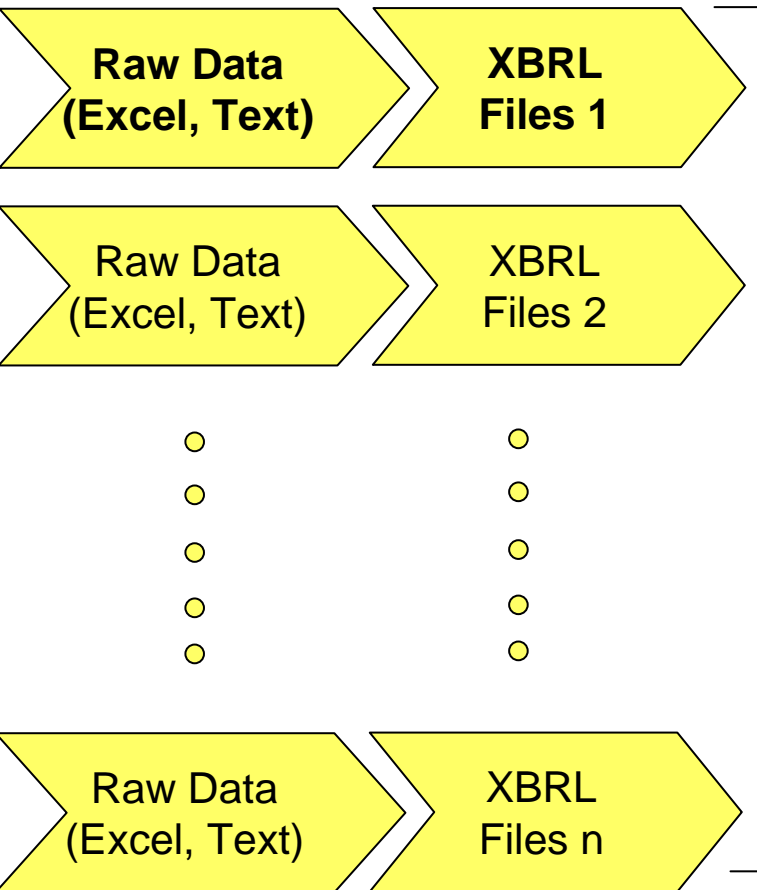
This paper presents a case study, where we applied business logic driven data quality rating system, and automatically tag a semantic data quality rank or score to each data element.

Objectives:

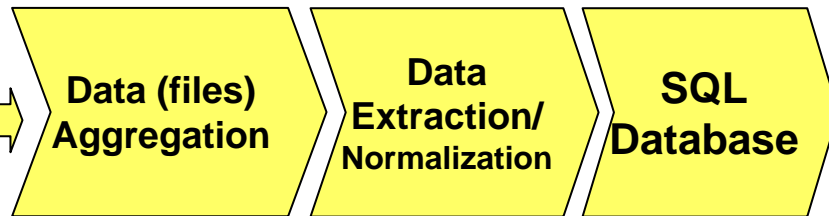
- ❖ **Contribute to the research on Information Quality in the area of capital markets financial data**
- ❖ **Describe challenges and opportunities**
- ❖ **Describe a practice-oriented solution using semantic tagging of an algorithmically assigned Data Quality Score as linked metadata**
- ❖ **Develop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

Financial Data Supply Chain

Filers (Companies)



Data Aggregation & Distribution

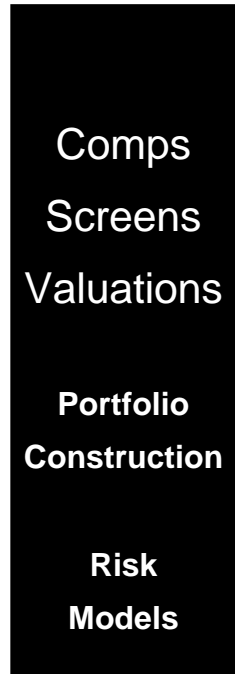


Non-XBRL Data



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

Investment Analytics

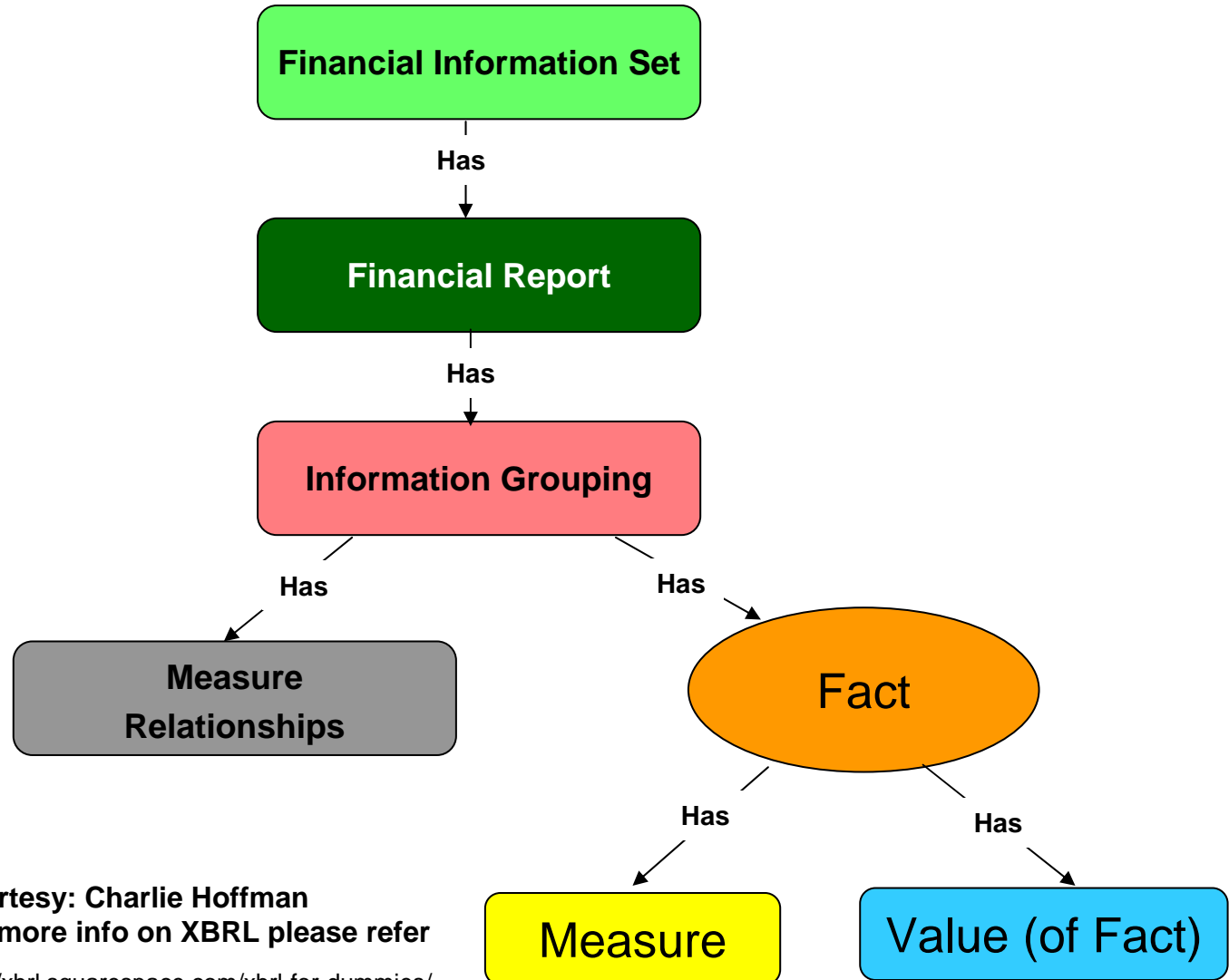


Financial Data is a Multi Dimensional Hypercube



Dimensions:

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



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.

	A	B	C	D
1				
2	Balance sheet			
3	at September 30, 2003			
4		Note	2003	2002
5			€ 1,000s	€ 1,000s
6	Fixed assets			
7	Intangible assets	15	7.3	11.7
8	Tangible assets	16	16.0	15.3
9	Investments	17	3.5	6.3
10				
11				

Thousands of euros.

The numbers in this column relate to the 2003 financial year.

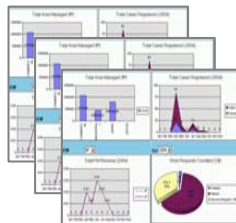
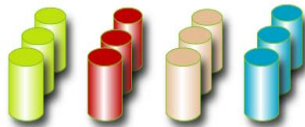
We are looking at "tangible assets."

Tangible assets are a type of fixed assets.

There is more information about this number in the notes to the accounts.

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

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**

Manual Process

Manual Process

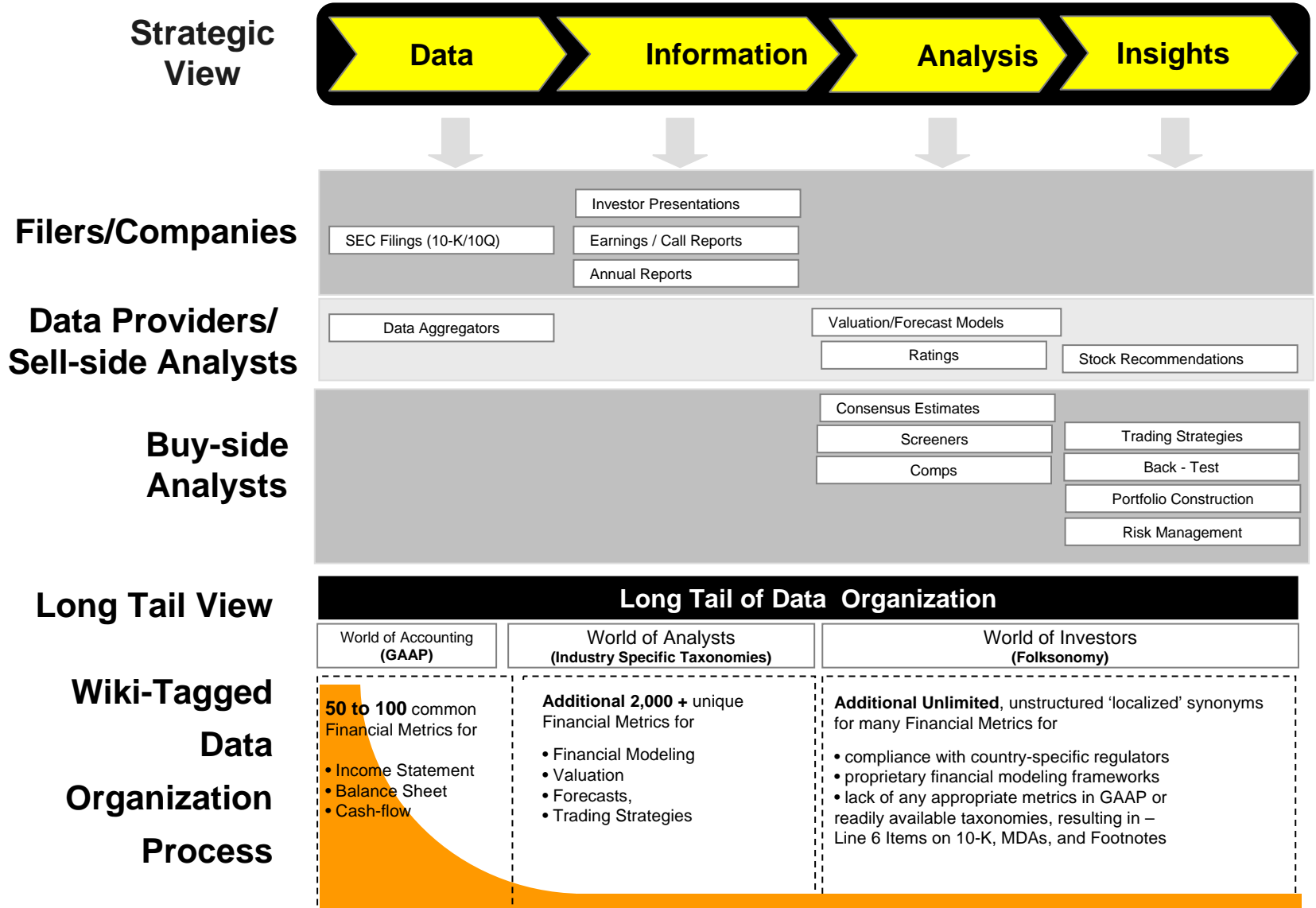
Manual Process

“biggest frustration is the turnaround time . . .”

Analyst at a top Wall Street firm

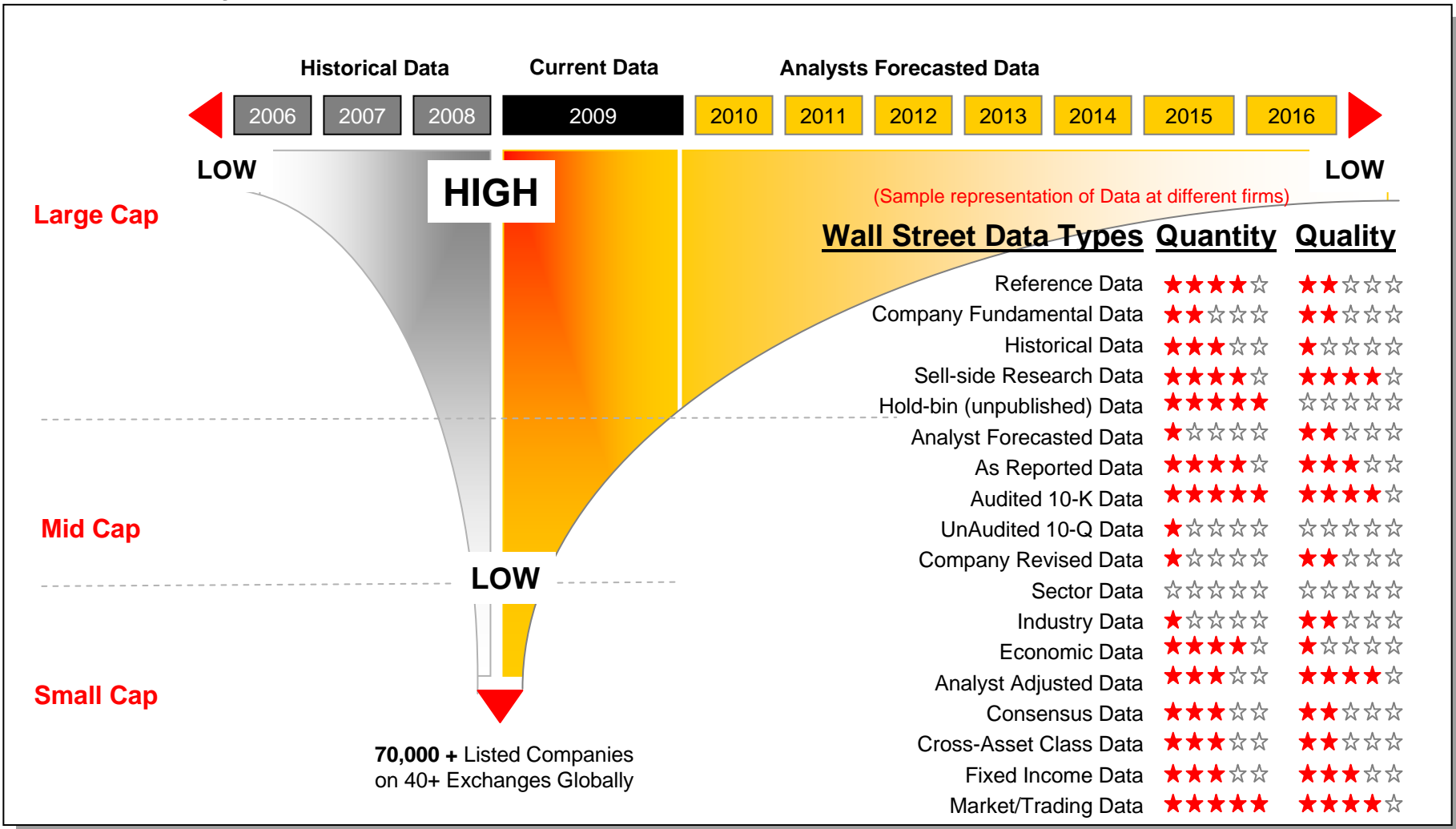
Help!

Challenge 3: Diversity of Semantics

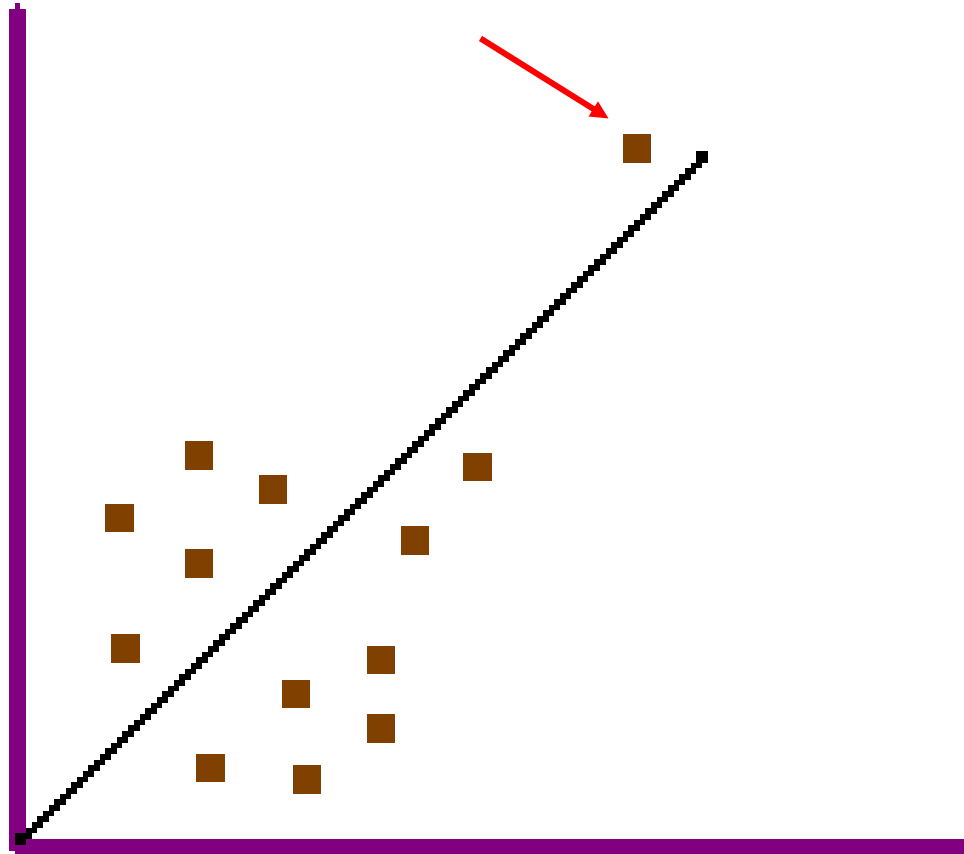


Challenge 4: Quantity & Quality of Data (Cost & Time Issue)

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

XML as a general data language

```
<?xml version="1.0"?>
<ORDER>
  <SOLD-TO>
    <PERSON>
      <LASTNAME>Smith</LASTNAME>
      <FIRSTNAME>William</FIRSTNAME>
    </PERSON>
  </SOLD-TO>
  <SOLD-ON>19980605</SOLD-ON>
  <ITEM>
    <PRICE>15.95</PRICE>
    <BOOK>
      <TITLE>Logistics Made Easy</TITLE>
      <AUTHOR>Dantzig, Tobias</AUTHOR>
    </BOOK>
  </ITEM>
</ORDER>
```

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

	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

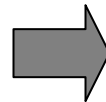
A small example of XBRL

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

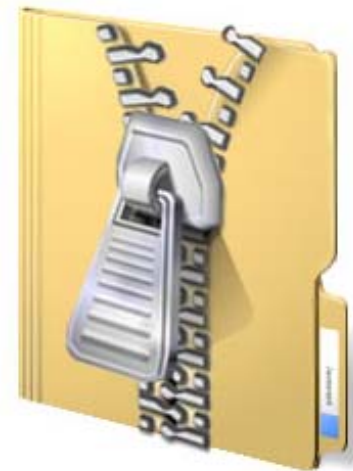
```
<ifrs-gp:AssetsHeldSale contextRef="Current_AsOf" unitRef="U-Euros"
  decimals="0">100000</ifrs-gp:AssetsHeldSale>
<ifrs-gp:ConstructionProgressCurrent contextRef="Current_AsOf"
  unitRef="U-Euros" decimals="0">100000</ifrs-
  gp:ConstructionProgressCurrent>
<ifrs-gp:Inventories contextRef="Current_AsOf" 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

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

Author Mode

(1) Authoring, Reviewing, Editing Mode

Member Authentication

Your user ID and password

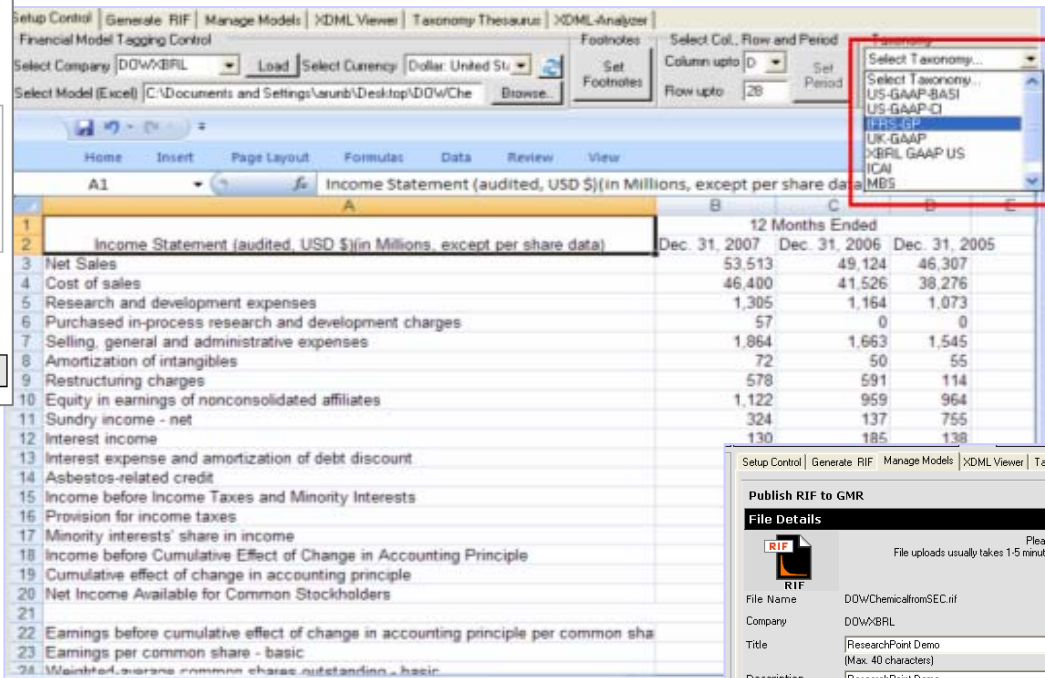
[Help for Login](#)

Username

Password

Domain

[Running ResearchPoint-XML for the first time](#)
You must have an active internet connection to authenticate your account information on ResearchPoint-XML server.




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

Publish RIF to GMR

File Details

 Please note: File uploads usually takes 1-5 minutes per MB on a high speed connection.


File Name: DOWChemicalromSEC.rif
Company: DOW/BRL
Title: ResearchPoint Demo (Max 40 characters)
Description: ResearchPoint Demo
Source(Optional):

Tags

Enter one or more tags, separated by space. Tags are keywords used to describe your file.

Wiki-Tags:
My-Tags(Optional):

Use default thumbnail Use my thumbnail



Collaboration

Visible To: Public Domain Private

Sharing Option

Comments: Yes, allow comments. No, don't allow comments.

Rating: Yes, allow ratings by others. No, don't allow this to be rated.

Allow Tags: Yes, allow tagging by others. No, don't allow tagging.

Web User Mode

ResearchPoint
XBRL and Semantic Web (RDF) integrated

XBRL Rendering Interactive Charts View

(2) Web Rendered/User/Analytics Mode

ResearchPoint Demo

Added: 8/19/2009
From: GMR_Editor
File Name: DOWChemicalfromSEC.rif
Company: DOWXBRL
Wiki-Tags: [ResearchPoint Demo](#)
Add a Tag
My-Tags: [ResearchPoint Demo](#)
Add a Tag
FileType: Refreshable Intermediate Format
Description: ResearchPoint Demo
Source:
Collaboration: Public [Share/Distribute](#)
Export: [MS Excel](#) [XBRL](#)
Embed: `<iframe src=http://202.71.128.81`
URL: `http://202.71.128.81/godrx/Net`

Metric	2005	2006	2007
Purchased in-process research and development	USD57	USD10	USD72
Amortization of intangibles	USD50	USD55	USD578
Restructuring charges	USD591	USD114	USD959
Equity in earnings of nonconsolidated affiliates	USD324	USD137	USD964
Sundry income - net	USD755		

XBRL Rendering Tabular View

ResearchPoint Demo

File Name	Sheet Name	Company Name	Language	Parent Metric	Currency	Taxonomy Label	Taxonomy Id	Wiki Tags	Notes
DOWChemicalfromSEC.rif	Income Statement	DOWXBRL	English	Restructuring charges	Dollar United States	Restructuring Costs	ifrgg_RestructuringCosts	researchpoint demo	

Proof of Concept: Data Quality Heat Map Visualization

Heat Maps

- Data Quality Rating
- Data Alerts
- Properties Modified

Legend

- ★★★★★
- ★★★★☆
- ★★★☆☆
- ★★☆☆☆
- ★☆☆☆☆
- ★
- Untagged

Avg. Semantic Data Quality Score

389

★

Dow for XBRL Conversion

Added: 9/25/2009
 From: amar
 File Name: DOWforXBRL.rif
 Company: Demo XBRL

S.No	A	B	C	D	E	F	G
1	Dow Chemical Balance Sheet						
2	Year-end, \$mn (unless otherwise notes)	1998	1999	2000	2001	2002	2003
3	Assets:						
4	Cash & equivalents	2839	1903	235	123	506	215
5	Marketable securities	611	399	302	267	706	89
6	Accounts receivable, net	4109	4396	4958	4537	4614	5385
7	Inventories	2748	2815	2921	2810	2786	3463
8	Deferred taxes	247	317	224	303	235	108
9	Total current assets	10554	9830	8640	8040	8847	9260

http://202.71.128.81 - Property Box - Microsoft Inter...

G4

Cell Semantic Data Quality Score: 900

File Name: DOWforXBRL.xls
 Sheet Name: Sheet1
 Company Name: Demo XBRL
 Language: English
 Parent Metric: Cash & equivalents
 Currency: Dollar: United States
 Taxonomy Label: Cash and Cash Equivalents
 Taxonomy Id: ifrs-gp_CashAndCashEquivalents
 Wiki Tags: Demo for XBRL Conversion
 Notes: Demo for XBRL Conversion

EDIT

Comments & Responses

Rating not allowed for owner ★★★★★

[Report Abuse](#)

FileSize: 131.7 KB Views: 4

GICS Classification

Industry Group: [Others](#)
 Industry: [Others](#)
 Sub Industry: [Others](#)
 Company: [Demo XBRL](#)

Data Quality Rating Implementation

Standard Taxonomy →

Think: FICO Score for Data



Software interface for Data Quality Rating Implementation. The interface includes a menu bar (Setup Control, Generate RIF, Manage Models, XDMML Viewer, Taxonomy Thesaurus, XDMML-Analyzer, Set Semantic Rating), a Taxonomy Metrics section with a tree view of IFRS-GP assets, an Attributes Penalty table, and a Business Rules Penalty table. A 'Penalty Update' dialog box is open, showing a successful update message. A yellow box labeled 'Quality Rating Set-up' is overlaid on the right side, with arrows pointing to the 'Attributes' and 'Rules' sections.

Attributes Penalty

Language	100
English	100
Other	50
Currency	100
Period	100
WikiTag	200
Notes	100
Alert	100
Tot. Attr. Penalty	700

Business Rules Penalty

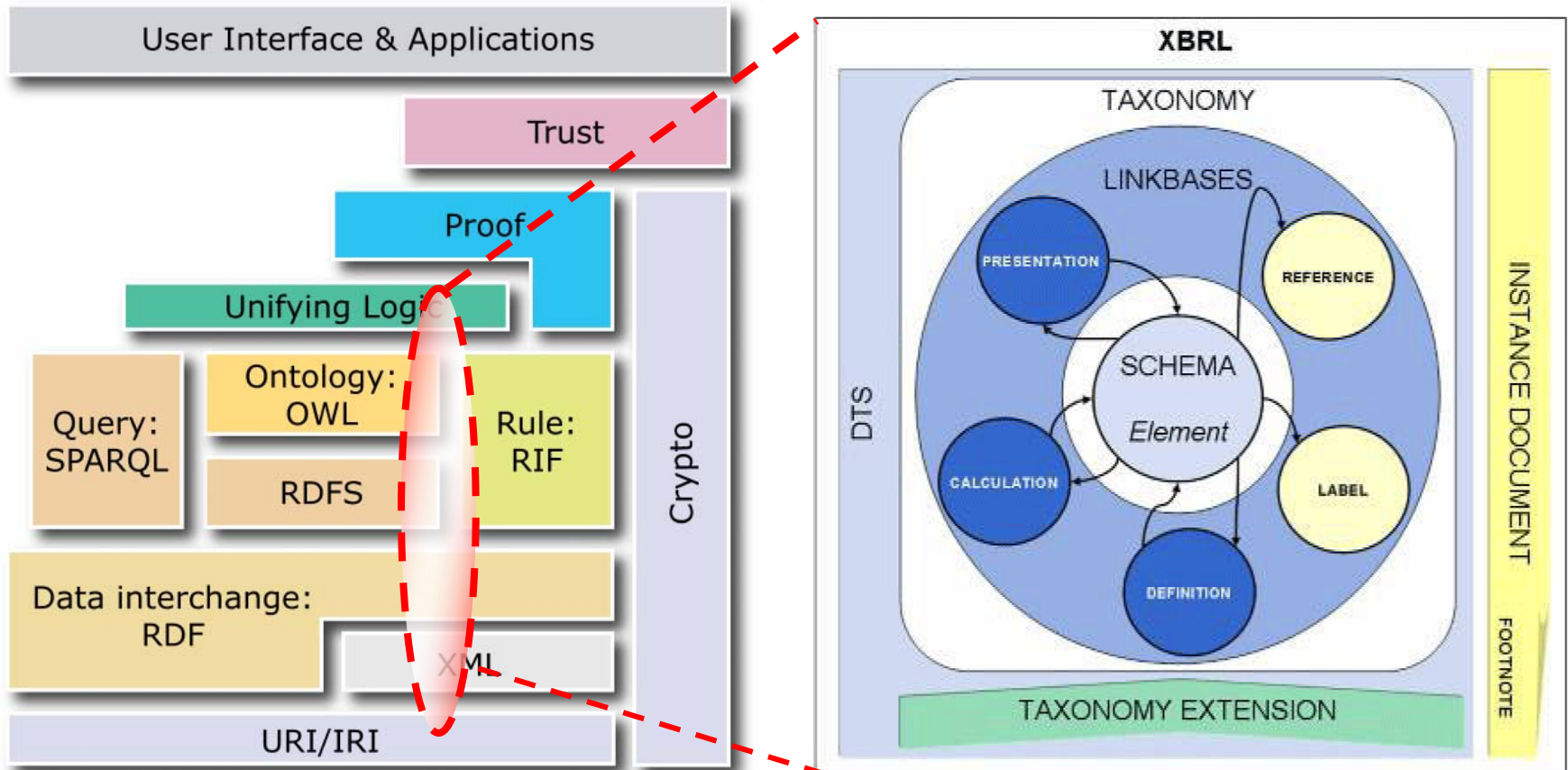
	Taxonomy Metrics	Operator	Data	Penalty
1	ifrs-gp_AssetsTotal	>	0	300
2	ifrs-gp_ResearchAndDevelopment	<	0	300
3	ifrs-gp_Inventories	<	0	300
4	ifrs-gp_RawMaterialsAndConsum...	>	0	300
5	ifrs-gp_AccruedLiabilitiesTotal	=	0	300
6	ifrs-gp_EmployeeExpensesByNat...	>	0	300
7	ifrs-gp_OtherOperatingExpenses	=	0	300
8	ifrs-gp_InvestmentsInJointVenture...	<	0	300
9	ifrs-gp_OtherLiabilitiesNonCurrent	>	0	300
11	ifrs-gp_CashAndCashEquivalents	<	100	200
1'	ifrs-gp_CashAndCashEquivalents	>	300	100
▶*				

Penalty Update dialog: Penalty updated successfully! OK

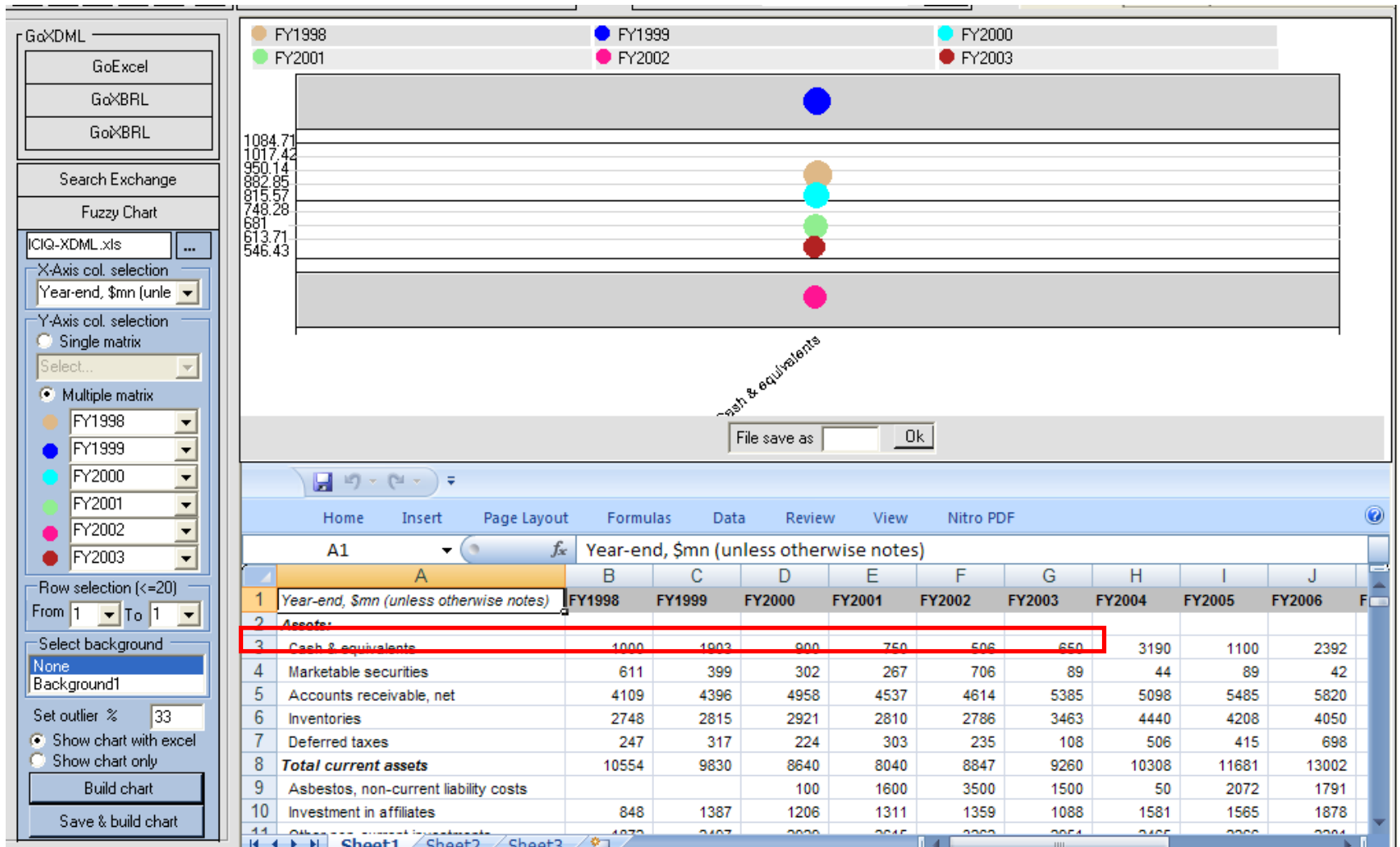
Update Penalty Score

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

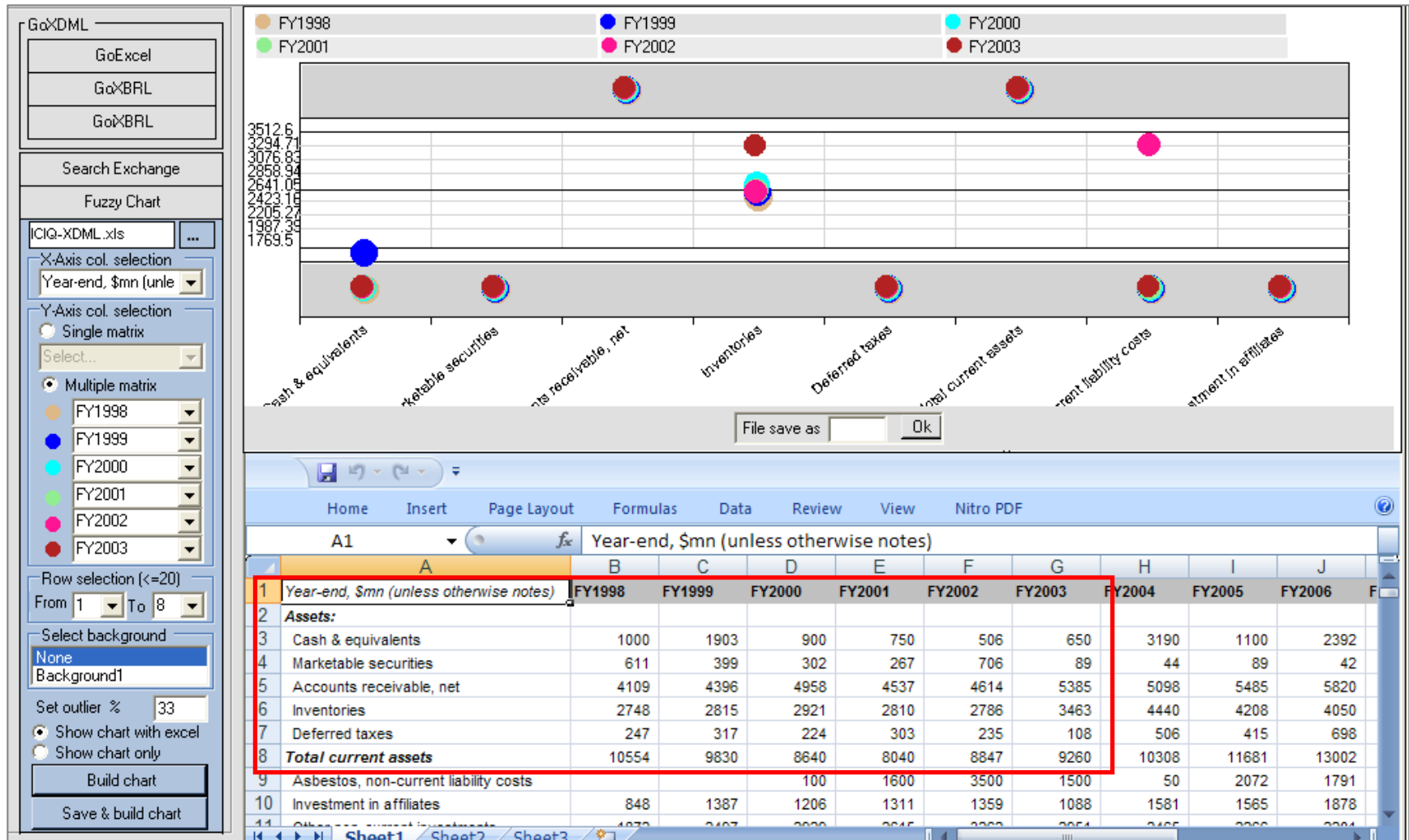
Semantic Web & XBRL Architectures



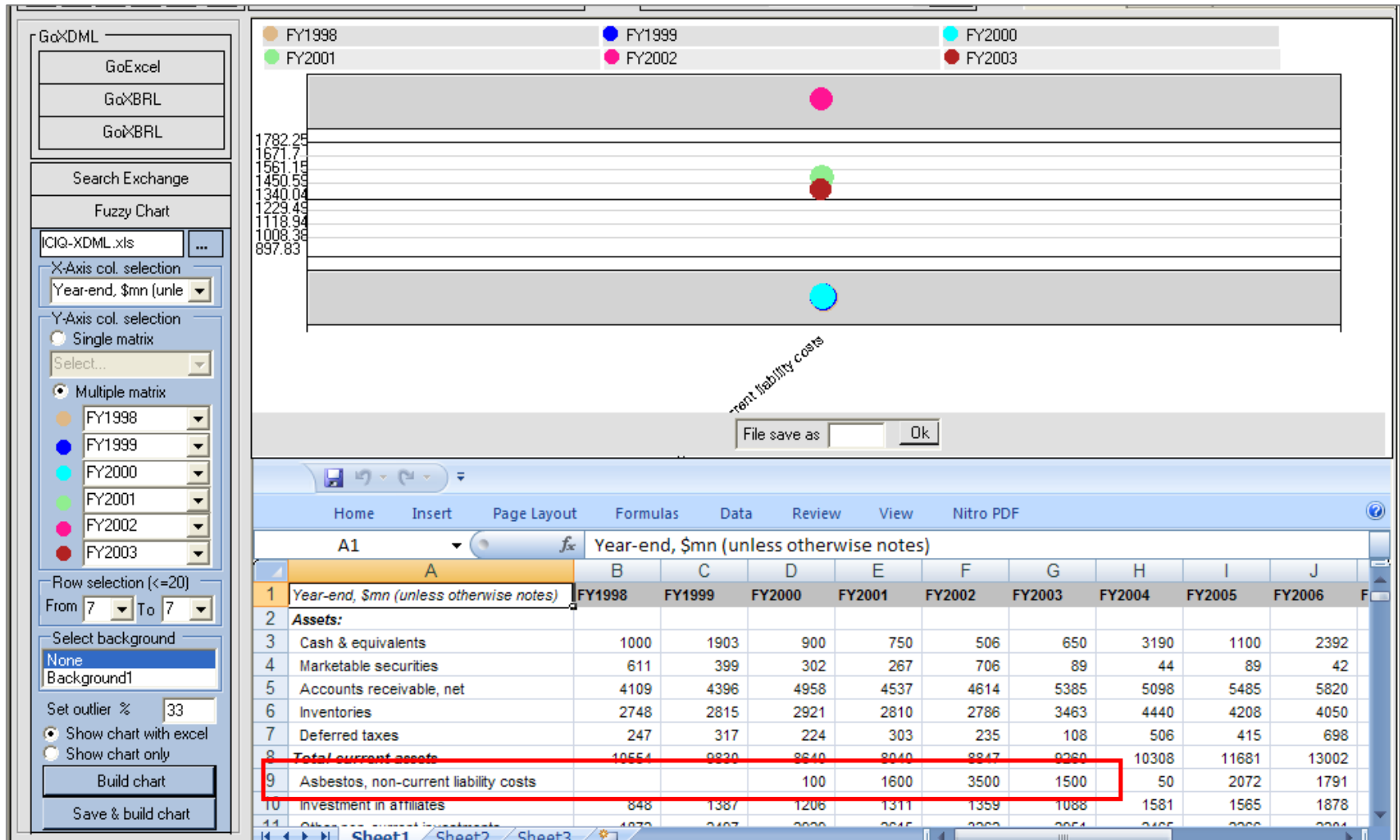
Use Case: Outlier Analytics – Single Company, Single Metric, Multi-Year



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



Use Case: Forensic Analytics – Alpha Insights



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.