XML: Data and Document Processing

XBRL

David Edmond
School of Information Systems
Queensland University of Technology

March 23, 2007

Introduction

Be afraid!

"Go to most information technology type people and say *debit* or *credit* and watch their hair stand on end!"

– Hitchhikers Guide to Understanding the IFRS-GP Taxonomy

XML flattened!

“All the information that would normally be encoded into the hierarchical structure of a standard XML document has been stripped out, and resides in an XBRL *taxonomy* . . . ”

– XML Flattened: The lessons to be learnt from XBRL

Enron: before and after

"Across corporate America, widespread corner cutting, steadily falling standards, and compromised financial discipline had been festering for close to a decade."

– Conspiracy of Fools

- Enron in Jan 2000
- Enron in Dec 2001
- Key people:
  - Kenneth Lay: Chairman and CEO
  - Jeffrey Skilling: CEO
  - Andrew Fastow: CFO
- Sarbanes-Oxley
- XBRL

Why XBRL? - the consultant's view

"Of all the decisions companies face, risk evaluation is perhaps the most fundamental. Yet, in the midst of an Internet-driven information revolution, most companies' risk-assessment processes are still *manual, data restricted and paper based*."

– How XBRL Web Services Impacts Credit and Risk Assessments

Why XBRL? - the regulator's view

"While electronic media have increased the accessibility of registrant information, that information is generally not available in a format that investors and other users who wish to perform technical data analyses can easily download and process using software applications or web services. In order to analyze financial information, these users of the information generally must either *copy data from financial documents into spreadsheets* or rely upon data that has been copied or otherwise *extracted and summarized by third-party sources*."

– Enhancing Commission Filings Through the Use of Tagged Data

XBRL

- XBRL is an “international standard for the communication of corporate financial and performance data”.
- It is XML-based, and makes extensive use of XML Schema and XLink.
- Version 1.0 of the XBRL specn. came out in July 2000. Version 2.0 in December 2001.
- The current specification of XBRL is version 2.1.
- This was originally released in December 2003, with the most recent update in November 2005.

– XBRL Website

But will it’s impact be, as suggested, “pervasive and positive”?

Example 1: Hello XBRL!

```
<xbrl
   xmlns="http://www.xbrl.org/2003/instance"
   xmlns:link="http://www.xbrl.org/2003/linkbase"
   xmlns:xlink="http://www.w3.org/1999/xlink"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xmlns:iso4217="http://www.xbrl.org/2003/iso4217"
   xmlns:extA="http://reallyuseful.info/ns/xbrl/extAssets">

 <link:schemaRef xlink:type="simple"
                 xlink:href="extAssets.xsd" />

 <context id="end2006">
  <entity>
   <identifier scheme="http://www.globalscot.co.uk/diaspora">
    David Edmond</identifier>
  </entity>
  <period>
   <instant>2006-12-31</instant>
  </period>
 </context>

 <unit id="OzDollars">
  <measure>iso4217:AUD</measure>
 </unit>

 <extA:car contextRef="end2006"
           unitRef="OzDollars" decimals="0">25000</extA:car>
</xbrl>
```

Arguments for using XBRL

- it uses XML, a widely supported computer standard for structured information;
- the document approach to data organisation that XML imposes is very well suited to retaining absolute integrity in the set of information;
- it uses XML in a way that ensures individual facts can be isolated from other facts in the source document without losing any information necessary to understand them;
- the XML representation of the data enables flexible querying of the source data for extracts that meet particular requirements; and, most importantly
- the XBRL format enables data capture to be extremely flexible.
- Complete financial reports are stored in XBRL format, *exactly as provided by customers*.

– Data management for credit risk - an XBRL equation

XLink

Links in HTML

Lots of HTML files contain hypertext links like the following:

```
1. <p>I am a student at <a href="www.qut.edu.au">QUT</a>.</p>
```

And many have references to images, written like the following:

```
1. <p>Here is my dog Winnie!<br/>
2.    <img src="winnie.jpg" alt="woof!"/></p>
```

- Both the <A/> and the <IMG/> elements may be considered as links from the document in which they appear to some external resource.
- The places to which they point may viewed as *remote* resources.
- The HTML file may be viewed as a *local* resource.

XLink

- XLink is the XML standard for describing links between resources.
- It is based on, but extends, the ideas of linking that appear in the `<A/>` and `<IMG/>` elements of HTML.
- Links can be attached to *any* element. This is because the links are all in the form of attributes.
- A *linkbase* is any XML document that contains collections of links.

✓ `http://www.w3.org/TR/xlink/`

Example 2: My QUT

Here is an idea of how the above anchor element might be recoded using XLink:

```
<p>I am a student at
   <span xmlns:xlink="http://www.w3.org/1999/xlink"
         xlink:type="simple"
         xlink:href="www.qut.edu.au"
         xlink:role="http://global.links.org/mainpage"
         xlink:title="QUT's home page"
         xlink:show="new"
         xlink:actuate="onRequest">
QUT</span>.</p>
```

- The link is specified by means of a number of dedicated attributes from the XLink namespace.
- The link is *simple* – it points directly from one resource to another.

Example 3: Woof Woof!

Write a simple link that replaces the HTML <IMG/> element above.

```
1. <p>Here is my dog <span
2.   xmlns:xlink="http://www.w3.org/1999/xlink"
3.   xlink:type="simple"
4.   xlink:href="..............."
5.   xlink:role="..............."
6.   xlink:title="..............."
7.   xlink:show="..............."
8.   xlink:actuate="...............">
9. Winnie</span>!</p>
```

Example 3: Woof Woof!

Write a simple link that replaces the HTML <IMG/> element above.

```
<p>Here is my dog <span
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xlink:href="winnie.jpg"
  xlink:role="http://global.links.org/pet/dog"
  xlink:title="David's pet poodle"
  xlink:show="embed"
  xlink:actuate="onLoad">
Winnie</span>!</p>
```

Components of XBRL

Basic framework

What you need to know to begin to understand XBRL:

- Instance
- Taxonomy Schema
- Linkbase
- Discoverable Taxonomy Set (DTS)

About instances

- An XML document encoded in line with some XBRL taxonomy is call an XBRL *instance*.
- It consists of a set of facts. A typical numeric fact connects:
  - A concept from that taxonomy, and
  - A number.
  - A context.
- A context consists of:
  - An entity of some kind.
  - A period of time.
  - A unit of measurement.
- As well as recording the basic facts to be reported, an instance must also contain a link to an XML Schema document.

About Schemas

A *taxonomy schema* defines a number of concepts from some domain:

- The schema takes the form of an XML Schema document.
- Each concept is associated with an XML Schema element definition.
- The corresponding element must be either an XBRL *item* or an XBRL *tuple*.
- Each element must be provided with a unique ID. This will be used for cross-referencing.

Role of LinkBases

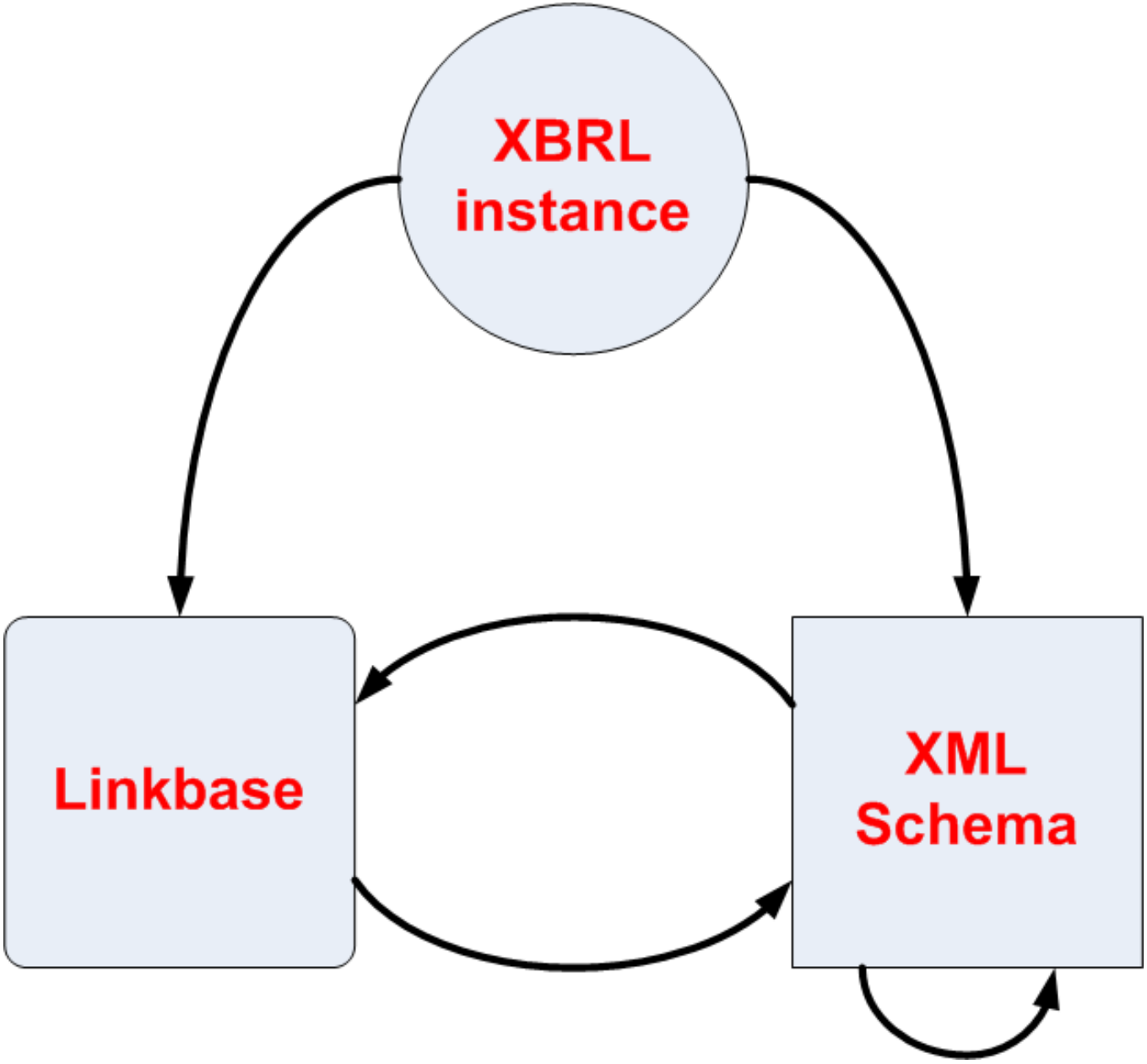
Five kinds of linkbase, divided into two categories, are suggested:

- For inter-concept relationships:
  - Presentation: how two concepts are related for presentation purposes.
  - Calculation: how two concepts are related for calculation purposes.
  - Definition: how two concepts are related for other purposes, e.g., man is a mammal.
- For relationships between concepts and their documentation:
  - Label: for short descriptions.
  - Reference: for authoritative sources.

The Discoverable Taxonomy Set

- The allowable content of a given instance is not governed by a single schema.
- A collection of documents is traversed and incorporated, according a prescribed set of rules.
- Discovery starts from the instance document.

A graphical view



A small example

Example 4: My worldly possessions

Consider this statement of somebody's assets:

House	300,000
Car	25,000
Cash	10,000
Total	335,000

In a conventional XML style, the data might be encoded as follows:

1. `<house>300000</house>`
2. `<car>25000</car>`
3. `<cash>10000</cash>`
4. `<total>335000</total>`

How can we express that the value of the `<total/>` element must equal the sum of the other three?

The instance

```
1.  <xbrl
2.   xmlns="http://www.xbrl.org/2003/instance"
3.   xmlns:link="http://www.xbrl.org/2003/linkbase"
4.   xmlns:xlink="http://www.w3.org/1999/xlink"
5.   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
6.   xmlns:iso4217="http://www.xbrl.org/2003/iso4217"
7.   xmlns:extA="http://reallyuseful.info/ns/xbrl/extAssets">
8.
9.   <link:schemaRef xlink:type="simple"
10.                  xlink:href="extAssets.xsd" />
11.
12.  <context id="end2006">entity and time involved</context>
13.
14.  <unit id="OzD"><measure>iso4217:AUD</measure></unit>
15.
16.  <extA:house contextRef="end2006"
17.         unitRef="OzD" decimals="0">300000</extA:house>
18.  <extA:car   contextRef="end2006"
19.         unitRef="OzD" decimals="0">25000</extA:car>
20.  <extA:cash  contextRef="end2006"
21.         unitRef="OzD" decimals="0">10000</extA:cash>
22.  <extA:total contextRef="end2006"
23.         unitRef="OzD" decimals="0">335000</extA:total>
24.
25.  </xbrl>
```

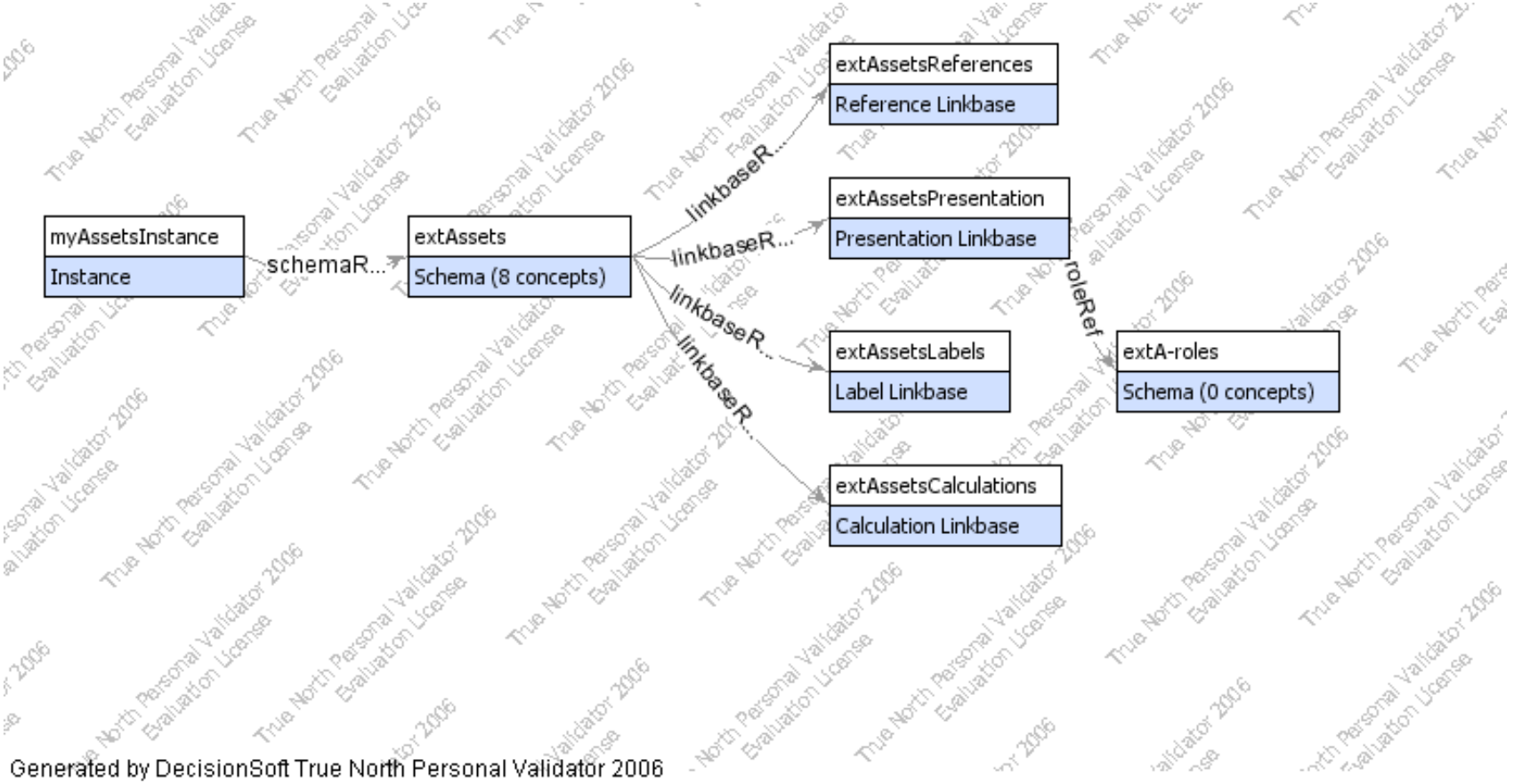
The concept of a ’house’

In the corresponding XBRL taxonomy, the *house* concept might be defined as follows:

```
<element
 id = "extA_house"
 name="house"
 type="xbrli:monetaryItemType"
 substitutionGroup="xbrli:item"
 xbrli:periodType="instant"
 nillable="true" />
```

All the other simple concepts (such as `car`, `cash` and `total`) would be defined similarly.

The ’extA’ Discoverable Taxonomy Set



The calculation linkbase

```
<linkbase
 xmlns="http://www.xbrl.org/2003/linkbase"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:xlink="http://www.w3.org/1999/xlink"
 xsi:schemaLocation
     ="http://www.xbrl.org/2003/linkbase
       http://www.xbrl.org/2003/xbrl-linkbase-2003-12-31.xsd">

<calculationLink
 xmlns="http://www.xbrl.org/2003/linkbase"
 xmlns:xlink="http://www.w3.org/1999/xlink"
 xlink:type="extended"
 xlink:role="http://www.xbrl.org/2003/role/link"
 xlink:title="Total asset calculation">

<!-- Part I:  All the concepts relevant to the
              calculation are identified and
              provided with labels.      -->

<!-- Part II: Then the concepts are linked in
              such a way as to specify the calculation
              total = house+car+cash                  -->

</calculationLink>
</linkbase>
```

Part I: Locate and label concepts

```
<loc
 xlink:type="locator"
 xlink:href="extAssets.xsd#extA_house"
 xlink:label="extA_house"/>

<loc
 xlink:type="locator"
 xlink:href="extAssets.xsd#extA_car"
 xlink:label="extA_car"/>

<loc
 xlink:type="locator"
 xlink:href="extAssets.xsd#extA_cash"
 xlink:label="extA_cash"/>

<loc
 xlink:type="locator"
 xlink:href="extAssets.xsd#extA_total"
 xlink:label="extA_total"/>
```

Part II: Form links that define a calculation

```
<calculationArc
  xlink:type="arc"
  xlink:arcrole
        ="http://www.xbrl.org/2003/arcrole/summation-item"
  xlink:from="extA_total" xlink:to="extA_house"
  weight="1.0"/>

<calculationArc
  xlink:type="arc"
  xlink:arcrole
        ="http://www.xbrl.org/2003/arcrole/summation-item"
  xlink:from="extA_total" xlink:to="extA_car"
  weight="1.0"/>

<calculationArc
  xlink:type="arc"
  xlink:arcrole
        ="http://www.xbrl.org/2003/arcrole/summation-item"
  xlink:from="extA_total" xlink:to="extA_cash"
  weight="1.0"/>
```

Continuing our example

Example 5: Down at the farm

Suppose that we add to our assets by buying a farm:

House	300,000
Car	25,000
Shed	15,000
Tractor	35,000
Cash	10,000
Total	385,000

Our XML representation might be as follows:

```
1. <house>300000</house>
2. <car>25000</car>
3. <shed>15000</shed>
4. <tractor>35000</tractor>
5. <cash>10000</cash>
6. <total>385000</total>
```

How can we validate this data when:

- In our taxonomy, there is no mention of either sheds or tractors – or any other farm-specific assets.
- Our calculation of the overall total makes no mention of them either.

Revised calculation

- The original calculation looks like this:

  total = house? + car? + cash?

  where the ? indicates an optional asset.

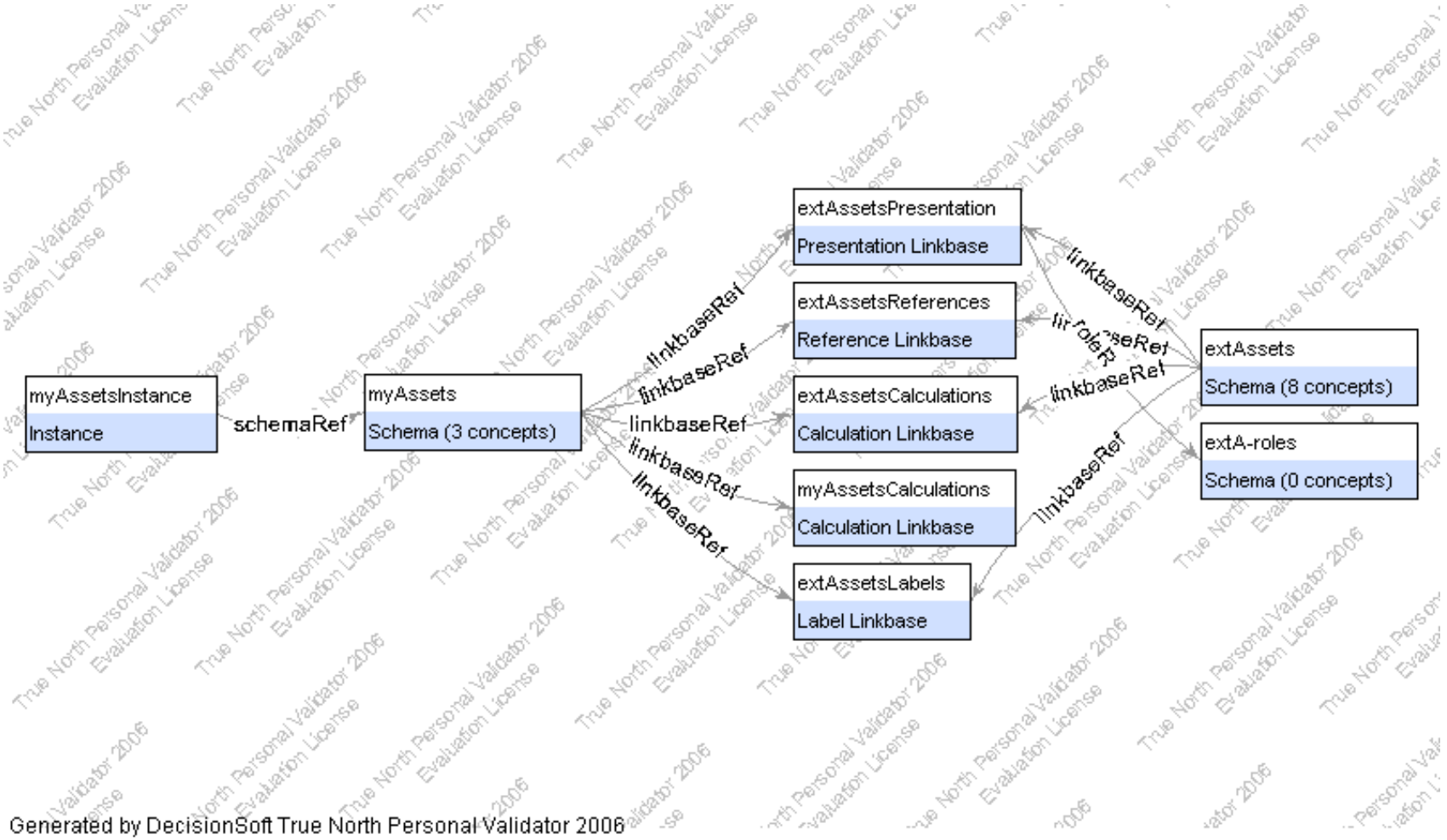
- We now want to extend it to:

  total = house? + car? + cash? + *shed? + tractor? + harvestor?*

The calculation extension

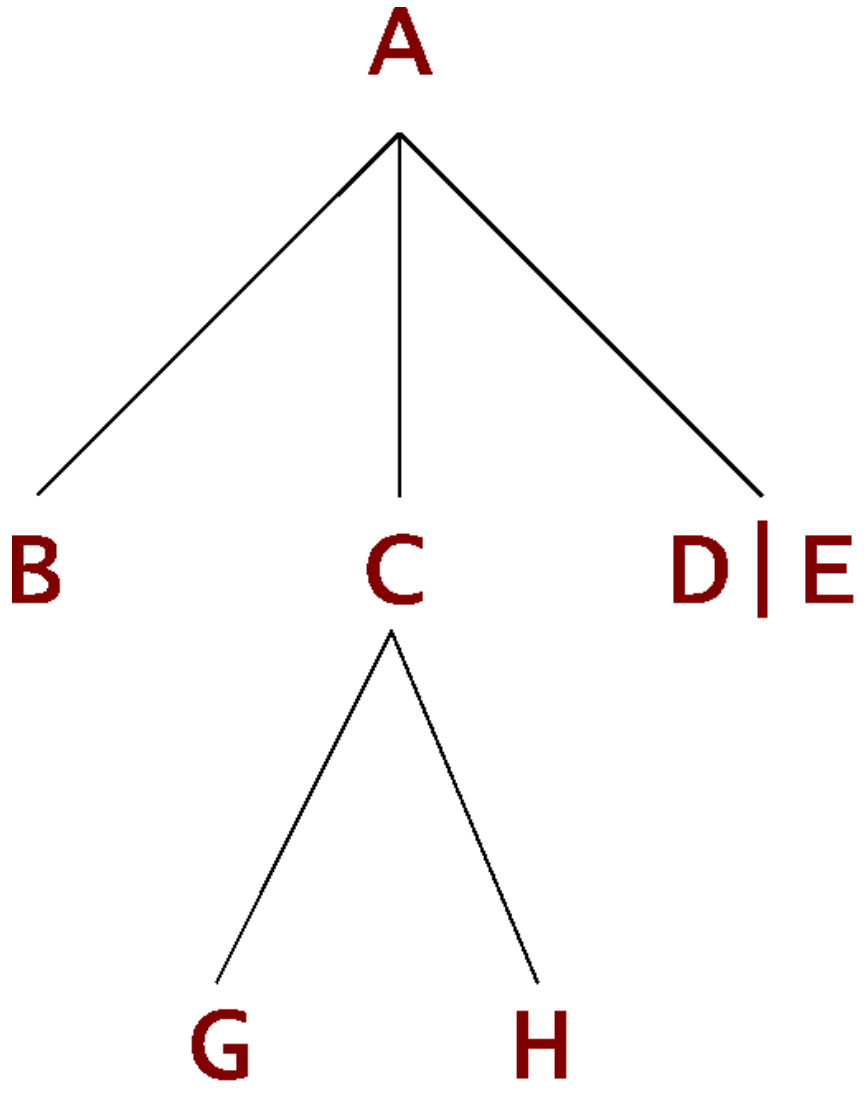
```
<calculationLink ... >

<loc
 xlink:type="locator"
 xlink:href="myAssets.xsd#myA_shed"
 xlink:label="myA_shed"/>

<locs ...  for tractor and harvester ... />

<loc
 xlink:type="locator"
 xlink:href="extAssets.xsd#extA_total"
 xlink:label="extA_total"/>

<calculationArc
  xlink:type="arc"
  xlink:arcrole
        ="http://www.xbrl.org/2003/arcrole/summation-item"
  xlink:from="extA_total" xlink:to="myA_shed"
  weight="1.0"/>

<calculationArcs ... for tractor and harvester ... />

</calculationLink>
```

The extension taxonomy



The ’X’ in XBRL

Presentation linkbases



A=(B C (D|E))
C=(G H)

A	B	1
A	C	2
A	D	3
A	E	3
C	G	1
C	H	2

Taxonomies

The IFRS-GP taxonomy

- IFRS is short for International Financial Reporting Standards.
- GP for General Purpose
- “The IFRS-GP taxonomy expresses concepts presented or disclosed in financial statements prepared . . . for profit-oriented entities”
- It is divided into 47 physical files.
- It contains 4,112 concepts:
  - From *Abandonment or withdrawal from plan previously reported as discontinued*.
  - To *Work performed by entity and capitalised*.

✓ `http://xbrl.iasb.org/int/fr/ifrs/gp/2005-05-15`

The APRA taxonomy

- The Australian Prudential Regulation Authority regulates the Australian financial services industry.
- As part of that oversight, it gathers statistics *directly* from relevant institutions.
- This information may be supplied in XBRL format. The 9539 concepts in this taxonomy include:
  - *Capital Adequacy Ratio*.
  - *Credit Risk Category - 20% Weight - Off Balance Sheet (Credit Equivalent Amount)*.
  - *Category IV claims on central banks, international banking agencies, regional development banks, ADIs in australia, and overseas banks - holdings of subordinated bonds issued by international banking agencies and multilateral regional development banks*.

– What is XBRL and what does it have to do with APRA?

The FFIEC Call Reports

This US regulatory authority is concerned with the contents of *Call Reports*:

- In the US, more than 8,000 banks are required to report quarterly on their financial condition.
- A *Call Report Modernization* project has been in operation since 2005.
- The intention is to “validate and release Call Report data faster.”
- “It is expected that the timeframe for releasing Call Report data will be shortened by a few weeks because of these changes.”

  – FFIEC: Central Data Repository Call Report Modernization

The outcome

The taxonomy includes:

- 3000 concepts
- 7000 business rules

The error rate has fallen from around 34% to nearly zero.

– XBRL Regulatory Reporting: Faster, Cheaper, & Better

The HMRC Computation Taxonomy

In the UK, from 2010, all company tax returns will be required to be in XBRL.

– Bringing XBRL Tax Filing to the UK

There are 4292 concepts in the taxonomy including:

Calculation Of CT Liability Financial Year 1 Chargeable At 19 Percent Profits.

Conclusions

This week's topics

In these notes, we discussed XBRL (the eXtensible Business Reporting Language). In particular, we looked at:

- The motivation behind XBRL.
- The most recent XBRL specification.
- The associated concepts: taxonomies and the discoverable taxonomy set.
- The XLink standard, and how it is used to extend XML Schema.
- An example.
- The IFRS-GP taxonomy.