

How can we exploit XBRL and Semantic Web technologies to realize the opportunities?

Track 3: Case Studies in XBRL Solutions

19th XBRL International Conference, Paris, 23rd June 2009

Dave Raggett <dsr@w3.org>

W3C Fellow – Financial data & Semantic Web

Member – XBRL INT Technical Standards Board

With some slides from Diane Mueller, JustSystems

Outline

- ▶ XBRL: adding semantics to business reports
- ▶ World Wide Adoption of XBRL
- ▶ Users and use cases for XBRL
- ▶ Realizing the potential
- ▶ Feeding the Semantic Web
 - XBRL, XLink, RDF, Turtle, SPARQL, OWL
 - Web APIs, Smart Searches, Web Scale Queries
 - Findings

So What is XBRL?

- ▶ **eXtensible Business Reporting Language**
 - a freely available electronic language for financial reporting.
 - based on XML, XML Schema and XLink
 - based on accepted financial reporting standards and practices to transport financial reports across all software, platforms and technologies
- ▶ Business reporting includes, but is not limited to:
 - financial statements,
 - financial and non-financial information
 - general ledger transactions
 - regulatory filings such as annual and quarterly financial statements.

“XBRL allows software vendors, programmers and end users who adopt it as **a specification** to enhance the creation, exchange, and comparison of business reporting information” *from xbrl.org*

Not just a number

- ▶ XBRL binds each reported fact to a concept in a reporting taxonomy e.g. US GAAP, IFRS
 - Each concept can be bound to a description and its definition in the accounting literature

Hierarchy of concepts

Terse label, EN

Currency, amount

Reporting period

Impairment of goodwill: \$M21 3 months to 2009-04-30

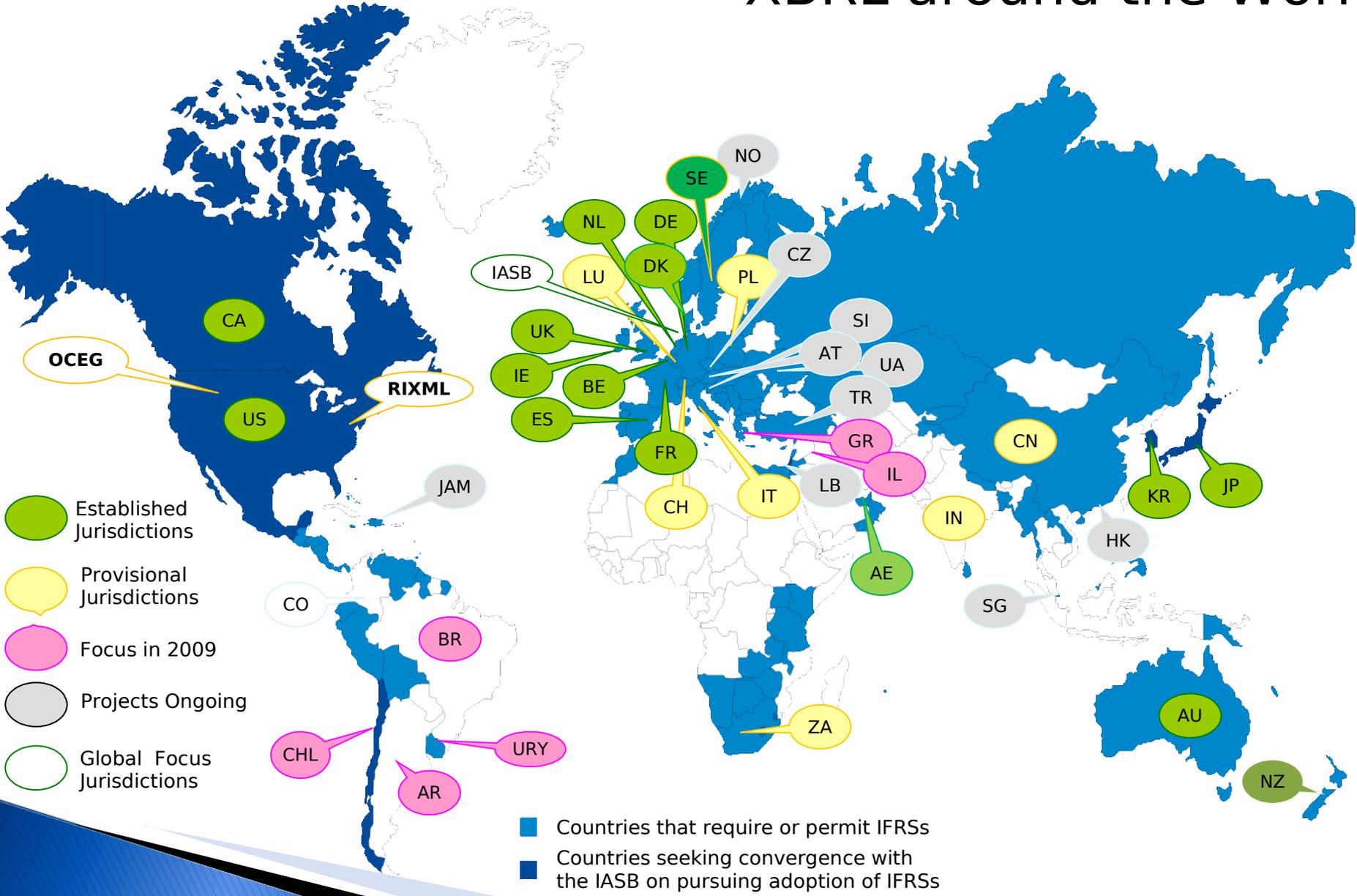
Description

Impairment of goodwill: Loss recognized during the period that results from the write-down of goodwill after comparing the implied fair value of reporting unit goodwill with the carrying amount of that goodwill. Goodwill is assessed at least annually for impairment.

Reference

Publisher: FASB; Name: Statement of Financial Accounting Standard (FAS); Number: 142; Paragraph: 47; Subparagraph: b; 2

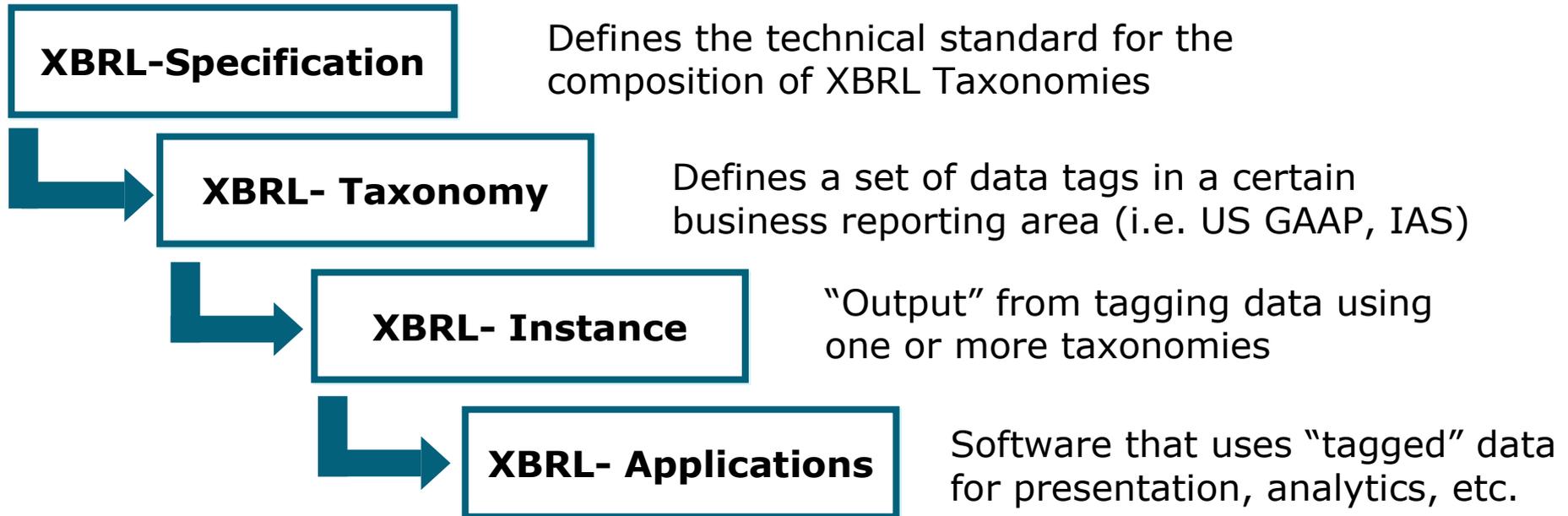
XBRL around the World



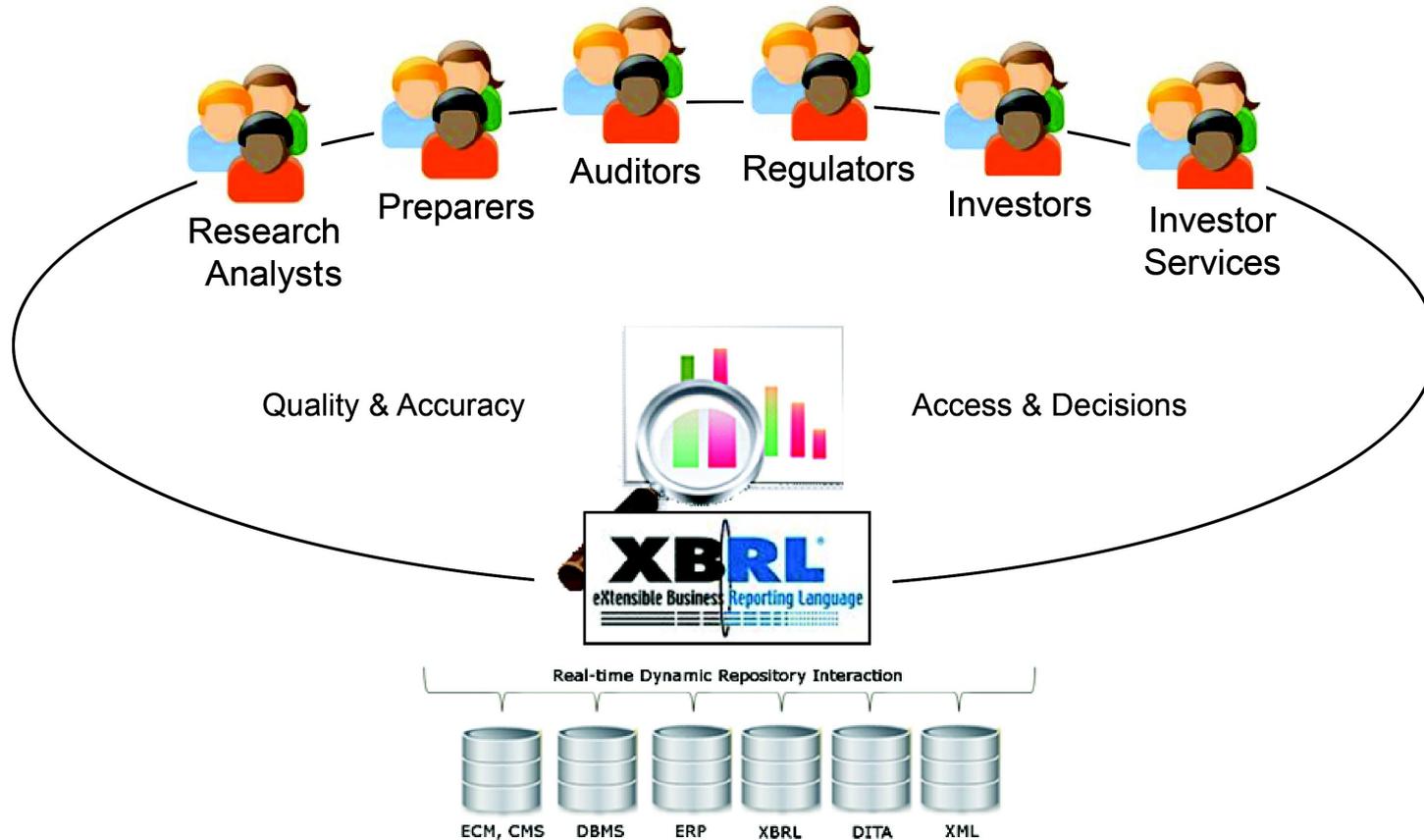
Purpose of XBRL

- ▶ XBRL provides users with a standard format in which to **prepare reports** that can subsequently be presented in a variety of ways.
- ▶ XBRL provides users with a standard format in which information can **be exchanged** between different software applications.
 - Integration of existing/legacy systems
- ▶ XBRL permits the automated, efficient and reliable **extraction of information** by software applications.
- ▶ XBRL facilitates the **automated comparison** of financial and other business information, accounting policies, notes to financial statements between companies, and other items about which users may wish make comparisons that today are performed manually.

XBRL Structural Overview



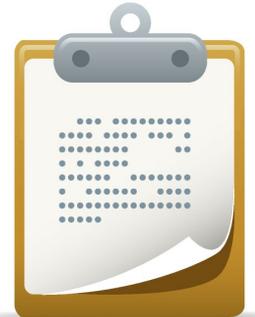
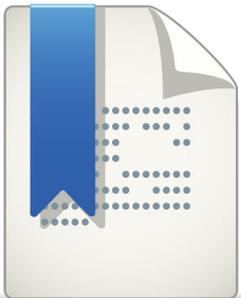
Who Benefits from XBRL?



Numbers of Users

▶ US Census

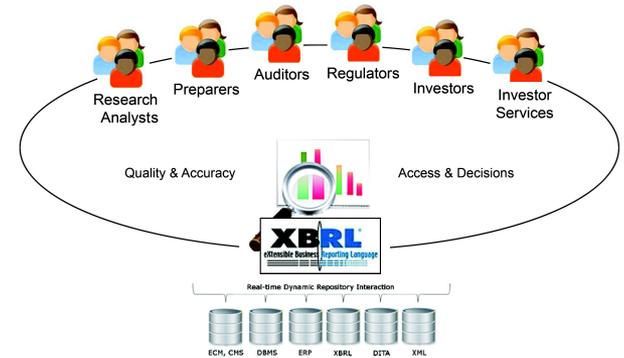
- Number of Accountants & Auditors: 1.3 million
- Number of Financial Analysts: 800K
- Number of Financial Journalists: 70K
- Number of Individual Investors: >40 million ⁽²⁾



<http://www.bls.gov/oco/ocos259.htm>

<http://pages.stern.nyu.edu> (as 1995 ²)
June 2009

Who Benefits from XBRL



- ▶ Four categories of users:
 - business information preparers,
 - Intermediaries in the preparation and distribution process,
 - users of this information and
 - the vendors who supply software and services to one or more of these three types of user..
- ▶ A major goal of XBRL is to add semantic precision to financial reports.
- ▶ It facilitates current practice; it does not change or set new accounting or other business domain standards.

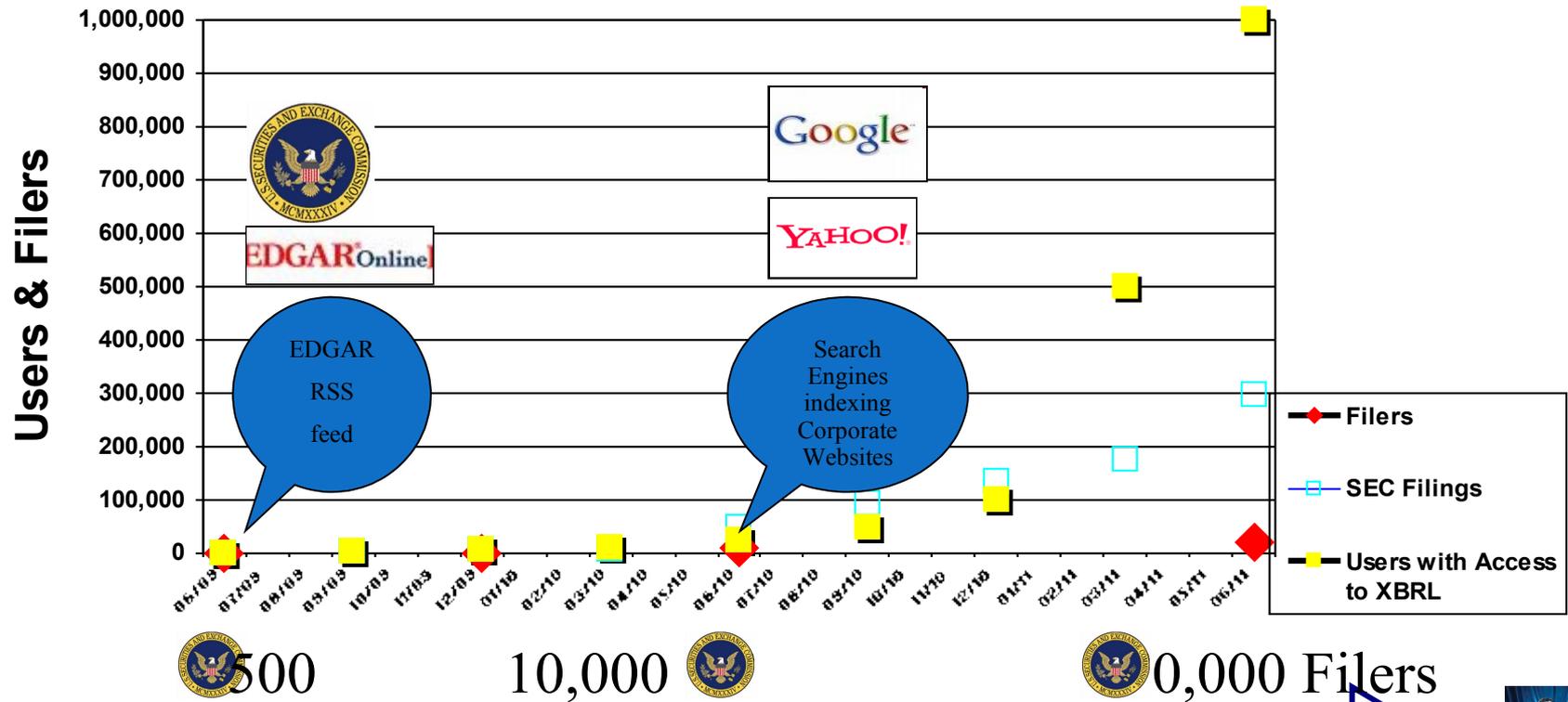
Impact

*The SEC mandate means that US companies will soon be producing a **tidal wave** of financial data in a machine interpretable format*

This is a a paradigm shifting event!

Adoption Curve

Influences: Access to & Amount of Data Available Publicly

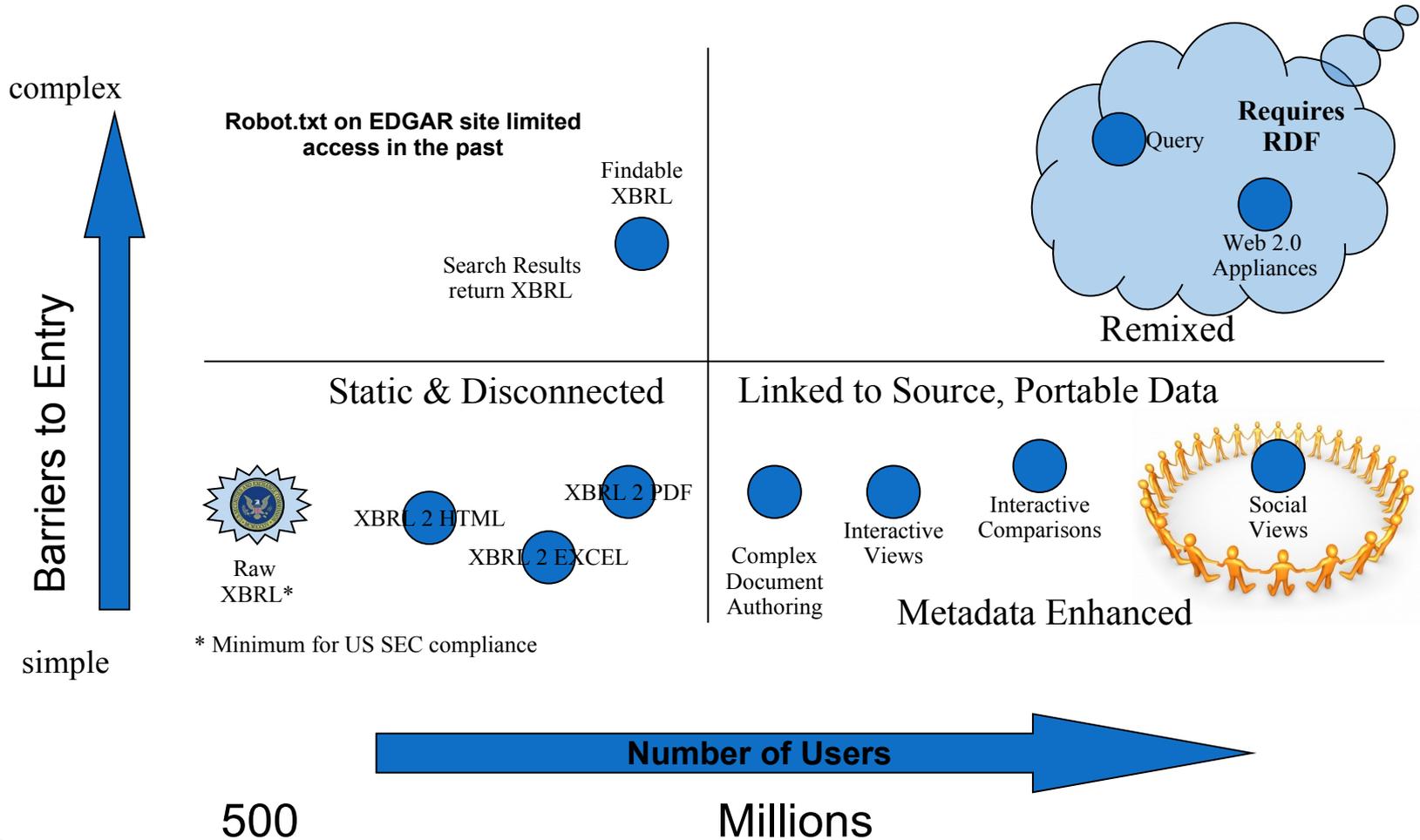


SEC 3 Year Mandate Roll out

* and this doesn't include e-gov content!



XBRL Consumption Technology Quadrants



Key Drivers

- ▶ As more XBRL content becomes available
 - More users start re-using it in their documents
- ▶ Historical XBRL Data available from EDGAR-Online
- ▶ US SEC data
 - Level 1 feeds (real-time) \$
 - New Mandate forces Corporations to host their XBRL filings publicly on Investor Relations Websites (free)
- ▶ Search Engines are currently blocked from indexing US SEC website
 - New Mandate makes financial content accessible to Search
 - XBRL Content is freely available and high quality

Exploiting XBRL

- ▶ How can you exploit reports filed in XBRL?
 - XBRL viewers and analysis tools
 - XBRL processing suites
 - Code your own analytic solution
 - Apply Semantic Web technologies and combine with other sources of information
 - 1) convert to RDF and OWL
 - 2) apply SPARQL and RIF
 - 3) define higher level query API
 - via server-side scripts

Some XBRL Viewers

- ▶ Web-based
 - SEC voluntary filing viewer and updated previewer
 - Rivet's *Dragon View*
 - CompSci's *iA*
 - *My experimental viewer*
- ▶ Standalone
 - JustSystem's *xfy XBRL Report*
 - Reporting Standard S.L. *XBRL Report Viewer*
 - Altova's *Missionkit* with *StyleVision*
(XML editor/viewer)



Activity <<

View Filings

Search Name or Ticker Symbol:

PAGE UP

- [-] American Strategic Income Portfolio Inc II
- [-] American Strategic Income Portfolio Inc III
- [-] Anadarko Petroleum Corp
- [-] ASML Holding Nv
- [-] Autodesk Inc
 - Annual Report (2009-01-31)**
 - Quarterly Report (2008-10-31)
 - Quarterly Report (2008-07-31)
 - Quarterly Report (2008-04-30)
 - Annual Report (2008-01-31)
 - Quarterly Report (2007-10-31)
 - Quarterly Report (2007-07-31)
 - Quarterly Report (2007-04-30)
 - Annual Report (2007-01-31)
- [-] Yearly Reports

PAGE DOWN

Company Comparison Report

Welcome !! ADSK 2009-01-31

Close All Tabs | Help

[Statement Of Income](#) | [Statement Of Financial Position Classified](#) | [Document Information](#) | [Entity Information](#) | [Notes to Financial Statements](#) | [Statement Of Cash Flows Indirect](#) | [Statement Of Shareholders Equity And Other Comprehensive Income](#) | [All Reports](#)

Autodesk Inc

[Company Summary](#) | [Filing Summary](#) | [Charting](#) | [Print Report](#) | [Export to Excel](#) | [SEC XBRL Filing](#)

Annual Report (2009-01-31)

> Freeze Headers

Statement Of Income (USD \$) (in Millions, except per share data)	12 Months Ended		
	Jan. 31, 2009	Jan. 31, 2008	Jan. 31, 2007
License and other	1,603.4	1,618.6	1,415.9
Maintenance	711.8	553.3	423.9
Total net revenue	2,315.2	2,171.9	1,839.8
Cost of license and other revenue	210.2	199.1	209
Cost of maintenance revenue	8.9	8.6	8.7
Total cost of revenue	219.1	207.7	217.7
Gross profit	2,096.1	1,964.2	1,622.1
Marketing and sales	900.7	847.7	700.4
Research and development	576.1	490.5	410.6
General and administrative	205.7	180.4	161.4
Impairment of goodwill and intangibles	128.9	0	0
Restructuring	40.2	0	0
Total operating expenses	1,851.6	1,518.6	1,272.4
Income from operations	244.5	445.6	349.7
Interest and other income, net	8	24.4	16.8
Income before income taxes	252.5	470	366.5
Provision for income taxes	(68.9)	(113.8)	(76.8)
Net income	183.6	356.2	289.7
Basic net income per share	0.81	1.55	1.26
Diluted net income per share	0.8	1.47	1.19
Shares used in computing basic net income per share	225.5	230.3	230.7
Shares used in computing diluted net income per share	230.1	242	243.2

The XBRL financial data on which this web site operates were submitted to the SEC on an "unaudited" or "unreviewed" basis primarily for the purpose of testing the XBRL format and technology, and do not constitute official SEC filings. Accordingly, users of this web site should not rely on the XBRL data or documents rendered by the web site in making investment decisions.

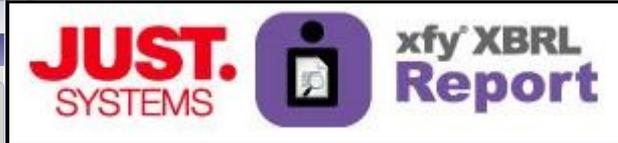
xfy Personal Client (New Document)(No Title)(*)

File Edit View Insert Format Data Bookmarks Tools Help

foo :=XPATHEX("xbrl-fnc:monetary(SalesRevenueNet',eol_5248_TTM_p12m_20081231_4','-6','actual','nullable','http://www.xbrl.org/us/fr/commo

	Jan. 1, 2008 Dec. 31, 2008	Jan. 1, 2007 Dec. 31, 2007	Jan. 1, 2006 Dec. 31, 2006
Income Statement			
Revenue			
Sales Revenue			
Sales Revenue, Net	25,269	24,462	22,923
Revenue	25,269	24,462	22,923
Cost of Goods and Services Sold			
Cost of Goods and Services Sold	13,379	12,735	11,713
Gross Profit	11,890		
Operating Expenses			
Selling, General and Administrative Expenses			
Selling, General and Administrative Expenses	5,245		
Research and Development Expense			
Research Development and Engineering Expense			
Research and Development Expense (Excluding In - Process)	1,404		
Research Development and Engineering Expense - Total	1,404		
Research and Development Expense - Total	1,404		
Operating Expenses excluding costs - Total	6,649		
Cost and Operating Expenses	20,051		
Operating Income/(Loss)	5,218		
Nonoperating Income/(Expense)			
Interest Income/(Expense)			
Interest Expense			
Interest Expense	215		
Interest Income	105		
Interest Income/(Expense), Net	(110)		
Nonoperating Gains/(Losses)			
Gain/(Loss) on Sale of Business	(23)		
Nonoperating Gains/(Losses)	(23)		
Nonoperating Income/(Expense)	(110)		
Income/(Loss) from Continuing Operations Before Income Taxes	5,108		
Income Taxes			
Income Loss Continuing Operations Before Income Taxes Minority Interest	5,108		
Provision for Income Taxes			

[Income Statement'.F6]



UBS Investment Research
United Technologies Corp

Reports Solid Third Quarter Earnings

- UTR reported Q3 EPS of \$0.99 (cons. \$0.96)
- UTR reported Q3 EPS of \$0.99, which includes a \$60 million pre-tax gain from the sale of a partnership interest and \$93 million in restructuring charges. Revenues grew 12% to \$12.2 billion, with organic growth of 8%. FCF was \$1.1 billion (118% of net income) in Q3.
- UTR raised its 2008 EPS guidance range to \$3.65-\$3.69
- UTR raised its 2008 EPS guidance to \$3.65-\$3.69 (cons. \$3.66), from \$3.55-\$3.65. Management's revised full year 2008 guidance implies a Q4 earnings forecast range of \$0.81-\$0.85 per share (cons. \$0.85). UTR expects full year FCF to equal or exceed net income. For 2007, UTR expects EPS growth of 10%, which implies EPS of \$4.02 at a minimum (cons. \$4.13).
- We are raising our 2008 and 2007 EPS estimates
- We are raising our 2008 EPS estimate to \$3.68 from \$3.65, primarily to reflect better-than-expected third quarter results. We are raising our 2007 EPS estimate to \$4.05 from \$4.03, primarily to reflect higher base pay in 2006.
- Valuation: We are raising our target price to \$90 per share
- We are raising our target price to \$90 per share (was \$85) to reflect our revised 2007 EPS estimate and a higher market multiple. Our revised target price reflects a 6%-10% premium (upgraded) to the market multiple on our revised 2007 earnings estimate, and excess cash of roughly \$3 per share. We maintain our Neutral rating on UTR.

Income Statement

	2008	2008	2008	2008	2008	2008
	12M	9M	6M	3M	Q3	Q3
Revenue	12,200	12,200	12,200	12,200	12,200	12,200
Operating Income	5,218	5,218	5,218	5,218	5,218	5,218
Operating Expenses	20,051	20,051	20,051	20,051	20,051	20,051
Operating Income/(Loss)	5,218	5,218	5,218	5,218	5,218	5,218

Please input the title of this document.

xy Personal Client JustInvestmentResearch3M.xml(+)

File Edit View Insert Format Table Bookmarks Tools Help

xHTML

JUST Investment Research

3M COMPANY : Large Accelerated Filer CIK CODE: 1

Competitive Information

Financial Ratios	0000088740	000030554
	3M Company Jan. 1, 2009 Mar. 31, 2009	Dupont Jan. 1, 2009 Mar. 31, 2009
Ratios		
Current Valuation Ratios		
Market Capitalization - Basic	34,480,820,000	20,179,703,704
Market Capitalization - Diluted	34,600,148,000	20,179,703,704
Earnings Per Share - Basic	1	1
Earnings Per Share - Diluted	1	1
Price / Earnings	11.56	15.62
Price to Sales	1.45	0.67
Dividend	2	2
Dividend Yield	0.0408	0.0743
Dividend Payout	0.4767	1.1000
Price to Free Cash Flow	24.6440	(114.0096)
Profitability Ratios		
EBITDA	1,074,000,000	1,254,000,000
Gross Margin	0.4937	0.2315
Operating Margin	0.1892	0.0683
Pre-Tax Margin	0.1838	0.0551
Profit Margin	0.1251	0.0430
EBIT	803,000,000	855,000,000

Market Capitalization

Findings: Reports Solid First Quarter Earnings

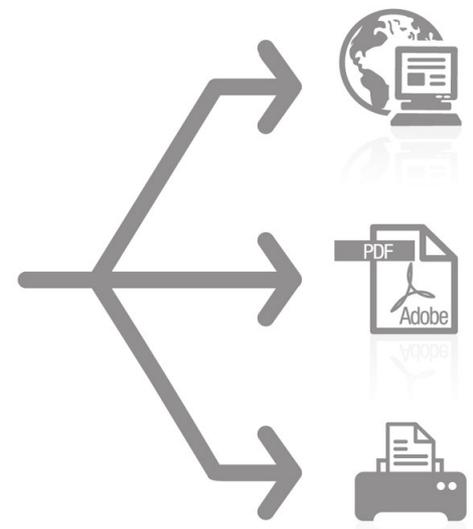
- MMM reported Q1 EPS of \$
- MMM reported Q3 EPS of \$0.99, which includes a \$60 million pre-tax gain from the sale of a partnership interest and \$93 million in restructuring charges. Revenues grew 12% to \$12.2 billion, with organic growth of 8%. FCF was \$1.1 billion (118% of net income) in Q3.
- MMM raised its 2008 EPS guidance range to \$3.65-\$3.69
- MMM raised its 2008 EPS guidance to \$3.65-\$3.69 (cons. \$3.66), from \$3.55-\$3.65. Management's revised full year 2008 guidance implies a Q4 earnings forecast range of \$0.81-\$0.85 per share (cons. \$0.85). MMM expects full year FCF to equal or exceed net income. For 2007, MMM expects EPS growth of 10%, which implies EPS of \$4.02 at a minimum (cons. \$4.13).

Basic Earnings Per Share Details

	Jan. 1, 2008 Dec. 31, 2008	Jan. 1, 2007 Dec. 31, 2007
Income Statement		
Basic Earnings Per Share Details		
Basic Earnings Per Share	4.95	5.70
Weighted-Average Shares Outstanding - Basic	664,200,000	748,300,000

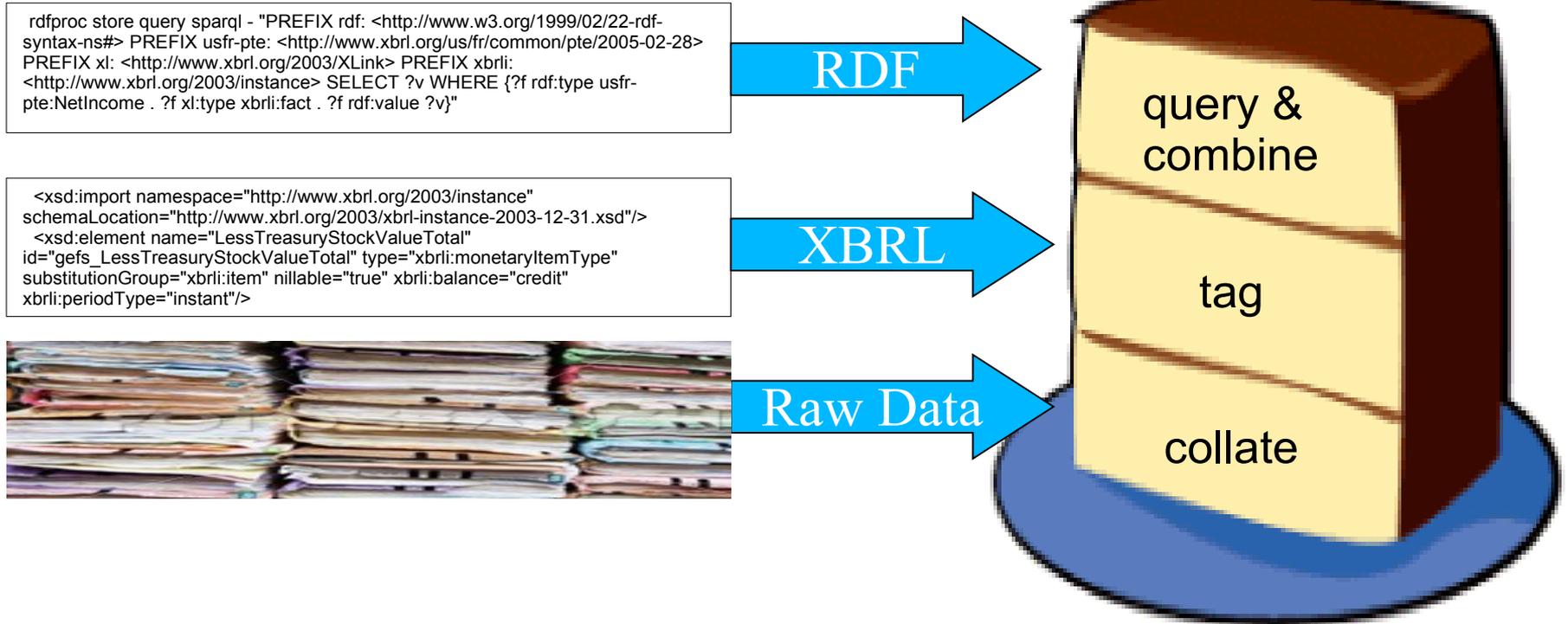
/html/body/div[3]/ul[1]/li[2] 68%

<http://na.justsystems.com/xyfXBRLReport>



Feeding the Semantic Web

The Semantic Web as a Layer Cake



How do we get there from here?

- ▶ Access to Raw Data - Recovery.gov, SEC.gov
- ▶ XBRL
- ▶ XBRL to RDF
- ▶ Semantic Web

Some more technical observations

Each US SEC XBRL filing consists of:

- ▶ Instance file with reported facts
 - ▶ Numeric and textual facts
 - ▶ Dates and periods
 - ▶ Footnotes
 - ▶ Reporting dimensions
- ▶ Schema for the instance file
 - ▶ Definitions of markup elements
 - ▶ References to the reporting taxonomy
- ▶ Taxonomy extensions
 - ▶ Labels for reporting concepts
 - ▶ Currencies, e.g. USD, EUR
 - ▶ Additional reporting concepts
 - ▶ Relationships between these concepts
- ▶ Around 10 to 50 MB including the taxonomy!

Xlink is your friend!

- ▶ XBRL was conceived by Charlie Hoffman in 1998
 - ▶ Before the Semantic Web took off
- ▶ Standardized outside of W3C by a new organization: XBRL International
- ▶ Makes heavy use of XML technologies
 - ▶ XML Schema
 - ▶ Xpointer
 - ▶ Xlink
- ▶ Designed for machine to machine transfer
- ▶ Inscrutable to human eyes
- ▶ But heavily dependent on URIs
 - ▶ Documents and namespaces

XBRL Instance snippet

```
<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 snippet

```
<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 snippet

```
<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 Quirks

- ▶ Mostly due to how XBRL is currently being used
 - ▶ Instance files
 - ▶ Taxonomy design
- ▶ Entangled hierarchies
 - ▶ To disentangle parent-child hierarchies for different tables, you also need to check that incoming and outgoing links have same role type
 - ▶ Similar problem for calculation links
- ▶ Paucity of presentation concepts and their relationships
 - ▶ Bad news for interactive access over the Web
- ▶ Reliance of Xlink for everything
 - ▶ e.g. taxonomies, versioning and formulae

Rendering XBRL

- ▶ Companies submit a rendered report in parallel with the XBRL source files
 - ▶ HTML and EDGAR filing process
- ▶ *New:* InlineXBRL, essentially XHTML with embedded XBRL semantics
 - ▶ Use XSLT to extract XBRL instance file
- ▶ No standard way to render XBRL filings
 - ▶ Reliance on rules of thumb
 - ▶ Text blocks and embedded HTML
 - ▶ Complex filings with lots of tables and heavy use of dimensions
- ▶ Is likely to improve in response to experience
 - ▶ Changes to standards and practices

Examples of problems

- ▶ Text blocks
 - Escaped HTML or plain text with/without formatting
 - Often contain embedded tables
 - Style elements can mess up whole report
- ▶ No relative ordering of tables
 - Need for additional presentation link roles
- ▶ Large text blocks within tables
 - Can be broken out of table for easier reading
 - But suffers from same paucity of ordering info

Suggested ways forward

- ▶ Better practices for filings
 - Avoid large text blocks and preformatted text
 - Only embed markup in text blocks when the taxonomy doesn't cover the associated semantics
 - Ensure such markup is valid! [no style element]
- ▶ Better practices for taxonomies
 - Extend taxonomy to fill the gaps, e.g. for executive compensation
 - Use genuine hierarchies for presentation and calculation linkbases [vis StatementTable]
- ▶ New ways to express dependencies
 - Between concepts and between facts

Why it matters and how to fix it

- ▶ Sloppy practices make it harder to realize the potential advantages of XBRL for investors and other consumers of business and financial data!
- ▶ Look after your interests by lobbying
 - Software vendors for better tools
 - Taxonomy developers for better coverage
 - XBRL standards committees to refine specifications
- ▶ Support XBRL International in meeting your needs
 - Best Practices Board
 - XBRL Standards Board
 - Working Groups

Semantic Web

*Translating XBRL into
RDF and OWL*

Why Import XBRL?

- ▶ It is very expensive to process 10-50MB of XML on each query
 - ▶ Memory and CPU intensive: about 1 second of CPU time per 10MB of XML source
- ▶ Better to pre-process filings into a persistent format designed to match needs of queries
 - ▶ Current tools use proprietary solutions
- ▶ RDF and OWL as natural choices
 - ▶ Mature standards
 - ▶ Facilitate mashing financial data with other kinds of information available over the Web
 - ▶ Web APIs and standards would enable an ecosystem of value adding players

Some Existing Work

- ▶ EU MUSING Project
 - ▶ <http://www.musing.eu/>
- ▶ Rhizomik Initiative
 - ▶ <http://rhizomik.net/semanticxbrl/>
 - ▶ XSD2OWL followed by XML2 RDF
- ▶ Dave Raggett
 - ▶ <http://sourceforge.net/projects/xbrlimport>
- ▶ This talk will focus on the latter
 - ▶ Open source translator of XBRL filings into RDF
 - ▶ Translates complete filing including taxonomy
 - ▶ Output as RDF in Turtle syntax

XBRL as Turtle

Use of blank nodes for concepts with many properties

```
_: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 as 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

- ▶ Rdfproc is a widely available command line tool

- ▶ Part of Dave Beckett's Redland suite

- ▶ Step 1 create the turtle for a given XBRL filing

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

- ▶ Step 2 use turtle to populate a triple store

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

- ▶ Step 3 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}"`

Experiments with rdfproc

With the following results:

- ▶ result: [v=937600000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=995300000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=937600000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=995300000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=483200000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=518900000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=507200000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=538200000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=507200000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=538200000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=243700000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ result: [v=216600000^^<http://www.w3.org/2001/XMLSchemainteger>]
- ▶ rdfproc: Query returned 12 results

Further work

- ▶ SPARQL queries rapidly get very complicated to write by hand
- ▶ The data returned needs further work before it is ready to be presented to the user
- ▶ Currently exploring this with a C program linked against the Redland and Rasqal libraries
 - ▶ Rendering even a single XBRL table takes many SPARQL queries
 - ▶ Lots of massaging of intermediate results
 - ▶ Followed by rendering to HTML or SVG
- ▶ Concerns over performance
 - ▶ Compared to experience with native C program

Some initial results

- ▶ Consider task of finding top-level concepts in a DAG of parent-child links of 600 bnodes
 - ▶ Nodes that are parents but not children
 - ▶ A few milliseconds with custom C code
 - ▶ Sparql query
 - ▶ takes 1000 seconds
 - ▶ $O(N^2)$ vs $O(N \log N)$
- ▶ Opportunities for optimisation
 - ▶ Smarter Sparql engines
 - ▶ Use of lower level API
 - ▶ Object-based approach
 - ▶ Persistent heaps

```
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))
}
```

XBRL and OWL

- ▶ XBRL Taxonomy loosely equates to OWL ontology
 - ▶ But note XBRL's taxonomy overrides
- ▶ Automated mapping is mostly feasible
 - ▶ As demonstrated by Rhizomik XSD2OWL
- ▶ XBRL's formal semantics are weak
 - ▶ OWL based on Description Logics
- ▶ XBRL versioning standard will describe differences between different versions of the same taxonomy, e.g. US GAAP 2008, 2009
 - ▶ Unaware of work on mapping this into OWL
- ▶ Reasoning across different taxonomies remains a major challenge
 - ▶ e.g. US GAAP vs IFRS

XBRL and Rule Languages

- ▶ Can be used to express analytic techniques as rules that are easier to maintain than code
 - Expert systems that spot unusual patterns
- ▶ Goal driven, top-down execution
- ▶ Data driven, bottom-up execution
- ▶ Combine with scripting as needed
 - Extension predicates
- ▶ RIF as W3C rule interchange format

What is XBRL good for?

- ▶ Ensuring Valid Data, Comparability, and Consistency
- ▶ So far virtually all the effort has gone into defining the taxonomies and tools for helping companies to prepare filings
- ▶ Driven by the SEC mandate
 - ▶ Experience gained with voluntary program
- ▶ As a flood of financial data in XBRL becomes available, work will ramp up on exploiting it
- ▶ Who would benefit?
 - ▶ Private investors with better access to information
 - ▶ Institutional investors through reduced costs
 - ▶ No more error prone re-keying of data
 - ▶ Businesses through easier access to investment
 - ▶ Greater transparency of risk to investors

Web APIs for Financial Data

- ▶ Support for an ecosystem of value-adding players
- ▶ First stage is data aggregators who pull XBRL from SEC and other sources and expose it as triples
 - ▶ Access to raw triples via SPARQL queries
 - ▶ Consumer uses scripts to add value
- ▶ High-level APIs for common queries, where the results are provided as charts or tables
 - ▶ For embedding in web pages
- ▶ Yet higher-level APIs for financial analytics that combine data from multiple filings
 - ▶ Complicated by variations across ontologies

Smart Search Engines

- ▶ Imagine search engines that provide selected financial highlights for each company that matches the search criteria you just entered
 - ▶ With salient numbers and charts
- ▶ The search results tailor the data provided according to your interests
 - ▶ Based upon analysis of the search criteria and other information gleaned from previous searches
 - ▶ Subject to your privacy preferences, of course!
 - **
- ▶ Interactive data you can drill down on

Web Scale Queries

- ▶ SPARQL & RDF offer generality but sacrifice speed for complex queries
- ▶ For predetermined models and queries a persistent object store can allow queries to execute at native C or Java speeds
 - ▶ Sub-second response times
- ▶ Use of cloud computing solutions for web scale performance
 - ▶ Executing a query across thousands of servers
 - ▶ Exploiting really large data sets
 - ▶ Changing the kinds of questions we can ask
- ▶ Dependent on ecosystem of players
 - ▶ Not a single algorithm unlike text-based search

Where next?

Next Steps

Short term

- ▶ Presentation @ Next week's XBRL Conference in Paris
- ▶ Plan to hold joint W3C/XBRL International workshops on interactive access to financial data
 - ▶ 28-29 September 2009, Washington DC, hosted by FDIC (Federal Deposit Insurance Corporation)
 - ▶ Another workshop is being planned for Europe in October
- ▶ Following the workshops,
 - ▶ W3C Interest Group on financial data
 - ▶ Focus on use cases and technical challenges
 - ▶ Identification of areas for further collaboration between W3C and XBRL International

Conclusions

- ▶ Out of the ashes come a new beginning
 - ▶ Financial scandals and crises stimulate new ideas
 - ▶ Greater transparency of risks and investment opportunities as drivers for economic growth
- ▶ National mandate for reporting in XBRL is creating opportunities for a new ecosystem for financial data
- ▶ XBRL is just one piece of this story
 - ▶ Others include FpML, RIXML, ...
- ▶ There are lots of challenges e.g. for reasoning across different taxonomies
- ▶ XBRL and cloud computing could radically transform how we think of the Semantic Web

Thanks for listening!

Questions?

Dave Raggett, W3C Fellow (JustSystems)
dsr@w3.org