Visualizing Financial Information Quality Using Heat Maps and Semantic Data Quality Rating System

ABSTRACT

Government regulators and investors in the global capital markets rely heavily on reported financial data and place implicit trust in the integrity and quality of this information. This financial data enters the information supply chain as manually processed data as company quarterly and annual filings. An intense focus on outlier data is of great interest as a potential compliance issue and as arbitrage opportunity for investment, also called alpha. Unfortunately, distinguishing between poor quality and valid outlier data is difficult for computers and requires manual screening. Given the vast amount of data it's also prone to human errors. We present a case study, where we applied business rules driven information quality ratings to automatically tag semantic data quality rank to each information element. Aggregated quality rank displays outliers as hot spots on heat maps enabling greater transparency and insights.

BIOGRAPHY

Ashu Batnagar
Chief Executive Officer
Good Morning Research

Ashu Bhatnagar is CEO of Good Morning Research, a Softpark company that specializes in building Semantic XBRL technology for Wall Street banks, hedge funds and government regulators in USA, UK, and India. Mr. Bhatnagar has over ten years of experience in working at Wall Street banks as product manager, and before that for fifteen years in the computer industry as VLSI designer, software product manager and Internet entrepreneur. Mr. Bhatnagar also taught Advanced VLSI Design and Supercomputer Architectures courses at graduate level as adjunct professor at University of Massachusetts, Lowell. Mr. Bhatnagar has an undergraduate degree in Electrical Engineering from Indian Institute of Technology, Roorkee, and Masters in EE from University of Rhode Island. He also studied as a graduate student at MIT in the area of Advanced VLSI Design. Mr. Bhatnagar has presented papers at International Conferences on Data Quality, Semantic Technology and W3C/FDIC Workshop on XBRL.

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Visualizing Financial Information Quality Using Heat Maps and Semantic Data Quality Rating System

Assets:												
Cash & Cash Equivalents	$3,016	$6,438	$14,304	$4,851	$6,714	$6,111	$10,339	$6,076	$5,505	$8,207	$14,546	$23,096
Short-Term Investments	35,636	42,610	46,288	32,900	27,447	17,300	13,323	25,371	31,283	39,345	43,789	48,233
Accounts Receivable	5,129	5,196	5,890	7,180	9,316	11,338	13,589	11,192	13,014	13,840	14,687	15,658
Inventory	673	640	421	491	1,478	1,127	985	717	740	828	900	973
Current Deferred Taxes	2,112	2,506	2,097	1,701	1,940	1,899	2,017	2,213	2,184	2,344	2,344	2,344
Other	2,010	1,583	1,566	1,614	2,115	2,393	2,989	3,711	2,950	3,024	3,175	3,348
Current Assets	**$48,576**	**$58,973**	**$70,566**	**$48,737**	**$49,010**	**$40,168**	**$43,242**	**$49,280**	**$55,676**	**$67,588**	**$79,448**	**$93,651**
Property & Equipment	2,268	2,223	2,326	2,346	3,044	4,350	6,242	7,535	7,630	8,428	8,625	8,729
Equity and Other Investments	14,191	13,692	12,210	11,004	9,232	10,117	6,588	4,933	7,754	10,081	11,241	12,401

Ashu Bhatnagar

Founder, CEO
GoodMorningResearch.com, and
FreeXBRLonline.com

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Special Acknowledgement:

To Colin Ritchie, portfolio manager at an Australian hedge fund for his contributions to the case study. In particular, his use of visualization tools – *RitchViewer*™ for analytics and *ResearchPoint*™ for heat map display of underlying data quality based on semantic ratings. The case study demonstrates the visualization of data quality in a realistic-day-in-the-life of a hedge fund manager.

Disclosures:

This presentation is not an offer to buy or sell any security or to participate in any trading strategy. All of the sample data and information in this presentation are based on public information and modified from the original sources for this study. While every effort was made to use reliable and comprehensive information, we do not represent that it is accurate or complete. All third-party trademarks, service marks and copyrights in this presentation belong to their respective owners.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

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
- Present a case study for visualizing data quality based on semantic ratings and displayed as heat maps for fast drill-down analysis.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Data Quality - Challenges

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

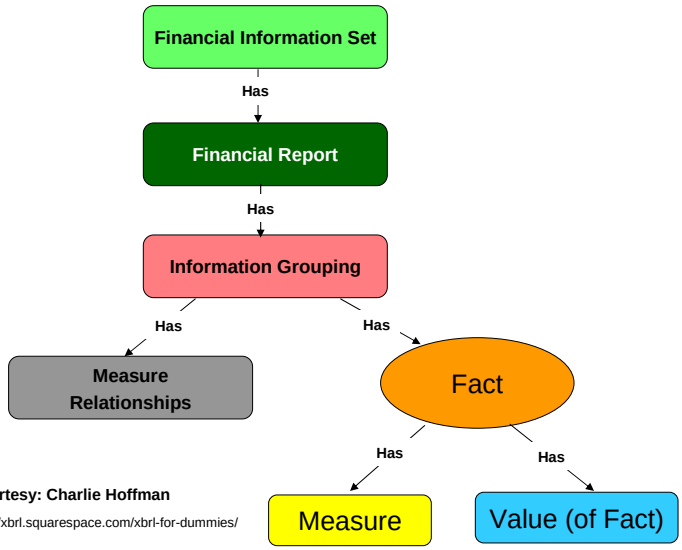
The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Financial Data is a Multi Dimensional Hypercube

Dimensions:

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

. . .



Courtesy: Charlie Hoffman
http://xbrl.squarespace.com/xbrl-for-dummies/

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

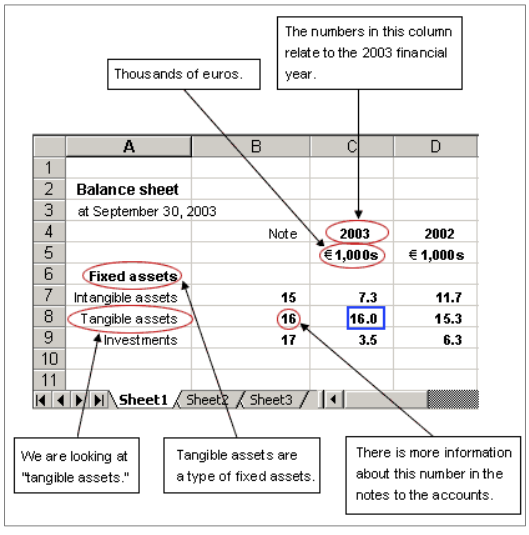
The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

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



Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

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!

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Challenge 3: Diversity of Semantics

Strategic View

Data → Information → Analysis → Insights

Filers/Companies

- SEC Filings (10-K/10Q)
- Investor Presentations
- Earnings / Call Reports
- Annual Reports

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

Long Tail View

Long Tail of Data Organization

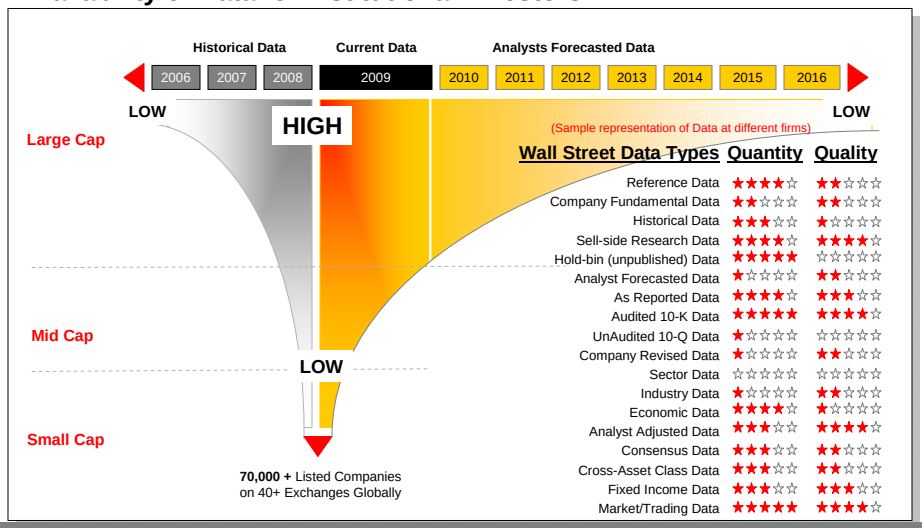
World of Accounting (GAAP)	World of Analysts (Industry Specific Taxonomies)	World of Investors (Folksonomy)
Wiki-Tagged Data Organization Process		
50 to 100 common Financial Metrics for • Income Statement • Balance Sheet • Cash-flow	**Additional 2,000 +** unique Financial Metrics for • Financial Modeling • Valuation • Forecasts, • Trading Strategies	**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

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

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

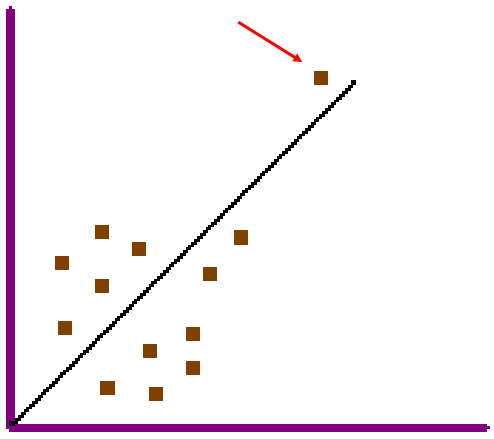
Availability of Data for Institutional Investors



Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Challenge 5: Distinguishing Data Error from Alpha Opportunity?



Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

A Practice-Oriented Solution Using

Semantic Tagging For Data Quality Ranking

And

Heat Map Visualization

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

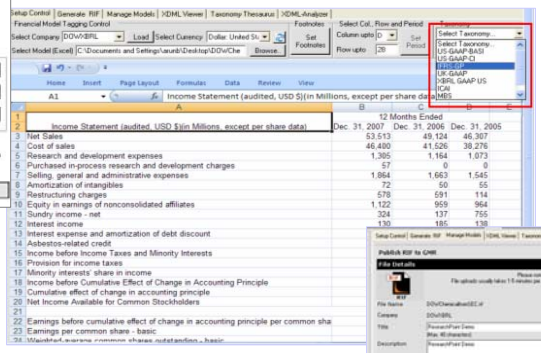
The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

ResearchPoint™

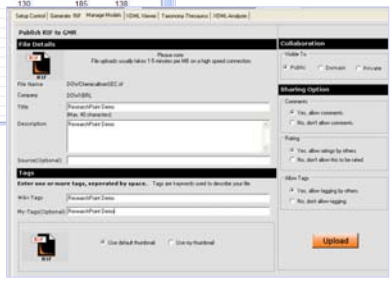
Excel, XBRL and Semantic Web (RDF) integrated

Login

Author Mode



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**

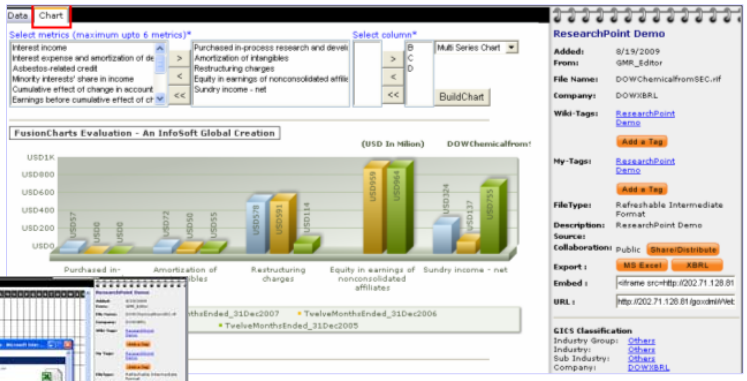
Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

ResearchPoint™
Excel, XBRL and Semantic Web (RDF) integrated

Visualization
Interactive Charts View

User Mode

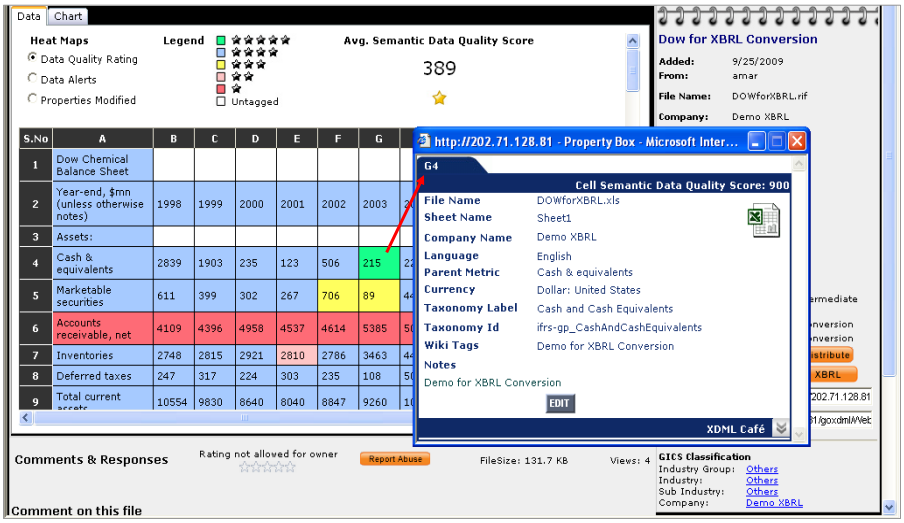


Visualization
Tabular View

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Proof of Concept: Data Quality Heat Map Visualization

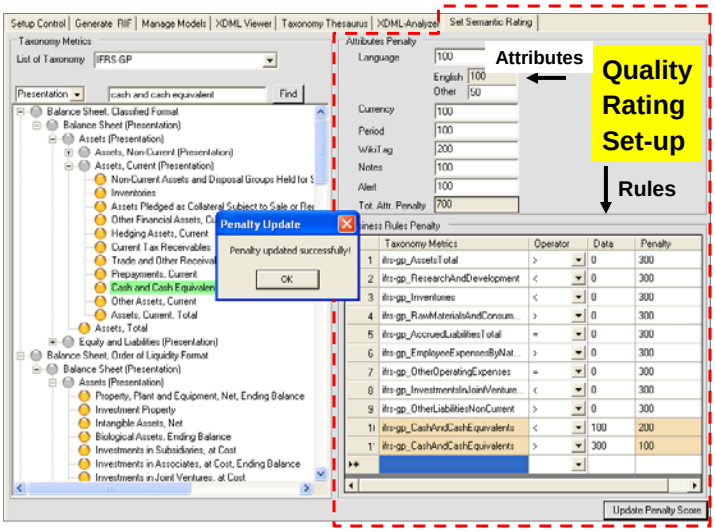


Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Data Quality Rating Implementation

Standard Taxonomy



Think:
FICO Score
for Data

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Case Study: **A day-in-the-life of a Hedge Fund Manager**

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Case Study: A day-in-the-life of a Hedge Fund Manager

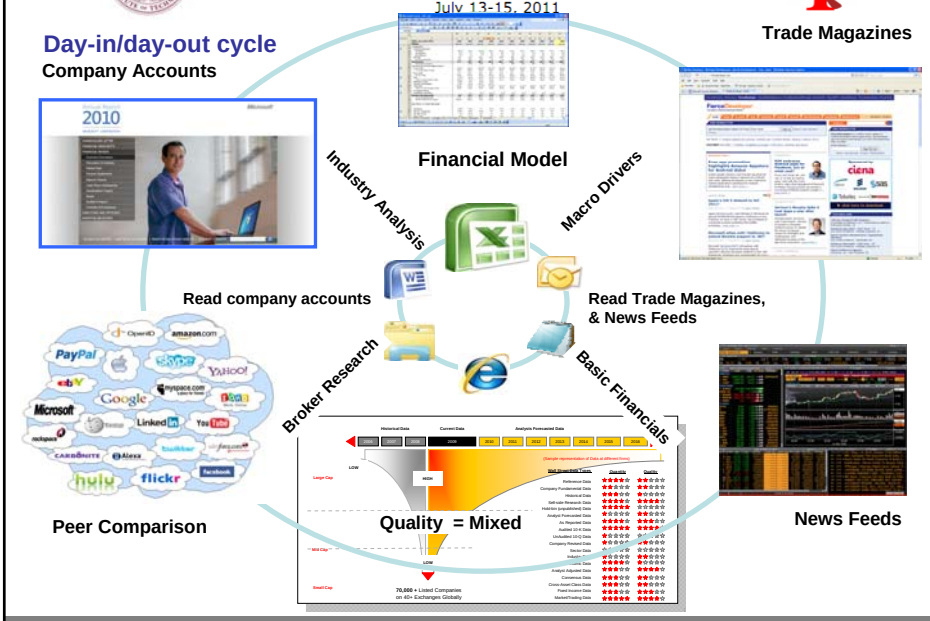
As part of stock Analysis, Portfolio Managers and Analysts:

- ❑ Read company accounts
- ❑ Read Trade Magazines, & News Feeds
- ❑ Read Broker Research
- ❑ Industry Analysis
- ❑ Peer Comparison
- ❑ Macro Drivers
- ❑ Triggers: (What will cause change)
- ❑ Basic Financials
- ❑ Detailed Financials (Model)

This is a day-in/day-out cycle.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011



Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Trigger:

Guinness World Records, the global authority on record breaking, today confirm that the *Company XYZ* product is the ***Fastest-Selling Consumer Electronics Device***.

The sales figures outstrip both the iPhone and the iPad for the equivalent periods after launch.

March 2011

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Start Analysis:

So the Fund Manager wants to look at a Model in more detail.

At the Model stage, the choice for a Fund Manager is to build and maintain an internal model, or reach out to a favourite Sell-side analyst(s) and obtain their latest model.

RitchViewer™ FinancePack Screen shots: Courtesy Colin Ritchie

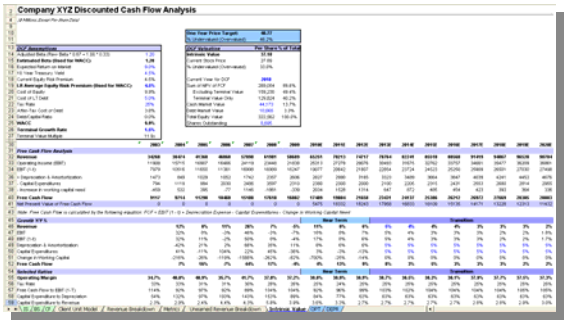
Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Analysis for Alpha:

To extract 'alpha' from the model, the Portfolio Manager needs to work through many components of the accounts being presented.

In this example, we are examining the Analysts' supplied Discounted Cash Flow Analysis.



A company's value can be described as the sum of all future cash flows discounted back to today's value.

Valuation = fcfY1/(1-dr) + fcfY2/(1-dr)^2 + … + fcfYn/(1-dr)^n-1

where:

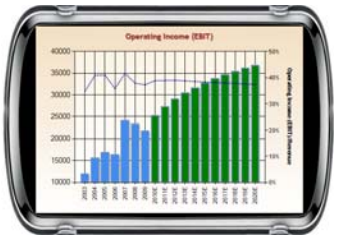
fcf = Free Cash Flow, and dr = Discount Rate

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Analysis (walk-thru-the-data):

- ✓ Model forecasts growth in revenue from 2009 to 2020.
- ✓ Revenue growth rate is falling away from double-digit in historic years to <2% by 2020.

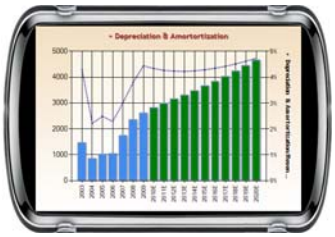


- ✓ EBIT growing from 2010 to 2020; and as importantly as a function of Sales Revenue.
- ✓ EBIT Margin (EBIT/Revenue) looks to be fairly constant over the forecast years.
- ✓ **So far the data looks reasonable**

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Analysis (walk-thru-the-data):



- ✓ Model shows Depreciation of Assets growing *(LHS),* and
- ✓ As a function of Revenue *(RHS line-chart*)

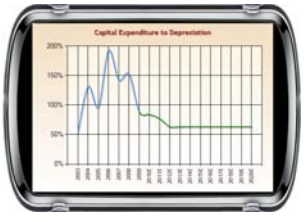
Capex to Revenue however looks odd as the model shows it taking until 2016 for the Capex spend to get back to 2009 levels – even though Revenue is growing!

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Analysis (walk-thru-the-data):

- ✓ Switching charts to depict Capex as a percentage of Revenue, we see the model is forecasting a step-down in future capex – even though we know that Company XYZ is investing to keep up with the competition.

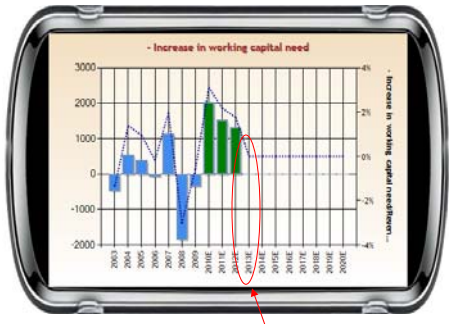


- ✓ Typically this item is a function of the rate the company is depreciating its assets (Capex/Depn) or a rate against future Revenue (Capex/Revenue)

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Drill-down Analysis



Moving down to Working Capital, we see a real issue with the data.

Working Capital is a function of the Capital that is need to be set aside for the Increase in Accounts Payable, Accounts Receivable, and Inventory that goes with any increase in Revenue.

Yet the Model has this data falling to **zero** in 2013.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Quantifying the Problem and its Impact:

Earnings per Share (EPS)	Jun-11	Jun-12
Mean Estimate (μ)	2.56	2.76
Standard Deviation (σ)	0.07	0.10
Coefficient of Variation (C_v)	2.7%	3.7%

$$C_v = \frac{\sigma}{|\mu|}$$

For a company like Company XYZ, this item is probably less noticed than an Industrial widget maker. A quick calculation shows a 6% impact on valuation by having this number missing.

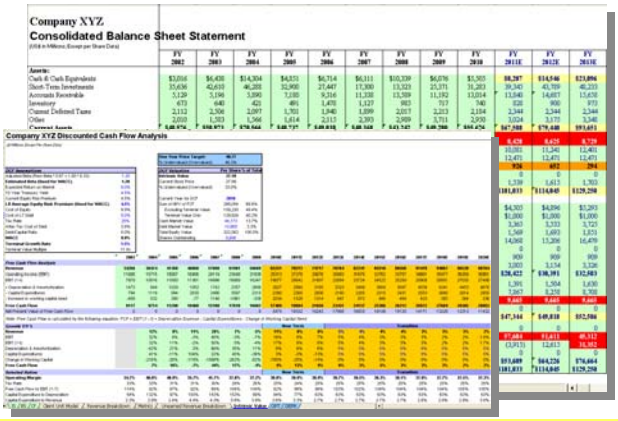
Now 6% may not appear much, but Finance Theory tells us that extracting true Alpha for a stock like Company XYZ is difficult due to the breadth of analyst's coverage.

If we take a group of 28 Analysts in the Market we find that the Coefficient of Variation is only 2.7% for 2011. This means the 6% impact on valuation is around double one standard deviation from the mean of Analysts estimates.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Visualizing Financial Information Quality Using Heat Maps and Semantic Data Quality Rating System



✓ Rather than walking through the Analysts model, a slow and tedious process, we upload the Excel model to *ResearchPoint*™ and examine the quality of the Balance Sheet via semantic rated Data Quality Heat Map.

Key finding: Quickly we find that the Balance Sheet is only forecast to 2013, where as the Discounted Cash Flow relies on Data to 2020.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Case Study Conclusion:
A day-in-the-life of a Hedge Fund Manager

Given this knowledge, the Fund Manager is left with several choices –

a) Source a new model, or
b) rework the Working Capital component of the Discounted Cash Flow extrapolating from Sales using Debtor Days, Creditor Days, and Days Inventory.

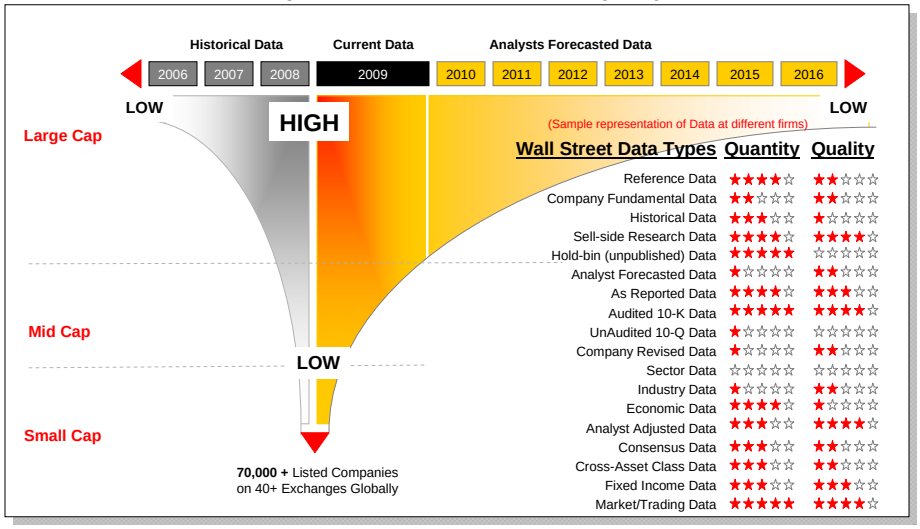
In the case, the Fund Manager chose the later and discovered the 6% impact on valuation due to missing data.

Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

In Conclusion:

Automated Semantic Ratings Enables Next Level of Data Quality Visualization – Faster!



Copyright © 2011. GoodMorningResearch.com – A Softpark Company

The 5th MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Information Quality Industry Symposium
Cambridge, Massachusetts, USA
July 13-15, 2011

Thank you

Ashu Bhatnagar

Email: ashu@softpark.com

FREE XBRL ONLINE

Copyright © 2011. GoodMorningResearch.com – A Softpark Company