



shaping tomorrow with you



Empower Flexibility via Big Data: Unleashing Constraints in Financial Analytics

Pierre-Yves Vandebussche – Fujitsu
pierre-yves.vandebussche@ie.fujitsu.com

17 April 2013

- Take advantage of **Open Big Data** published on the Web
- Integrate Large Scale **Holistic data** to provide a more comprehensive information
- Perform **knowledge analytics** while preserving data provenance
- Enable **concept** (from diverse integrated sources) based **metric calculation**

Challenges

■ Interoperability between Open Data

- lack of inter data sources links
- format heterogeneity (XML, schemas, API, WebSite crawling, CSV files, etc.)

```
<!-- Copyright (c) 2005-2011 EDGAR Online, Inc. All F  
<xbrl xmlns="http://www.xbrl.org/2003" instance="http://www.xbrl.org/2003/instance" xmlns:us-gaap="http://www.xbrl.org/2009/us-gaap" xmlns:dbpedia="http://www.dbpedia.org/ontology" xmlns:dbpedia-owl="http://www.dbpedia.org/ontology" xmlns:dbpedia-property="http://www.dbpedia.org/ontology" xmlns:dbpedia-class="http://www.dbpedia.org/ontology" xmlns:dbpedia-industry="http://www.dbpedia.org/ontology" xmlns:dbpedia-service_provider="http://www.dbpedia.org/ontology" xmlns:dbpedia-telecoms_equipment="http://www.dbpedia.org/ontology" />  
<xbrll:schem /www.xbrl.org  
<us-gaap:Sha mentByShareBe  
<us-gaap:Sha mentByShareBe  
<us-gaap:Pro fitedCurrent cc  
<us-gaap:CashAndCashEquivalentsAtCarryingValue cont
```

dbpedia-owl:foundedBy

- dbpedia:Sergey_Brin
- dbpedia:Larry_Page

dbpedia-owl:foundingDate

1998-09-04 (xsd:date)

dbpedia-owl:foundingDate

1998-09-01 00:00:00 (xsd:date)

dbpedia-owl:industry

- dbpedia:Telecoms_equipment

As Web Search Goes Mobile,

record all that you and your loved ones are going thr
April 03, 2013, Wednesday

Google Inc

NASDAQ: GOOG - 4 Apr 08:44 ET

806.20 +0.00 (0.00%)

805
800

10am 11 12 1 2 3 4pm

1 d 5 d 1 m 6 m 1 y 5 y max

Competitors

Technorati, IceRocket, Microsoft, Aductions, Zoho, Yahoo!, Danger,

Talaria, 3/13 1

Channel Intelligence, \$125M

2/13 2

- Lack of a unique **worldwide identifier** for a company

- **XBRL Taxonomies**

- updated by regulation in a frequent manner. Backward compatibility?
- taxonomy alignment (XBRL concepts \Leftrightarrow Banking Regulations)

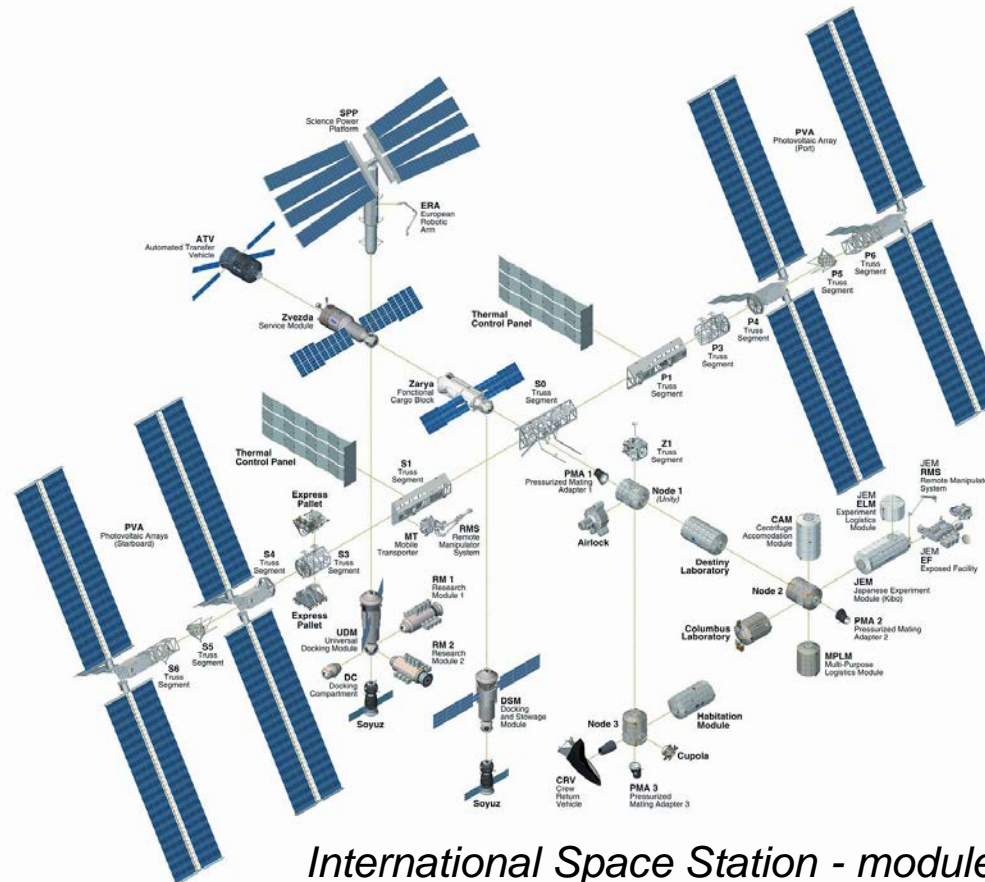


GAAP @en
IFRS @jp

The background of this section contains snippets of XML code, including: `<!-- copyright 11 EDGAR Online`, `<xlink:arcrole="http://"`, `<us-gaap:ShareBasedCompensationArrangem`, `<us-gaap:ShareBasedCompensationArrangem`, `<us-gaap:ProductWarrantyAccrualClassifi`, `<us-gaap:CashAndCashEquivalentsAtCarryi`, `<us-gaap:ShareBasedCompensationArrangem`, and `<us-gaap:ShareBasedCompensationArrangem`.

Challenges

- Integrate and represent all Data in a **standard model** to
 - process knowledge analysis algorithms over all data
 - allow concepts calculation regardless of the provenance (XBRL, New York Times, DBPedia, Stock Price, etc.)



International Space Station - modules integration

■ Use Linked Data as

- a technology to **bridge** Open Big Data set silos and facilitate **interoperability**
- a **graph model** to represent all the data in a **unique format** and provide a **standard access**

Existing Problem	Solution we offer
Entity identification and relations	Use of the LEI (legal Entity Identifier) to uniquely identify companies. Create some mapping between taxonomies
Disparate data formats	Represent all the data in a unique format (holistic data integration)
Knowledge analytics	Compute company sentiment based on New-York Times press articles

■ Customisable interactive data

Existing Problem	Solution we offer
Combining concepts coming from disparate data sources	Develop a KPI/formula editor for a user to build its own comparison metric regardless of the data provenance (e.g. XBRL and DBPedia)
Few company selection criteria	Use XBRL data and Open Linked Data information to filter companies to compare



Hikaku

A new company comparison experience

Filters

From: 07/01/2002
To: 11/30/2012

Datasources

- Company sentiment (extracted from NYTimes articles)
- Crunchbase (LOD dataset)
- DBPedia (LOD dataset)
- Financial data from Edgar filings (instant data)
- Financial data from Edgar filings (period data)
- New-York Times (LOD dataset)
- Stock prices

- ^ Sentiment Value + -
- ^ Funding (USD) + -
- ^ Number of Employees + -
- ^ Assets (USD) + -
- ^ AssetsCurrent (USD) + -
- ^ InventoryNet (USD) + -
- ^ Liabilities (USD) + -
- ^ LiabilitiesCurrent (USD) + -
- ^ StockholdersEquity (USD) + -
- ^ EarningsPerShareBasic (USD) + -
- ^ EarningsPerShareBasic (USD/share) + -
- ^ EarningsPerShareDiluted (USD) + -
- ^ EarningsPerShareDiluted (USD/share) + -
- ^ GrossProfit (USD) + -
- ^ NetIncomeLoss (USD) + -
- ^ OperatingIncomeLoss (USD) + -
- ^ SalesRevenueNet (USD) + -
- ^ Number of Articles + -
- ^ Stock Price (USD) + -

Introduction | **Chart** | Timeline

Hikaku application gives you the ability to compare companies:

- By adding concepts on the left from various data sources.
- By creating your own KPI formula (button to the right).

You also can select the indicators you want to display in the charts by choosing (View Chart) from column's header menu and selecting companies (using checkbox to the left).

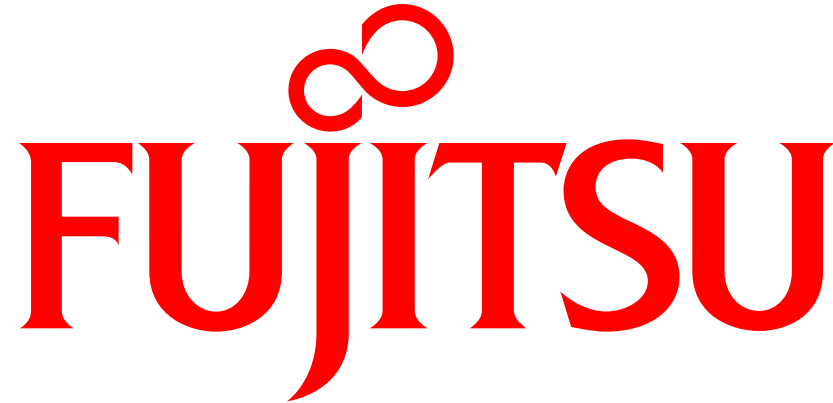
<input type="checkbox"/>	Company	Info	Symbol	CIK	LEI	Description	Founding Date	Depiction
<input type="checkbox"/>	Google Inc.		GOOG	0001288776	7ZW8QJWV	Google Inc. i	1998-09-04	http://upload
<input type="checkbox"/>	QUALCOMM INC		QCOM	0000804328	H1J8DDZKZ	Qualcomm In		http://upload
<input type="checkbox"/>	MICROSOFT CORP		MSFT	0000789019	NR2EJN1ER	Microsoft Co	1975-04-04	http://upload
<input type="checkbox"/>	ADOBE SYSTEMS INC		ADBE	0000796343	FU4LY2G49	Adobe Syst		http://upload
<input type="checkbox"/>	ORACLE CORP		ORCL	0001341439	54930074LK	Oracle Corpr	1977-06-16	http://upload
<input type="checkbox"/>	CISCO SYSTEMS, INC.		CSCO	0000858877	8E6NF1YAL	Cisco System		http://upload
<input type="checkbox"/>	HEWLETT PACKARD CO		HPQ	0000047217	WHKXQACZ	Hewlett-Pac		http://upload
<input type="checkbox"/>	EMC CORP		EMC	0000790070	KDMDXRWU	EMC Corpor		http://upload
<input type="checkbox"/>	SYMANTEC CORP		SYMC	0000849399	YF6ZV0M6A	Symantec Ci	1982-03-01	http://upload
<input type="checkbox"/>	YAHOO INC		YHOO	0001011006	549300ZJC0	Yahoo! Inc. i	1995-03-01	http://upload
<input type="checkbox"/>	INTERNATIONAL BUSINESS MACHINES CORP		IBM	0000051143	9LH5FXPWK	International	1911-06-16	http://upload
<input type="checkbox"/>	APPLE INC		AAPL	0000320193	HWUPKROMI	Apple Inc. is	1976-04-01	http://upload
<input type="checkbox"/>	DELL INC		DELL	0000826083	3E70L4WYA	Dell, Inc. is a	1984-11-04	http://upload
<input type="checkbox"/>	VERIZON COMMUNICATIONS INC		VZ	0000732712	2S72Q52U0	Verizon Com	1983-10-07	http://upload

- Using Linked Data technology to tackle Open Data interoperability issue
 - Use this technology to encompass more data sources (EDINET, etc.)

- Experiment innovative Sentiment Value from text
 - Use this algorithm to detect facts from XBRL free text description and other Open data sources

- Enhance financial analysts experience by allowing KPI calculation using concepts regardless of their source
 - Enable dynamic addition of data sources

Thank you for your attention.
Also demonstrating at Fujitsu booth!



shaping tomorrow with you

Acknowledgement

Fujitsu Team: Goto Masatomo, Gofran Shukair, Vivian Qian Li, Jurgen Umbrich, Yutaka Mitsuishi, Aisha Naseer

DERI Team: Sean O’Riain, Richard Cyganiak, Vit Novacek

Components Architecture

