Applying Linked Data Technologies to Financial Reporting

Benedikt Kämppgen, André Freitas, Andreas Harth, Seán O’Riain
Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- UC2: Understanding the Origin and Trustworthiness of XBRL
- Relation to DDI Lifecycle
Financial analysis

Task: Find sales and operating income for each segment of company 三精輸送機

Thanks to Craig Weber for this example
Financial analysis (2)
Extensible Business Reporting Language (XBRL): Describes business reporting information in an XML format, for exchange, and comparison [XBRL SPEC].
RAYONIER INC had a sales revenue net of 377,515,000 USD from 2010-07-01 to 2010-09-30 [XBRL example]
U.S. Securities and Exchange Commission database

RAYONIER INC had a sales revenue net of 377,515,000 USD from 2010-07-01 to 2010-09-30 [XBRL example]
Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- UC2: Understanding the Origin and Trustworthiness of XBRL
- Relation to DDI Lifecycle
UC1: Integrating XBRL with External Data Sources

CalcBench: Comparison of heterogeneous reports

<table>
<thead>
<tr>
<th>RAYONIER INC</th>
<th>VIEW IN SPREADSHEET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condensed Consolidated Statements of Income and Comprehensive Income - Quarterly</td>
<td></td>
</tr>
</tbody>
</table>

2Q 2012 4/1 - 6/30/2012

<table>
<thead>
<tr>
<th>Name</th>
<th>2Q 2012</th>
<th>WELLPOINT, INC 2Q 2012 4/1 - 6/30/2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>$371,926,000</td>
<td>$3,314,200,000 (88.78%)</td>
</tr>
<tr>
<td>Costs and Expenses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of sales</td>
<td>$262,555,000</td>
<td>$6,700,003,818.73%</td>
</tr>
<tr>
<td>Adjusted Cost of Sales</td>
<td>$262,000,000</td>
<td>6,500,000</td>
</tr>
<tr>
<td>Selling and general expenses</td>
<td>$16,250,000</td>
<td>$2,081,800,000 (99.22%)</td>
</tr>
<tr>
<td>Other operating (income) expense, net</td>
<td>($5,299,000)</td>
<td>-</td>
</tr>
<tr>
<td>Costs and Expenses, Total</td>
<td>$273,506,000</td>
<td>$</td>
</tr>
<tr>
<td>Equity in income of New Zealand joint venture</td>
<td>$170,000</td>
<td>$0</td>
</tr>
<tr>
<td>Operating Income</td>
<td>$98,590,000</td>
<td>$14,358,700,000 (99.31%)</td>
</tr>
<tr>
<td>Operating Margin %</td>
<td>26.51%</td>
<td>433.25%</td>
</tr>
<tr>
<td>Interest expense</td>
<td>($16,056,000)</td>
<td>$117,600,000 (86.35%)</td>
</tr>
<tr>
<td>Interest and miscellaneous income, net</td>
<td>$85,000</td>
<td>$1,048,600,000 (99.99%)</td>
</tr>
<tr>
<td>Income Before Income Taxes</td>
<td>$82,619,000</td>
<td>$1,048,600,000 (92.12%)</td>
</tr>
<tr>
<td>Income tax expense</td>
<td>($13,540,000)</td>
<td>$405,000,000 (96.66%)</td>
</tr>
<tr>
<td>Net Income</td>
<td>$69,079,000</td>
<td>$643,600,000 (89.27%)</td>
</tr>
<tr>
<td>Net Income Margin %</td>
<td>18.57%</td>
<td>19.42%</td>
</tr>
</tbody>
</table>
UC1: Integrating XBRL with External Data Sources

- CalcBench: Comparison of heterogeneous reports

**Rayonier Inc**

<table>
<thead>
<tr>
<th></th>
<th>2Q 2012</th>
<th>COMPARE TO COMPANY: WELLPOINT, INC 07/30/2012</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SALES</strong></td>
<td>$371,926,000</td>
<td>$3,314,200,000 (88.78%)</td>
</tr>
<tr>
<td><strong>Costs and Expenses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of sales</td>
<td>$262,555,000</td>
<td>$6,700,000,3,818.73%</td>
</tr>
<tr>
<td><strong>Name:</strong> Adj. Cost of Sales</td>
<td>X share</td>
<td></td>
</tr>
<tr>
<td>Selling and general expenses</td>
<td>$16,250,000</td>
<td>$2,081,800,000 (99.22%)</td>
</tr>
<tr>
<td>Other operating (income) expense, net</td>
<td>($5,299,000)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Costs and Expenses, Total</strong></td>
<td>$273,506,000</td>
<td>-</td>
</tr>
<tr>
<td>Equity in income of New Zealand joint venture</td>
<td>$170,000</td>
<td>$0</td>
</tr>
<tr>
<td><strong>OPERATING INCOME</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Operating Margin %</strong></td>
<td>26.51%</td>
<td>433.25%</td>
</tr>
<tr>
<td>Interest expense</td>
<td>($16,056,000)</td>
<td>$117,600,000 (86.35%)</td>
</tr>
<tr>
<td>Interest and miscellaneous income, net</td>
<td>$85,000</td>
<td>$1,048,600,000 (99.99%)</td>
</tr>
<tr>
<td><strong>INCOME BEFORE INCOME TAXES</strong></td>
<td>$82,619,000</td>
<td>$1,048,600,000 (92.12%)</td>
</tr>
<tr>
<td>Income tax expense</td>
<td>($13,540,000)</td>
<td>$405,000,000 (96.66%)</td>
</tr>
<tr>
<td><strong>NET INCOME</strong></td>
<td>$69,079,000</td>
<td>$643,600,000 (89.27%)</td>
</tr>
<tr>
<td><strong>Net Income Margin %</strong></td>
<td>18.57%</td>
<td>19.42%</td>
</tr>
</tbody>
</table>

Computation of missing values (e.g., total revenues)
UC1: Scenarios

Can we group SEC financial companies by operation industry from Freebase and get the sales revenue net and cost of goods sold to compute the average Gross Profit Margin?

Can we compare the sales revenue net and cost of goods sold of SEC companies with expenses for automobile loans reported in FFIEC call reports?
UC1: Approach

- Using RDF Data Cube Vocabulary (QB) for representing XBRL data

UC1: Approach – Architecture

- Saiku OLAP client
- Olap4Ld
- Open Virtuoso
- LDSpider crawling
  - SEC XBRL database (Edgar Linked Data Wrapper [EDGAR])
  - Freebase (already represented as Linked Data)
  - FFIEC XBRL data (FDIC Linked Data Wrapper [FFIEC])

For demo, see http://xbrl.us/research/appdev/pages/275.aspx#
UC1: Approach – Linking of Data Sources

SEC XBRL Data

Freebase Data

FFIEC XBRL Data

CIK number

freebase:Company

freebase:BusinessOperatingIndustry

edgar:Issuer

fdic:Issuer

xsd:date

dgar:Concept

ffiec:Concept

qb:Observation

B. Kämpgen – Applying Linked Data Technologies to Financial Reporting
Result: Many more Sales revenue net than Cost of goods sold.

Open problem: Manual effort in creation of additional dimensions.
UC1: Integrating SEC XBRL with FFIEC XBRL

- Result: Selected companies do not provide both SEC and FFIEC information.

- Open problem: Missing links between companies in SEC and FFIEC.
Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- UC2: Understanding the Origin and Trustworthiness of XBRL
- Relation to DDI Lifecycle
UC2: Understanding the Origin and Trustworthiness of XBRL

ETL over heterogeneous data sources
UC2: Understanding the Origin and Trustworthiness of XBRL

- ETL over heterogeneous data sources

1. Print Log
   - Units Graph
   - People Graph
   - Paper Usage
   - Commute Survey Results
   - Travel Requests

2. Generate Print Jobs Graph
3. Aggregate Print Jobs and Load the Print Usage Cube into the DWH
4. Print Cube
5. Do Rollup Aggregations to generate KPIs

<table>
<thead>
<tr>
<th></th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>printing emissions</td>
<td>600</td>
<td>503</td>
</tr>
<tr>
<td>paper usage</td>
<td>4165</td>
<td>3968</td>
</tr>
<tr>
<td>travel emissions</td>
<td>534000</td>
<td>429193</td>
</tr>
<tr>
<td>commute emissions</td>
<td>456</td>
<td>391</td>
</tr>
</tbody>
</table>

Carbon dioxide emission by kg
UC2: Scenario

- How **current** is this information?
- What **aggregation functions** were used?
- What **source** do the raw data come from?

<table>
<thead>
<tr>
<th></th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>printing emissions</td>
<td>600</td>
<td>503</td>
</tr>
<tr>
<td>paper usage</td>
<td>4,165</td>
<td>3,968</td>
</tr>
<tr>
<td>travel emissions</td>
<td>534,000</td>
<td>429,193</td>
</tr>
<tr>
<td>commute emissions</td>
<td>456</td>
<td>391</td>
</tr>
</tbody>
</table>

Carbon dioxide emission by kg
UC2: Approach

- RDF vocabulary for representing ETL elements
- Simple workflow structure of processes, artifacts, agents
- Different abstraction levels through nested workflows

https://sites.google.com/site/co4svocab/
UC2: Cogs – Structure

- Taxonomy of common ETL concepts, e.g.,
  - Execution
  - State
  - Extraction, Transformation, Loading
  - Layer

Cogs:
151 concepts
17 properties
## UC2: Visualising XBRL Report Values

### Report Context


<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Greenhouse Gas Emissions By Weight Resulting From Commute in kgco2e</td>
<td>44399.86068376993</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Paper Usage Resulting From Printing in sheet Per FTE</td>
<td>269.055181196995</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Distance Resulting From Commute in km Per FTE</td>
<td>1675.12573821098</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Energy Consumption in kwh Per FTE</td>
<td>4517.37663287567</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Greenhouse Gas Emissions By Weight Resulting From Travel in kgCO2e Per FTE</td>
<td>3784.130765108943</td>
<td></td>
</tr>
<tr>
<td>Total Distance Resulting From Travel in km</td>
<td>882896.375</td>
<td></td>
</tr>
<tr>
<td>Total Greenhouse Gas Emissions By Weight Resulting From Energy Consumption in kgco2e</td>
<td>266461.2808</td>
<td></td>
</tr>
<tr>
<td>Total Energy Consumption in kwh</td>
<td>512425.64</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Paper Usage in sheet Per FTE</td>
<td>2120.452678873376</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Greenhouse Gas Emissions By Weight Resulting From Paper Usage in kgCO2e Per FTE</td>
<td>34.9674692014107</td>
<td></td>
</tr>
<tr>
<td>Total Greenhouse Gas Emissions By Weight Resulting From Printing in kgco2e</td>
<td>503.5122985839844</td>
<td></td>
</tr>
<tr>
<td>Total Distance Resulting From Commute in km</td>
<td>189991.3844122141</td>
<td></td>
</tr>
<tr>
<td>Total Greenhouse Gas Emissions By Weight Resulting From Paper Usage in kgco2e</td>
<td>3968.25</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Greenhouse Gas Emissions By Weight Resulting From Travel in kgCO2e</td>
<td>4291.93</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Greenhouse Gas Emissions By Weight Resulting From Printing in kgCO2e Per FTE</td>
<td>4.439329841281084</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Greenhouse Gas Emissions By Weight Resulting From Energy Consumption in kgCO2e</td>
<td>2349.349424989754</td>
<td></td>
</tr>
<tr>
<td>Total Paper Usage in sheet</td>
<td>240500</td>
<td></td>
</tr>
<tr>
<td>Average Per FTE Greenhouse Gas Emissions By Weight Resulting From Commute in kgCO2e</td>
<td>391.4669576567966</td>
<td></td>
</tr>
</tbody>
</table>
UC2: Visualisation of Provenance Descriptors

KPI Details & Provenance Information

KPI Name
Total Greenhouse Gas Emissions By Weight Resulting From Printing

Context URI

Unit
http://sustainable.deri.ie/measurementunits#kgco2e

Value
503.5122985839844

GRI KPI Compliance
EN16 - Total direct and indirect greenhouse gas emissions by weight.
UC2: Visualisation of Provenance Descriptors

4. Aggregation

1. Lookup
UC2: Visualisation of Provenance Descriptors

Open problems:

- More user-friendly provenance visualizations
- Integration in ETL tools (e.g., Pentaho Data Integration, Google Refine)

4. Aggregation

1. Lookup
Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- UC2: Understanding the Origin and Trustworthiness of XBRL
- Relation to DDI Lifecycle
Relation to DDI Lifecycle

Study (Can we compute GPM?)

SEC, FFIEC, Freebase

Add dimensions, unit conversion

Finding relevant reports

Business Intelligence platform

Data Mining, Visualisations, Link analyses with publications

Data Archiving

Report archived

Figure: Combined Life Cycle Model

Finding relevant reports

Add dimensions, unit conversion

Data Analysis

Data Discovery

Data Distribution

Data Processing

Data Collection

Study Concept
Conclusions

- XBRL interesting data source to be published as Linked Data for integration with other finance data.
- Open challenges in „mechanics“ and „analysis“ of Financial Linked Data.
- Preprocessing, matching, and understanding of data
- DDI interesting for XBRL use cases
Thank you for your attention, questions, and feedback

Study (Can we compute GPM?)

SEC, FFIEC, Freebase

Add dimensions, unit conversion

Finding relevant reports

Report archived

Business Intelligence platform

Data Mining, Visualisations, Link analyses with publications

SEC, FFIEC, Freebase

Add dimensions, unit conversion

Finding relevant reports

Study Concept

Data Collection

Data Processing

Data Distribution

Data Discovery

Data Analysis

Data Archiving

Repurposing

Figure: Combined Life Cycle Model
References

- [EDGAR RAY] http://edgarwrap.ontologycentral.com/archive/52827/0001193125-10-238973#ds
Backup: UC1: Why is QB such a good fit?

- Using RDF Data Cube Vocabulary (QB) for representing XBRL data

- Aggregate Data: 1) avg, count, sum 2) XBRL formulas/calculation arcs
- Numeric: financial facts
- Time series: valid start and end date of facts
- Multidimensional: filing, issuer, segment…
- Highly structured: concepts from taxonomies

http://www.w3.org/TR/vocab-data-cube/
UC2: Approach – Architecture with Provenance-aware ETL Applications

ETL Application

ETL Application

ETL Application

LightProv4J

Sustainability Data

<provURI>

OPMV + Cogs provenance descriptors

Provenance Store
Backup: Future Work Summary

- Further reduce “mechanics” needed
  - Creation of additional dimensions
  - Filtering for certain companies
  - Identification of identical companies in SEC and FFIEC

- Support more expressive “analyses”
  - Concept hierarchies
  - Formulas/business rules for missing values
  - Segment information (e.g., salary of CEO)
Relation to DDI Lifecycle and Use Cases

- “DDI For Use within a Research Project”
- Describe analysis goals and results
- Publish results of analyses (including queries) again as Linked Data and allow other users to enhance it, e.g., add annotations, add additional information to dimensions.

Figure: Combined Life Cycle Model
Relation to DDI Lifecycle and Use Cases (2)

- “Data Dissemination/Data Discovery”
- “Finding and Linking Publications related to Data”
- Link analyses with publications
- Many data sources are finance-relevant, the question is to find'
Relation to DDI Lifecycle and Use Cases (3)

- “Metadata Mining for Comparison, etc.”
- “Links to external thesauri”
- Linking comparable financial information
- Publishing taxonomy information and extensions
- Integration of “Web Tables”

Figure: Combined Life Cycle Model