
Linked Business and Financial Data on the Web

Dave Raggett <dsr@w3.org>
WWW2010, W3C LOD Camp

Current Situation

- Easy access to stock prices and news stories, but impoverished access to business and financial information in general
- Limited access to free sources of information
- High costs of pulling info into analysis tools
- Difficulties in combining different information sources
- Challenges with quality and provenance
- Effects institutional investors and everyone else

Opportunity

- The Web has the potential to offer greatly improved access to business and financial info
- Improved transparency of corporate, government and non-profit organizations
- Lower costs and easier combination of information sources
- Better flow of capital to where it's needed
- Search engines, social websites, corporate relations websites, investor websites,

Key Challenges

- Practices for entity naming and metadata
 - Holding companies, subsidiaries, mergers, splits,...
 - Essential for combining different information sources
- Persuading institutions to share the data freely
- Variations in corporate structures and accounting practices
 - US GAAP – rules based system of accounts
 - IFRS – principles based system of accounts
- Disruptive effect of move to the Web
 - Understanding the drivers is key for realizing the potential

XBRL

- Financial reporting language (XML+XLink)
 - Mandated by SEC and national reporting authorities world wide
 - Justified by improvements in internal processes
 - Unrealized promise for public consumption
 - Annual reports of publicly owned companies, mutual funds and potentially, government accounts
- Availability of reports filed in XBRL
 - Free from SEC website
 - Mostly non-free in other countries, but ...

XBRL Filings

- Facts, Contexts, Reporting Taxonomies
 - Instance document
 - Facts as numbers, currencies, dates, dimensions, footnotes
 - Concepts, relationships, labels
 - Taxonomies as semantically weak ontologies
 - Schema documents
 - Concepts & Relation Types
 - XML elements as instances of concepts
 - Linkbase documents
 - Defining relationships using XLink
 - Concept to concept
 - Concept to label or accounting literature
 - URIs for relationships and XML namespaces
 - Makes it practical to translate filings to RDF
 - Around 10 - 50 Mb including the taxonomy
 - For SEC filings

XBRL Instance fragment

```
<context id="ThreeMonthsEnded_30June2008_Unaudited">
  <entity>
    <identifier scheme="http://www.sec.gov/CIK">
      0000013610
    </identifier>
    <segment>
      <xbrldi:explicitMember
        dimension="us-gaap:StatementScenarioAxis">
        bne:Unaudited
      </xbrldi:explicitMember>
    </segment>
  </entity>
  <period>
    <startDate>2008-04-01</startDate>
    <endDate>2008-06-30</endDate>
  </period>
</context>
```

```
<us-gaap:RestructuringSettlementAndImpairmentProvisions
contextRef="ThreeMonthsEnded_30June2008_Unaudited"
unitRef="USD" decimals="-3">
  -17479000
</us-gaap:RestructuringSettlementAndImpairmentProvisions>
```

XBRL Schema fragment

```
<link:roleType roleURI="http://www.edgar-  
online.com/taxonomy/role/IMetrix_NotesToFinancialStatements"  
id="IMetrix_NotesToFinancialStatements">  
  <link:usedOn>link:calculationLink</link:usedOn>  
  <link:usedOn>link:presentationLink</link:usedOn>  
</link:roleType>
```

```
<link:linkbaseRef xlink:type="simple"  
xlink:arcrole="http://www.w3.org/1999/xlink/properties/linkbase"  
xlink:role="http://www.xbrl.org/2003/role/calculationLinkbaseRef"  
xlink:href="aa-20081231_cal.xml" xlink:title="Calculation Links, all"/>
```

```
<import namespace="http://xbrl.us/us-gaap-all/2008-03-31"  
schemaLocation="http://xbrl.us/us-gaap/1.0/elts/us-gaap-all-2008-03-  
31.xsd"/>
```

```
<element name="IncreaseDecreaseInNetAssetsHeldForSale"  
id="aa_IncreaseDecreaseInNetAssetsHeldForSale"  
type="xbrli:monetaryItemType" abstract="false"  
xbrli:periodType="duration" xbrli:balance="credit" nillable="true"  
substitutionGroup="xbrli:item"/>
```


XBRL Linkbase fragment

```
<labelLink xlink:type="extended"
xlink:role="http://www.xbrl.org/2003/role/link">
<loc xlink:type="locator"
xlink:href="http://xbrl.us/us-gaap/1.0/elts/us-gaap-2008-03-31.xsd#us-
gaap_AccountsPayable" xlink:label="us-gaap_AccountsPayable"/>
<label xlink:type="resource"
xlink:role="http://www.xbrl.org/2003/role/label" xlink:label="us-
gaap_AccountsPayable_lbl" xml:lang="en-US">Accounts payable,
trade</label>
<labelArc xlink:type="arc"
xlink:arcrole="http://www.xbrl.org/2003/arcrole/concept-label"
xlink:from="us-gaap_AccountsPayable" xlink:to="us-
gaap_AccountsPayable_lbl"/>
. . .
</labelLink>
```

Which just says use “*Accounts payable, trade*” as the US English label for the US GAAP reporting concept: AccountsPayable

XBRL and the Semantic Web

- XBRL is expensive to process and query
 - Memory and cpu intensive 1 second per 10Mb
 - LibXML2 + HTTP cache + hash indexing
- Preprocess to easier to manipulate formats
- Work on mapping filings to RDF and OWL
 - MUSING Project, Rhizomik Initiative, etc.
 - My own C code for mapping to RDF/Turtle
- Challenges for generating reports across filings
 - Non-additive taxonomy extensions
 - Changes and/or differences in reporting structures

XBRL in Turtle

```
_:context_FY07Q3
  xl:type xbrli:context;
  xbrli:entity [
    xbrli:identifier "0000789019";
    xbrli:scheme <http://sec.gov/CIK>;
  ];
  xbrli:period (
    [ xbrli:startDate "2007-01-01"^^xsd:date;
      xbrli:endDate "2007-03-31"^^xsd:date; ]
  ).

_:unit_usd xbrli:measure iso4217:USD.

_:fact209
  xl:type xbrli:fact;
  xl:provenance _:provenance1;
  rdf:type us-gaap:PaymentsToAcquireProductiveAssets;
  rdf:value "461000000"^^xsd:integer;
  xbrli:decimals "-6"^^xsd:integer;
  xbrli:unit _:unit_USD;
  xbrli:context _:context_FY07Q3.
```

XBRL in Turtle

Part of US GAAP taxonomy

```
@prefix usfr-pte: <http://www.xbrl.org/us/fr/common/pte/2005-02-28>.
```

```
usfr-pte:ChangeOtherCurrentAssets
  rdf:type xbrli:monetaryItemType;
  xbrli:periodType "duration".
usfr-pte:ChangeOtherCurrentLiabilities
  rdf:type xbrli:monetaryItemType;
  xbrli:periodType "duration".
```

```
_:link155 arcrole:parent-child [
  xl:type xl:link;
  xl:role role1:StatementFinancialPosition;
  xl:use "prohibited";
  xl:priority "1"^^xsd:integer;
  xl:order "1.0"^^xsd:decimal;
  xl:from usfr-pte:IntangibleAssetsNetAbstract;
  xl:to usfr-pte:IntangibleAssetsGoodwill;
].
```

Experiments with rdfproc

a) import filing to turtle*

```
xbrlimport http://example.com/filing.xml rdf.ttl
```

b) add to triple store

```
rdfproc -n store parse rdf.ttl turtle
```

c) query for facts, e.g. usfr-pte:NetIncome

```
rdfproc store query sparql - "PREFIX rdf:  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX usfr-pte:  
<http://www.xbrl.org/us/fr/common/pte/2005-02-28> PREFIX xl:  
<http://www.xbrl.org/2003/XLink> PREFIX xbrli:  
<http://www.xbrl.org/2003/instance> SELECT ?v WHERE {?f rdf:type  
usfr-pte:NetIncome . ?f xl:type xbrli:fact . ?f rdf:value ?v}"
```

* <http://sourceforge.net/projects/xbrlimport/>

Performance

- Find top-level concepts: 600 concepts with parent-child relationships (DAG)
 - Nodes that are parents but not children
 - Few milliseconds with C code
 - 1000 seconds with Sparql query
- Opportunities for optimization
 - Smarter engines that know about the ontology, the kinds of queries and the patterns of data in the triple store
 - $O(N^2)$ vs $O(N \log N)$
 - Fallback to lower-level APIs

```
SELECT ?p
WHERE {
  ?L1 arcrole:parent-child ?b1 .
  ?b1 xl:type xl:link .
  ?b1 xl:from ?p
  OPTIONAL {
    ?L2 arcrole:parent-child ?b2 .
    ?b2 xl:type xl:link .
    ?b2 xl:to ?p
  }
  FILTER (!BOUND(?b2))
}
```

Realizing the Potential

- Importing XBRL and other data sources into common stores for efficient query processing
- Supporting multiple levels of conceptual models
 - Bridging the gap between the raw data and the financial analytics and end-user presentations
 - “Recipes” rather than formal logic
 - Uploadable scripts as financial apps
 - Export data or presentations
- Use by search engines, investor sites, SNS, etc.
- Chicken and egg problem for funding this ...