The Data Point Model methodology in the European Supervision: COREP/FINREP

Ignacio Boixo. Eurofiling / University of Huelva. boixo@eurofiling.info

Abstract:

This paper describes the tagging needed when linking regulatory frameworks with automated reporting systems, the automated tagging problems, tagging in hierarchical and dimensional (tabular) reports, dimensional tagging characteristics, and the Data Point Model methodology as an evolutionary tagging approach.

While the *eXtensible Business Reporting Language* (XBRL) has been originally designed for addressing the tagging needs on accounting systems, a data normalization methodology is more and more required, especially for large regulatory frameworks.

The Data Point Model methodology has been developed for this purpose. Despite the fact that the Data Point Model will play a crucial role in the upcoming European supervision, the available references are however quite limited, at best. The author had compiled a very basic list of references[1] but this is still a long way from a fully comprehensive description.

The research topic of this paper is therefore a starting point for a formal description of the Data Point Model methodology, analyzing potential advantages and drawbacks. Examples are given illustrating the different topics addressed. The examples used in tagging come mainly from US-GAAP XBRL taxonomy, while the examples on Data Point Model come from European Supervision reporting frameworks, such as Basel II (COREP), Financial reporting (FINREP), Banking statistics and Solvency II.

Keywords: Reporting framework, European supervision, Data Point Model, data normalization

Paper (draft version) presented at 24th World Continuous Auditing and Reporting Symposium "Continuous Auditing and XBRL Coming of Age", May 3-4, 2012, Malatya / Turkey

Forthcoming *International Journal of Digital Accounting Research*

[1] "Data point model references" compiled by the author at www.eurofiling.info/dpm

1.- Tagging

1.1.- Tagging for humans and for computers

Tagging requirements are synonymous with reporting and public disclosure. Transmitting business reports from one entity to another requires a common understanding about the semantic meaning of the different tags used in the report, such as “Cash” or “Assets”. The IFRS and national GAAPs would be understood in this sense as guidelines about reporting tags.

While business reports were designed only for human reading, modern evolutions of the Renaissance double-entry bookkeeping system on paper (Pacioli, 1494) are in general use. Lists of accounting tags are generally accepted and commonly understood. The reporting entities file paper reports that in turn are computerized as images in implementations as EDGAR (Electronic Data Gathering Analysis and Retrieval), the filing system used by the USA Securities and Exchange Commission (SEC) to electronically receive all registrant filings that were earlier filed on paper.

With languages such as XBRL, the information unit was changed from a piece of paper to letters and digits. The convention about how to represent amounts in computerized digits is well established, as the semantic meaning of what the amount represents is not needed. But how to translate paper-based tags into computerized strings of letters and digits is more complex, as the semantic meaning is to be preserved.

The American Institute of Certified Public Accountants (AICPA) states “*There are some higher order skills and competencies that CPAs need to demonstrate in order to become licensed, and those are analytical and judgment skills. In terms of XBRL in the general corporate world, CPAs will be called upon to look at their financial and nonfinancial information and identify which tags go with it*” (Thomas, 2012).

The Securities and Exchange Commission explains that “*filers must select tags from the US GAAP taxonomy which best represent their financial reporting concepts*” (SEC 2011). The HM Revenue and Customs (UK Tax Agency) defines Taxonomy as “*The ‘dictionaries’ containing the unique XBRL tags*” while synthesizes: “*The requirement to file a Company Tax Return online with accounts and computations in iXBRL format does not mean you need to change your accounting processes. It is only the final figures that need to be tagged*” (HMRC 2011).

One of the main advantages of the XBRL is the tagging (Bonsón *et al.*, 2009). Basically, tagging in this context would be understood as the process of the definition of a set of unique mappings from company’s accounting items to XBRL items. Some accounting items would already exist in the company’s accounting while others would be specifically created for a particular filing purpose. Some XBRL items would already exist in the applicable taxonomy while others would be specifically created for a particular filing purpose (taxonomy extension).

In such a way, the accountants and the computers work with the same tags for the same terms.

1.2.- Tagging the financial statements of XBRL International inc.

We will use as an example the recently published financial statements of XBRL International Inc. (*XII*) available in its website (XII, 2012). Under the corporate statement "Transparency in Business & Financial Reporting", the XBRL International financial statements from 2003 to 2011 (pdf format) are available, as well the XBRL instance document with financial statements 2011 (xbrl format).

In the example of XII financial statements, the "paper style" financial account is as follows:

ASSETS	2011		2010	
CURRENT ASSETS:				
Cash and equivalents	$ 721,596	2	624,658	1
Dues receivable	369,262	4	332,680	3
Contributions receivable	-		50,000	3
Co-sponsorship conference fees receivable, net of allowance	53,353	4	9,595	3
Investments	21,253	7,8	16,800	5,6
Prepaid expenses and other assets	7,193	10	11,080	9
Total Current Assets	1,172,657		1,044,813	

Figure 1: XII financial statements 2011, "paper style" view (fragment)

The numbers in red represent the line in which the original financial item has been tagged according to the US GAAP 2011 XBRL taxonomy:

label	us-gaap:Cash
⊞ contextRef	c001
⊞ unitRef	u001
decimals	0
precision	
xsi:nil	false
value	624.658

Cash	1	c001	USD	0	624.658
Cash	2	c002	USD	0	721.596
Accounts Receivable, Net, Current	3	c001	USD	0	392.275
Accounts Receivable, Net, Current	4	c002	USD	0	422.615
Other Short-term Investments	5	c001	USD	0	16.800
Short-term Investments	6	c001	USD	0	16.800
Other Short-term Investments	7	c002	USD	0	21.253
Short-term Investments	8	c002	USD	0	21.253
Prepaid Expense and Other Assets, Current	9	c001	USD	0	11.080
Prepaid Expense and Other Assets, Current	10	c002	USD	0	7.193

Figure 2: XII financial statements 2011, XBRL document, view with Arelle

The lines 1 and 2 are quite easy: US GAAP 2011 "Cash" is the XII financial statement "Cash and equivalents". Idem for lines 9 and 10, where US GAAP 2011 "Prepaid Expense and Other Assets, Current" is the XII financial statement "Prepaid expense and other assets".
The US GAAP 2011 tag for the lines 5 and 7 is "Other Short-term Investments" while the US GAAP 2011 tag for the lines 6 and 8 is "Short-term Investments". Both are the same XII financial statement "Investments".

The US GAAP 2011 tag for the lines 3 and 4 is "Accounts Receivable, Net, Current", as the sum of the XII financial statement "Dues receivable" plus "Contributions receivable" plus "Co-sponsorship conference fees receivable, net of allowance".

In order to maintain the internal XBRL tags to be as aligned as closely as possible with accounting tags, the Label Camel Case Concatenation LC3 convention is recommended in the commonly used Financial Reporting Taxonomy Architecture (Hamscher *et al.*, 2005), in which *"an important design goal for financial reporting taxonomies is to maximize the usability of the taxonomy to the non-technical (from a computer science perspective) users and experts of the reporting domain"*


```
<xbrli:context id="c001">
<xbrli:entity><xbrli:identifier scheme="www.xbrl.org">XII</...></...>
<xbrli:period><xbrli:instant>2010-06-30</...> </...></...>
<xbrli:context id="c002">
<xbrli:entity><xbrli:identifier scheme="www.xbrl.org">XII</...> </...>
<xbrli:period><xbrli:instant>2011-06-30</...> </...></...>
<xbrli:unit id="u001">  <xbrli:measure>iso4217:USD</...>   </...>

<us-gaap:Cash  contextRef="c001" unitRef="u001">624658</...>
<us-gaap:Cash  contextRef="c002" unitRef="u001">721596</...>
<us-gaap:AccountsReceivableNetCurrent  contextRef="c001" unitRef="u001">392275</...>
<us-gaap:AccountsReceivableNetCurrent  contextRef="c002" unitRef="u001">422615</...>
<us-gaap:OtherShortTermInvestments  contextRef="c001" unitRef="u001">16800</...>
<us-gaap:ShortTermInvestments  contextRef="c001" unitRef="u001">16800</...>
<us-gaap:OtherShortTermInvestments  contextRef="c002" unitRef="u001">21253</...>
<us-gaap:ShortTermInvestments  contextRef="c002" unitRef="u001">21253</...>
<us-gaap:PrepaidExpenseAndOtherAssetsCurrent  contextRef="c001" unitRef="u001">11080</...>
<us-gaap:PrepaidExpenseAndOtherAssetsCurrent  contextRef="c002" unitRef="u001">7193</...>
```

Figure 3: XII financial statements 2011, XBRL instance document, raw view

In a balance sheet it is not difficult to define the tags for each item. In the above example, "AccountsReceivableNetCurrent" is to be translated as "Accounts Receivable, Net, Current" according to LC3 convention. Therefore, all the accountants understand "Accounts Receivable, Net, Current" for US-GAAP and all the computers understand the XBRL tag "AccountsReceivableNetCurrent" with exactly the same meaning in relation with US GAAP.

1.3.- Tag uniqueness

Humans have certain flexibility to use slightly (or not so slightly) different words for the same meaning: this is an intrinsic characteristic of natural languages (Chomsky, 1956). However, tag uniqueness is a practical requisite in computing.

A specific characteristic of XBRL is the uniqueness of tags (Valentinetti, 2011). For example, to tag the item "Operating Profit", a company uses the label <Operating Profit>, other <Operating.Profit> while another <Operating_Profit>. From this point of view, one of the specific tasks of XBRL community is to "Standardize" the labels used to communicate electronically financial data (Nel and Steenkamp, 2008; Bonson *et al*., 2009; Gray and Miller, 2009).

Operationally, once the data contained in a report has been "labeled" with XBRL tags, applications can "recognize" what is the accounting entry for each of them; in this way, a report produced by a particular application package can be easily imported into any other application. This characteristic derives from the fact that the labels created in XBRL are constituted by simple strings of text which, by their nature, can be read without resorting to special interfaces (Bovee *et al*., 2002). It should be stressed, moreover, that every element identifies a single encoded concept in XBRL accounting: the uniqueness is a prerequisite for being able to store, retrieve, analyze, share and reuse information between computers (Ramin and Prather, 2003).

As is evident in many financial statements, the item used does not necessarily “fit” with predefined tags. This is especially true in “principles based” accounting systems. A solution is provided by the inherent “Extensibility” of the XBRL, at the cost of creating the “XBRL's paradoxical power”, the trade-offs between customizing to better parallel existing paper reports and compromising to more closely match the standards, and the research needed for the transition from free form to customized reports (Cohen, 2004). Some interoperability advantages are lost if each company is allowed to create tags for its own reporting (Valentinetti, 2011).

In some cases, specifically in mandatory filing supply chains, the regulation defines exhaustively the items to be reported; even if not all the items are applicable to all the reporting entities (*proportionality principle*). For the purposes of this paper, only this simplest scenario is studied, in which extensions are not allowed.

The tagging is a compromise between generalization (to be understood by everybody) and particularization (to better reflect the entity detailed situation). The tags particularly designed for an enterprise are probably unique, so it is more advantageous to use common tags. On the other hand, if only the tags provided by the regulator are used, the enterprise may be unable to properly reflect its own particularities. Recent academic research has illustrated this dilemma.

In practical terms, there are other sorts of discrepancies where several regulators can be cited. XBRL Spain jurisdiction undertook an exercise, years ago, to extend the IFRS taxonomy with the ES GAAP 2007 definitions. As a first step, each accountancy expert received one piece of the work in order to tag ES GAAP 2007 according to IFRS. In turn, other different accountants reviewed the work according to the “four eyes principle”. The results were disappointing. The discrepancies among accountants were so high than this particular approach was discontinued. A well-known academic expert was hired to carry out the matching. The result was that the vast majority ES GAAP 2007 tags did not match exactly, with total accuracy, with the equivalent IFRS tag. In some cases, the differences were minimal for a number of reasons (legal, customary…), and the absolute matching was not formally achieved. At the end of the process, XBRL Spain decided to use different tags for the brand new ES GAAP 2007 taxonomy.

However, the Business Register Working Group of XBRL Europe is currently matching tags among the GAAP of different countries.

The degree of precision doesn’t need to be absolute for Business Register comparative purposes, and the matching is now achievable.

1.4.- Reusing XBRL tags in the same report: dimensions

An inherent limitation of the tagging uniqueness is that each tag should be used for whichever entity and period. A tag labeled as “CashXII2010” identifying the entity and period obviously lost generalization, as the tags must be different for each reporting entity and period of time.

The XBRL “context” is used to define the entity and the period of the information. In the example, for the context “C001”, the entity and the period to which the financial statement is reported is, in XBRL, **"www.xbrl.org">XII** and **2010-06-30** while in the “paper style” report, the text is XBRL INTERNATIONAL, INC. STATEMENTS OF FINANCIAL POSITION. JUNE 30. ASSETS 2010.

In the case of the period of time, the convention of tagging beginning and end of period is not so evident. For instance, tagging as “CashBeginningYear” or “CashEndYear” still maintains the uniqueness. Why not use such kind of tags? It is a decision of modeling the financial statements tags. In the “paper style” example, the tag “Cash” is used in a tabular format, with the amount 624,658 for 2010 and 721,596 for 2011. The tagging in XBRL follows such convention, BUT at the cost of splitting the information.

The Cash for 2010 is reported as `<us-gaap:Cash contextRef="c001" >624658</…>` . The year 2010 is not evident in the reporting line. It is needed go to the XBRL instance header to realize that **c001** references **2010-06-30,** while the context **c002** is used for **2011-06-30**

The repeated expression `unitRef="u001"` is used to indicate than the figures are USD in the previous line `<xbrli:unit id="u001"> <xbrli:measure>iso4217:USD</…> </…>`. The currency unit is also not evident in the reporting line. Moreover, using different currencies, the same Cash tag would be reported several times, for USA, EUR, JPY and so on.

Each XBRL tag would be reused in an XBRL instance document for different Entities, Periods and Currencies, or whatever combinations among them. It is the dimensional (tabular) internal structure of XBRL, as these three dimensions are defined (hard coded) in the roots of the current XBRL language specification[2] dated 2003. The additional specification for XBRL dimensions, generalizing the use and definition of dimensions, was released in 2006.

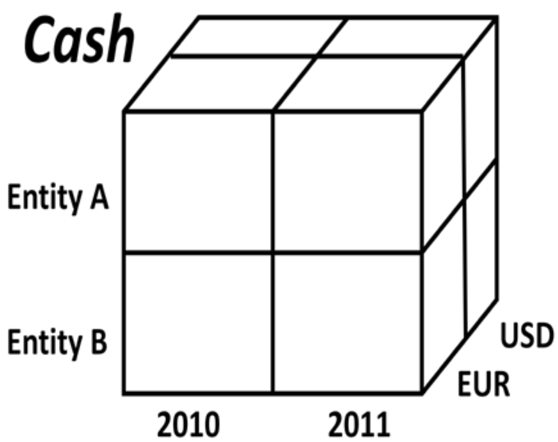

Figure 4 Inherent basic dimensional structure of XBRL: Entity, Period, Unit (Currency)

As in the equivalent “paper style” reports, it is forbidden to report different figures for the same combination of tag, entity, period and currency. The reason is evident: if two different figures are reported for exactly the same accountancy concept, at least one is false.

Therefore, the same tag would be reused several times in the same report, for different reporting entities, periods or currencies. This capacity of reusing the same tag for different contexts has been generalized as “dimensions” in XBRL.

[2] See XBRL Recommended Specifications at http://www.xbrl.org/SpecRecommendations

A hierarchical expression is isomorphic to an equivalent tabular expression. Both expressions have the same meaning even if the hierarchical expression is a label (string of letters and digits) and the tabular expression includes coordinates. Both expressions are isomorphic as any property that is true of one of the hierarchical expression is also true of the tabular expression and vice versa.

In fact, the US GAAP taxonomy uses the dimensional expression as in this **example of dimension** used in the XII financial statements, as follows:

2010
Accumulated Depreciation, Depletion and Amortization: Property, Plant and Equipment: Machinery and Equipment……………………………………………………………1,166

2010 Accumulated Depreciation, Depletion and Amortization	
	Machinery and Equipment
Property, Plant and Equipment	1,166

```
<xbrli:context id="c005"><xbrli:entity>
<xbrli:identifier scheme="www.xbrl.org">XII</...>
xbrli:segment><xbrldi:explicitMember dimension="us-gaap:PropertyPlantAndEquipmentByTypeAxis">
us-gaap:MachineryAndEquipmentMember</...></...></...>
<xbrli:period> <xbrli:instant>2010-06-30</...></...></...>

<us-gaap:AccumulatedDepreciationDepletionAndAmortizationPropertyPlantAndEquipment
contextRef="c005" unitRef="u001">1166</...>
```

Figure 5 Hierarchical, tabular and XBRL dimensional representation (example)

Additionally, the XBRL design had included some accounting characteristics in the underlying roots of the language. An XBRL item, in addition to a tag, has characteristics as a type (e.g. monetary item) and indication whether it is a credit/debit (assets or liabilities), if it is a stock or flow (balance sheet cash or sales per year).

The problem is exacerbated when the number of potential figures to be reported is several thousands.

The US-GAAP 2011 taxonomy has 15,000 concepts and 37,000 defined presentation items. The tag matching is not an easy task (Cangialosi, 2011)

The Basel II reporting framework definitions by the European Banking Authority, codenamed COREP (COmmon REPorting) has about 17,500 cells defined in the templates. Considering the open lists (entities of the group, securitizations, risks…) the potential number of data elements to be reported would reach the amount of 115,000 figures. In the case of the FINREP (FINancial REPorting) regulatory framework, the figures are 4,500 cells defined in the templates, with the potential number of data elements being about 25,000 (Weller, 2012)

Defining tag by tag, in an exhaustive hierarchical list, is not practical. Very quickly, the regulators decided to use tables instead of typical hierarchical list traditional in the balance sheets.

1.5.- Automated tagging

Why not automate the tagging process? Can a computer read an accounting statement and carry out the tagging process with an applicable taxonomy?

The exercise is simply imitating a computer programmed to carry out the tagging.

The source information will be the financial statements in the well know and widely used pdf format, created by Adobe Systems. The pdf format basically represents the information in "paper style image" ready to be printed or displayed on a screen.

Most of the financial statements are pure "paper style images" as the financial statements are first printed and then hand signed by the auditor. The hand signed financial statements are finally scanned and formatted as pdf documents. The original letters and digits are now images of letters and digits.

The first point is to transform financial statements from the pdf "paper style image" into a more computer-manageable file, as a text file. When opening the financial statements pdf 2011 with Acrobat (the product developed by Adobe Systems to read pdf files), it results that such "paper style image" is a pure image (as photography). The way in which a computer translates a photographical image into letters and digits is an Optical Character Recognition (OCR) process. The OCR process is less effective for a computer than for a human. This difference is used by "Completely Automated Public Turing test to tell Computers and Humans Apart" CAPTCHA doors (Ahn et al, 2008) requiring the user (human or computer) to type letters or digits from a distorted image that appears on the screen. Computers are currently unable to read letters or digits from images where trained (alphabetized) humans have no such problems.

When applying Acrobat OCR feature to the pdf image of the XII financial statements 2011, the results are pretty good in terms of translation into letters and digits. Practically all the letters and digits are correctly identified. Moreover, the paragraphs and figures are correctly separated by line breaks. Only some unexpected spaces were introduced in the figures (e.g. 721 ,596 instead 721,596).

The problem is in the internal order in the financial statements. The original "paper style image" has tree columns: item description, year 2011 and year 2010. The automated OCR is unable to maintain such table structure, being the result that item descriptions and figures are mixed with no order, being impossible to reconstruct the original table. The translation provides item lines and figures, but NOT what in are the figures for 2011 and 2010 for each item description. The tagging has been lost, and the translated financial report is now meaningless. See the text below where the line breaks have been substituted by "/" symbols for easier reading:

1,172,657 / 5,051 / 9,583 / $1,187,291 / LIABILITIES AND NET ASSETS / CURRENT Liabilities: / Accounts payable and accrued expenses / Unearned dues / Reserve - jurisdictional direct participants / Co-sponsorship conference fees payable / Deferred compensation / Total Liabilities / Net assets - unrestricted / $ 146,590 / 341 ,931 / 75,671 / 31,822 / 596,014 / 591,277 / $1 ,187,291 / See accompanying notes to financial statements. / 2 / 2010 / $ 624,658 / 332,680 / 50,000 / 9,595 / 16,800

Maybe then the problem is related to the use of "paper style image" financial statement. We would use the XII financial statements 2010, that are not "paper style image" only (the document has embedded the original letters and digits; even the format is not modifiable anymore). The widely used "copy and paste" functionality is operative, and these are the results (not too much better) when applied to the page 2:

"XBRL TNTERNATIONAL, INC. STATEMENTS OF FINAIICIAL POSITION JUNE 30. ASSETS CURRENT ASSETS: Cash and equivalents Dues receivable Contributions receivable Co-sponsorship conference fees receivable, net of allowance Investments Prepaid expenses and other assets Total Current Assets Property and equipment, net of accumulated depreciation of$1,167 and $700, respectively LIABILITIES Ai\D NET ASSETS CURRENT LIABILITIES: Accounts payable and accrued expenses Unearned dues Reserve- jurisdictional direct participants Co-sponsorship conference fees payable Deferred compensation Total Liabilities Net assets-unrestricted 20t0 2009 $ 624,658 332,680 50,000 9,595 16,900 I1,090 7,044,813 233 s 45,255 297,423 4,500 37,438 24,845 409,461 635,595 $ 1,045,046 s 712,954 364,131 154,854 114,730 6,656 r,353,325 700 $ 178,255 267,136 4,500 ll,3gl 17,559 478,840 875, Ig 5 $1,354,025"

As one conclusion, the most important problem when translating from accounting "paper style" documents to computer files is the tagging. When a list of words and figures has not a correspondence, the semantic meaning is lost, and the information becomes useless.

The traditional procedure to solve this translation from accounting "paper style" documents to computer files is using humans as "re-typing" machines, as currently the capacity of trained humans to "read and understand" accurately accounting "paper style" documents is much bigger than the computer-based programs. This is well known from empirical evidence in business reporting supply chains participants, as supervisors, business registries or financial information intermediaries.

It is accordingly assumed in this paper that the "paper style" accounting documents having a semantic meaning for humans are useless for computers. However, the computers are routinely programmed to print or display "paper style" documents from computer files. Therefore, the main restriction in the computerized accounting information supply chain is the asymmetry between the easy/cheap transformation from computer file to "paper style" document against the difficult/expensive transformation from "paper style" to computer file to document.

As a consequence, the topic to be investigated is how to define tags in computers with a functional equivalence to the "paper style" document.

2.- Tables and dimensions

A table is not free of notation problems. For instance, the decision of what elements are in the columns and which of them are in the rows are in general purely arbitrary. Moreover, a table following a typical spreadsheet is a bi-dimensional structure with rows and columns. It's easy to imagine that additional dimensions may apply to a table. An example would be sales by country, by size and by color. There is no way to represent these three dimensional structures in a bi-dimensional spreadsheet without "hiding one dimension". For instance, we may design a spreadsheet with sizes in the columns and colors in the rows, and repeat the spreadsheet per each country.

The problem here is what it is the "computer" tag. Basically, we have a single base item "sales" that would be reported by different breakdowns (dimensions), as country (France, Japan), size (Small, Large) and color (Red, Green) in several combinations.

One potential solution is defined a specific tag for each "cell", as follows:

SalesFranceSmallGreen	
SalesFranceSmallRed	
SalesFranceLargeGreen	
SalesFranceLargeGreen	
SalesJapanSmallGreen	
SalesJapanSmallRed	
SalesJapanLargeGreen	
SalesJapanLargeRed	

Figure 6: One fully explicit tag per cell

The resulting list is the complete enumeration, or Cartesian product, in which all the valid combinations receive an individual tag.

Another solution would be a bi-dimensional form, spreadsheet style, by arbitrary selecting a dimension (or combination of dimensions) as columns and the combination of the other dimensions as rows. There are different possibilities of draw a bi-dimensional tables under such restrictions, being one of them the following:

	Red	Green
SalesFranceSmall		
SalesFranceLarge		
SalesJapanSmall		
SalesJapanLarge		

Figure 7: Moderately dimensional tags

This approach is known as "moderately dimensional" or "form centric" among practitioners.

The advantage of this approach is the preservation of the "referential integrity". It is possible design the dimensions and base items in such a way that the shape of the original forms is more or less maintained in the resulting "computer" tags. This is feasible as part of the dimensions remain "collapsed" into the tags identifying base items. With an intelligent selection of "collapsed" dimensions, the shape of the original bi-dimensional templates is basically preserved.

The disadvantage with this solution is that the selection of which is in the base item tag and which is on the dimension is purely arbitrary. Moreover, the selection between base items and dimensions is not commutative. A table would collapse the dimension D into the base item, while other would make independent such dimension D. As result, the same dimension D is represented in two different and incompatibly ways in the same regulatory framework: collapsed as part of the base items and simultaneously defined independently.

Another approach is to reduce to a minimum the number of base items and express the all different combinations as dimensions. In the example the base item is “Sales” and all the disaggregations are modeled as dimensions, in a tri-dimensional cube, as follows:

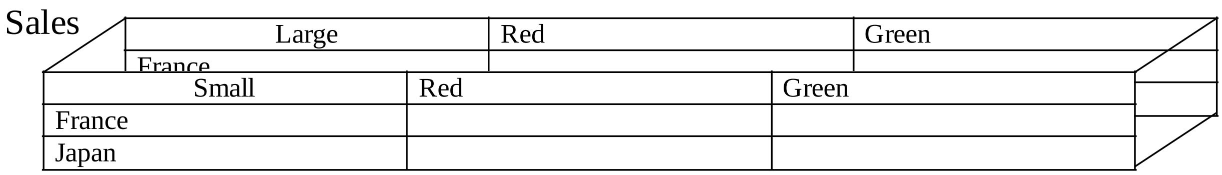

Figure 8: Data centric tags

This approach is known as “data centric” or “highly dimensional” among practitioners.

The advantage of this solution is that no selection of which disaggregations are in the tags and which are on the dimensions is needed any more: all is dimensional.

When the number of dimensions is two or less, the resulting model is an standard bi-dimensional spreadsheet. However, for three or more dimensions the resulting model is a “hypercube” with no easy projection into bi-dimensional tools.

The dimensions are commutative by nature: the metamodel is consistent irrespective of the order of the dimensions used for graphical representation.

In the example, colors are in the axis X, countries in the axis Y and sizes in the axis Z. This is an arbitrary and irrelevant use case. In other use case the axis would be selected in a different order with exactly the same metamodel.

The disadvantage of this approach is the loss of the “referential integrity”, or property by which the relationships between the original regulatory framework and the dimensional representation (metamodel) remain consistent. As everything is commutative dimensions and very few base items, the information about the shape of the original forms is lost. This lost of information is due that the design or particular order in which the cells are positioned in the original forms is, in general, purely arbitrary (Matherne, 2012). The transformation of an arbitrary positional model into a normalized non-positional model produces the loss of the positional information.

The usual solution is creating a data dictionary with a double entry table, in which each original cell has its translation into dimensional coordinates. In the Chapter “Using the Data Point Model” are show several examples of data dictionaries.

As the reporting framework cells are arbitrarily positioned, is deductible that such positional arrangement in not unique. In fact, it is not impossible that each particular observer would design a particular positional rendering arrangement for each particular use. The metamodel is stable and unaffected for such different positional renderings. The only requirement is the creation of a rendering data dictionary for each use[3].

In a reporting framework table, there are no generally accepted conventions that define what desegregations are to be represented into columns versus what desegregations are represented in the rows and what are maintained in the header, forcing new tables for each occurrence. Moreover, the same figure may appear in several tables. This is typical when a table with details includes a total and this total is also to be reported in the primary table.

For instance, we will have a primary table with two countries (France, Japan), and one table per country with details (Size, Color, Totals). The figure of the total sales in the details of France must match with the sales of France in the summary table.

If we use tags with full descriptions, we will have two different items for the same value: SummaryFrance and SalesDetailFranceTotal

However, using a disaggregation notation we will have a single definition: Sales (France).

This tabular (aka dimensional) approach offers advantages and drawbacks. The number of definitions is reduced. Only the base item (Sales) and the desegregations (France, Japan, Small, Red, Green) need to be defined. These advantages are more evident as larger is the number of applicable disaggregations.

Another advantage is that the use of this formalization has the commutative property. It's easy to detect that sales (France, Small, Red) are the same as sales (Red, Small, France). However, if we use the Cartesian explosion for tagging, the commutative property will be lost. For instance, a computer cannot detect that SalesFranceSmallRed is the same than SalesRedSmallFrance. The commutative property is only applicable to the desegregations expressed as indexes but not when "hardwired" or "collapsed" in to the tag. For instance, SalesFrance (Small, Red) is different for a computer than SalesRed (Small, France).

[3] In a balance sheet, when the tag AssetsCurrentCash is found it is easy to find the original item in the balance sheet. But when an XBRL expression such as Assets (N5, J43) is found, it is not evident to realize that it is Assets(Current,Cash). One of the main objectives of the original XBRL design was to maintain compatibility between the paper documentation and the XBRL stuff, in the sense that the XBRL tags displayed with basic tools will provide enough information to point to the original definition on paper. This is not a problem when the tags are composed of the full enumeration of desegregations such as AssetsCurrentCash. But when we have a very limited number of tags and a large number of desegregations coded in a meaningless structured reverse from the XBRL instance document to the original regulatory framework paper, it is extremely difficult without proper tools. The current efforts in fostering the XBRL rendering specification are oriented to providing a standardized way to link the original reporting frameworks with the XBRL instance documents expressed as tags plus desegregations. The XBRL consortium is developing the "rendering" specification exactly for this use. See the Table Linkbase Public Working Draft at http://www.xbrl.org/SpecPWDs (accessed in April 2012).

3.- Tagging tabular reporting frameworks.

One of the problems is the empirical evidence (well known by XBRL practitioners) in relation with the difficulties for IT XBRL experts to understand the regulatory frameworks and, vice versa, the difficulties for regulators to read XBRL taxonomies. For an accountant, even with limited English knowledge, it is easy identify tags as “Cash” or “Accounts Receivable, Net, Current”. However, for an IT expert but accountancy illiterate, even with English as the mother tongue, the difference among accountancy tags is usually inextricably arcane.

Creating a regulatory framework is mainly an issue for regulation experts. The tool generally available is the typical bi-dimensional spreadsheet. However, the bi-dimensional-only spreadsheet cell representation cannot capture easily the full set of breakdowns. Of course, IT experts may help by providing tools and expertise. But the semantic definition of regulatory frameworks is definitely a task for the regulators. And the only tool widely available among regulatory experts is the standard bi-dimensional spreadsheet.

The situation in Europe with 27 national regulators plus the European regulator is a specific case about how to define items with exactly the same meaning in all the 27+1 jurisdictions.

In the case of the cross-border Dexia Group, the Head of Prudential Policy stated: “*as the regulations are different, a slight shift in the definition is present and leads to distortions of the presentation of the same concept in the different reports. Furthermore, no precise and harmonized definition does exist, some reporting companies are considering full exposure, other “non-depreciated” exposure etc.. Above the problem of definition, another challenge is to comply with all the different interpretations for each country. Cross-border reporting turns out to be a real nightmare.*” (Pellizzari, 2009)

When the regulators design forms, they use bi-dimensional Excel spreadsheets, in which a large number of breakdowns are collapsed, and therefore hidden. What actually is the full set of breakdowns applicable to each cell is, for sure, in the brain of the regulation experts, but not always clear on the paper for everybody.

A strong preference has been voiced by reporting entities in recommending that the same cell of the same template should have identical interpretation, definition and use, and therefore the same value for the same circumstances, irrespective of the National Supervisory Authority jurisdiction. Clarifying the exact content of each cell with a full description of the context of breakdowns and scope is therefore of the utmost importance.

Once the regulatory framework is ready, the work of IT experts starts in such basis. The problem is basically that the regulatory frameworks are, in general, arbitrarily designed.

The European data modeling for Banking Supervision goes back to February 2005 when the kick-off meeting was held (Boixo and Schmehl, 2010) under the guidance of Pierre-Yves Thoraval as COREP Chair, Adrian Abbot of the UK-Financial Services Authority and Frédéric Marié of the Banque de France, along with international XBRL experts, such as Charlie Hoffman and Walter Hamscher. It was agreed the use of XBRL and discussed how to apply XBRL to capture regulatory frameworks. Required functionalities were analyzed, especially, more dimensional features that the existing at such point of time in XBRL specification.

During the initial COREP and FINREP development of XBRL taxonomies, it was never evident if breakdowns in different tables were equivalent or not. A first step was the concept of a Data

Matrix, showing the breakdowns applicable to the different tables. An excellent refinement came when the Bank of Italy contributed the much more robust Matrix Schemas (Excel files with breakdowns in a structured form) that have since been routinely published with each XBRL taxonomy release, thus providing an important step for quality control. Thanks to this contribution to the process, the Data Model much facilitates the creation of the corresponding XBRL taxonomy.

The Data Point Model is a collective result of this common effort.

The origins of the Data Point Model are rooted in the Matrix Schema of the Bank of Italy, already implemented in the 90's (Del Vecchio, Di Giovanni & Pambianco, 2007). The Matrix Schema approach could be broadly described as a strategy to collect granular reporting information (at a very basic level, highly disaggregated) from reporting entities and aggregate it at supervisor level. The Matrix Schema is the definition of required disaggregate information defining the disaggregation in a particular notation, equivalent to dimensions. The Eurofiling initiative complemented the Matrix Schema methodology with the Data Matrix approach (Boixo and Flores, 2005) that basically links each breakdown with all the tables in which it is used. Further developments and the availability of proof-of-concept tools created the methodology described as the Data Point Model, in which each quantitative value (the Data Point) is described by a basic meaning (monetary asset, numerical percentage…) and the intersection of all applicable breakdowns (time, currency, country, collateral, amount type….).

The European Banking Authority (EBA, 2011), proposes that the *Reporting according to this regulation shall be done by institutions in accordance with the specifications included in the Data Point Model and the XBRL taxonomies*, while defining the Data Point Model *as a structured formal representation of the data, identifying all the business concepts and its relations, as well as validation rules, oriented to all kind of implementers. It contains all the relevant technical specifications necessary for developing an IT reporting format*.

In turn, the European Insurance and Occupational Pensions Authority (EIOPA, 2012) *envisages awarding a direct service contract for the development of Data Point Modeling for the Solvency II XBRL taxonomy.* Other examples may be found in different Financial Supervisors, as Bermuda, Moldavia or Peru.

However, the perception of the Data Point Model as being a silver bullet for these purposes is not unanimous. One critic of the Data Point Model warned that *the link between the taxonomy and the business use is more difficult to establish and recommending staying closer to the business side so that business users can read the taxonomy without a dictionary by their side*, thus proposing *a mixture between Data Point Modeling and the business use* (Benari et al, 2012).

Linking bi-dimensional *human-oriented* representation with formal *computer-oriented* data models has been an open issue since the 70's, when the *computer-oriented* Relational Data Base model was introduced, competing with the initial, more intuitive and *human-oriented* hierarchical models. But that is another story[4].

[4] See an explanation about hierarchical vs. relational models at http://en.wikipedia.org/wiki/Database_model

4.- Definition of Data Point Model

The Data Point Model is a methodology to analyze the regulatory frameworks and formally determine what the exact definition of each data point is as an expression of the basic meaning and the desegregations applicable. **The Data Point Model is not an XBRL artifact, but a methodology for understanding the reporting framework meaning.**

A Data Point Model would be defined as the exhaustive definition of each cell of a reporting framework (typically tabular) with a BASE ITEM and a list of applicable DIMENSION MEMBERS

A data point of a reporting framework is identified by the MEMBERS of the BASE ITEM and DIMENSIONS that characterize it.

Data Point

Identification of a base item and a member (component) of each applicable dimension (breakdown) describing this base item in order to explicitly define a piece of information (e.g. a cell in a template). The data point definition is therefore the intersection of the definitions of each applicable member projected over the definition of the base item.

Base Item

Basic [financial/prudential/statistical…] meaning (nature) of the data from the conceptual point of view of the reporting framework (e.g. FINREP: asset, liability, income, …; COREP: capital, exposure, …).

Basic characteristics of the data:

1. Type of data (monetary, percentage, number, date, …)
2. Period (instant, duration)
3. For monetary items (credit, debit)

Dimension

Each of the additional type of “characteristics/breakdowns/disaggregations/attributes” that identify in detail the information included in the data points (e.g. type of asset, currency of the instruments, sector of the counterparty, residence of the counterparty, location of the activity,...).

Every “dimension” must have at least two or more possible members (values).

The list of members would be explicit (when exhaustively pre-defined) of typed (when only the type is predefined, but not the exhaustive list of values, as dates, identification codes of securitizations, percentages, ...)

Every “dimension” has a “default member”: “Not applicable/All”. When a “data point” is not identified by a specific dimension, it is assumed that the member of that dimension is the default member.

The number of dimensions is a matter of opinion: it is possible to state a dimension for any attribute or to combine more than one attribute in a single dimension.

Member

Each “value” or concept of a single dimension (e.g. “Cash”, “Loans”, “Shares and other equity” are “members” of the dimension “Type of asset”).

Two concepts having the same label and sharing some characteristics, but being not exactly the same (resulting in different figure in some circumstances) must be identified as two different members.

A “member” can be used in more than one dimension with exactly the same meaning, although used in a different context (e.g. the member “Spain” can be used in the dimensions “residence of the counterparty”, “location of the activity”, …)

The set of “members” that share a common semantic nature would grouped (in “domains”) to facilitate its use by more than a dimension (e.g. the domain “Geographical area” is used to include all members related to countries and regions, regardless of the dimension in which they are used).

The members of a domain would also be ordered in a given hierarchy and represented as a tree (with nodes and children). The tree provides additional information regarding the relationship between the members included in a node and its children: equal than, higher than,… (e.g. Currencies “All” is equal than “Euro” and “Currencies other than euro”). The tree of members simplifies the definition of the valuation rules between the different data of a reporting framework: The relation between the different nodes and their children reflects the valuation rules implicit in the templates.

The inclusion of the members of some dimensions in a domain is a matter of opinion. They can be included in the same domain provided that they share the semantic nature, but, at least in some cases, it could be more useful to split them into more than one domain according to its conceptual nature.

Method

If two cells in the reporting framework are disaggregations, breakdowns or share characteristics of the same business concept, the differences are to be expressed as members of dimensions. At the end of the process, only a few of “pure” base items and a number of dimensional members are defined to express in a combinational structure the original regulatory framework.

5.- Dimensional Properties

Members growing arithmetically define potential data points exponentially

The Cartesian product (potential data points) defined by a number of desegregations (dimensions and its members) may be expressed according to the following formulae:

n_i = number of members of the dimension i

d = number of dimensions

$$\text{Total number of members} \leq \sum_{i=1\ to\ d} n_i$$

$$\text{Total number of potential data points} = \prod_{i=1\ to\ d} n_i$$

$$\text{Typically } \sum_{i=1\ to\ d} n_i \ll \prod_{i=1\ to\ d} n_i \text{ as it is equivalent to:}$$

$$\bar{n} = \frac{\sum_{i=1\ to\ d} n_i}{d} \qquad \bar{n}^d \approx \prod_{i=1\ to\ d} n_i \quad \text{and } d\bar{n} \ll \bar{n}^d$$

Figure 9: Number of members vs. number of resulting cells

The total number of potential data points (Cartesian product) has grown exponentially while the number of members has grown only arithmetically. As the typical number of dimensions and the average number of members of a dimension is greater than 10, it is only needed define 10x10=100 members to define 10^{10} = 10,000,000,000 different data points.

The reduction of different definitions using the dimensional approach versus an exhaustive definition item by item is clear.

Each cell defines one data point; each data point is defined by one cell (at least).

Each cell in the reporting framework should be defined as a ONE data point.

Several cells definitions would result in the SAME data point. For example, the total of a table of detail would be repeated in a summary table. The cells are different, but the meaning (and therefore the value) is always the same. The same data point is defined ONCE AND ONLY ONCE by the reporting framework, regardless of whether it is included in more than one table.

It is (usually) forbidden in the reporting frameworks to define (*invent*) any data point not used for at least one cell. Of the complete set of potential data points existing in the Cartesian product of dimensions, the Data Point Model methodology select only the subset of data points effectively defined by the regulatory framework.

In formal terms, the set of the reporting framework cells is a subjective function over a set of data points. The resulting set of data points is the Data Point Model. The other potential data points defined by the Cartesian product of dimensions-members are ignored.

C = set of cells in the reporting framework

D = set of data points

The data points function ¦ **: C ® *D*** is defined, as " ***c* Î *C,*** $***d*** Î ***D,*** ¦ (***c***) = ***d***

The data points function ¦ **: C ® *D*** is subjective (not injective), as " ***d* Î *D,*** $***c*** Î ***C,*** ¦ (***c***) = ***d***

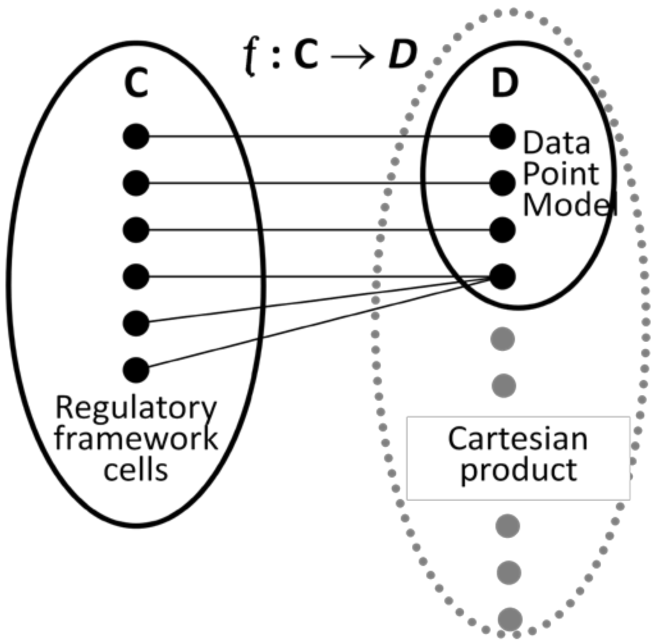

Figure 10: Subjective function of the Regulatory framework cells set into the Data Point Model

Each dimension would be used only zero or one times in each data point

It is not possible to use the same "dimension" more than once to identify a data point.

The data point is to be identified only by the strictly necessary dimensions. When these are not used in their identification, the Data Point Model assumes that they take the default member of the dimension (Not applicable/All).

Using the same dimension two or more times is not allowed. This usually happens because a specific domain member needs to be defined.

For instance, if the Exposure is Government, the Exposure cannot also be simultaneously Retail. If the Exposure is the sum of Government and Retail, not being the Total, a specific GovernmentPlusRetail member of Exposure Domain should be defined.

A member would be shared among several dimensions

Some dimensions can share the same “members” when they have exactly the same meaning, although they are used in a different context (e.g., the member “Spain” can be used in the dimensions “residence of the counterparty”, “location of the activity”, …)

The dimensional expression is commutative

The order in which the dimensions are expressed in the data point is not relevant. As a dimension can’t appear twice in the data point, there is no possibility of confusion.

For example, *Assets; Type of Instrument=Loan; Immediate Borrower=Credit Institution;* is identical to *Assets; Immediate Borrower=Credit Institution; Type of Instrument=Loan;*

The number of dimensions used in a data point is only limited by the number of existing dimensions

Of course, as each dimension would be used only once in each data point, the number of dimensions used in a data point cannot exceed the number of existing dimensions. But this is the only limit.

There is no fixed number of dimensions in the model. This implies that it is possible to add any new dimension to the reporting framework and consequently to the Data Point Model when needed, without restrictions.

Form Centric vs. Data Centric approaches

The design of the Data Point Model of a template-based reporting framework depends on the form in which its data points are identified. There are two main different approaches for identifying the financial information:

- **Form centric approach:** Data is identified according to a specific context (template, row and column) designed for requiring the data. Therefore, the data has no meaning without the positional context.
- **Data centric approach:** Data is identified by the values of a set of dimensions/attributes, regardless of any context. The data is meaningful on its own; no positional context (template, row and column) is needed for its identification.

6.- Using the Data Point Model

6.1.- Form Centric vs. Data Centric example

In reporting terms, the list of labels for the cell 0800:

> *Balance Sheet; Spanish Activities; Nominal; Assets; Residents in Spain; Euro; Loans; Credit Institutions*

is defined in the Data Point Model as

> *Assets; Type of Instrument=Loan; Immediate Borrower=Credit Institution; Residence of Counterparty=Spain; Location of the Activity=Spain; Currency=Euro; Amount Type=Nominal*

Form Centric coordinates:

> *Table 4000, cell coordinates (0400, 200) / cell number 0800*

Data Centric coordinates:

> *Base(mi6) Dimensional expression(AS5,CT4,RCes,RAes,CUeur,AT3)*

BALANCE SHEET (Spanish activities-Nominal)			**Table 4000**
ASSETS	C / R	**Residents in Spain**	
		Euro	**Rest**
		010	**020**
Loans	100	0400	0410
•Credit institutions	200	0800	0810

IDENTIFICATION OF THE DATA					
BY ITS LOCATION ON THE TABLES			BY ITS ATTRIBUTES		
			ATTRIBUTES	MEMBERS	CODE
Table 4000	Row (R) (200)	Column(C) (010)	▪ **Basic item**	Assets	mi6
Table 4000	Cell (0800)		▪ **Main category**		
			◆ *Type of instrument*	Loan	AS5
			▪ **Counterparty**		
			◆ *Inmediate borrower*	Credit institution	CT4
			▪ **Geographical area**		
			◆ *Residence of counterparty*	Spain	RCes
			◆ *Location of the activity*	Spain	RAes
			▪ **Currency**	Euro	CUeur
			▪ **Amount type**	Nominal	AT3
The same data has a different code in each table			Every data is identified by its attributes independently of the table		
The code identifies the location of the data in a table					

Figure 11: Example of Form cell and translation into a Data Point Model

6.2.- COREP Basic example

In the example below we can see the Capital Adequacy table in the reporting framework COREP, as well as its definition as a preliminary draft of Data Point Model[5]. The original is a summary table, basically a traditional hierarchical structure. The cells are identified with a four-digit code. The Data Point Model defines each cell as a list of characteristics.

As example, the cell 0010 "TOTAL OWN FUNDS FOR SOLVENCY PURPOSES" is defined in a data point with three components: "Own funds for solvency purposes"; Main category (Own funds: Total own funds); Amount type (Outstanding).

Cell	ID	LABEL
0010	1	TOTAL OWN FUNDS FOR SOLVENCY PURPOSES
	1.8	MEMORANDUM ITEMS
1270	1.8.1	IRB provision excess (+) / shortfall (-)
	2	CAPITAL REQUIREMENTS
1420	2.1.1.1.01	Central Governments or Central Banks
	3	MEMORANDUM ITEMS
1980	3.2.a	Solvency ratio (%)

Cell	Base	Main category	Portfolio	Approach	Exposure class	Amount type
0010	Own funds for solvency purposes	Own funds: Total own funds				Outstanding
1270	Memorandum item	Memorandum items: IRB provision excess (+)/ shortfall (-)				Outstanding
1420	Capital requirements: Credit risk	Type of exposure: Total exposures	Banking book	Credit risk: SA	SA approach: Central Government or Central Bank	Capital requirement
1980	Memorandum item (%)	Memorandum items: Solvency ratio (%)				Percentage

Figure 12: Capital Adequacy table (fragment) and equivalent Data Point Model (Rodriguez and Gutierrez, 2010)

[5] See "Ongoing intermediate release" (2010-11-10) at www.eurofiling.info/corepTaxonomy/taxonomy2012

6.3.- Using open source tool: FINREP example

An ExGen open source tool (Ruiz, Fiedura, *et al.*, 2012) used in the initiative Eurofiling is publicly available[6] to facilitate the creation, maintenance and representation of the Data Point Model.

In the example[7] below, corresponding to the table 6 of FINREP, coordinates (3,3), the cell is defined as Assets; Portfolio (Measured at amortized cost); Reporting scope (CRD –*Capital Risk Directive*- consolidation); Category of assets (Debt securities held); Counterparty sector (General governments); Amount type (Unimpaired); Delinquency (> 90 days≤ 180 days).

The cell meaning is fully described in the context of the FINREP reporting framework by the base item (Assets) and the intersection or "perimeter" of the applicable breakdowns (Portfolio, Reporting scope, Category of assets, Counterparty sector (General governments), Amount type, Delinquency). The definition (explicit or contextualized) of the cell in the original reporting framework is now replaced by an exhaustive list of applicable disaggregations. The tag representing the cell is now the nature of the amount (Assets) and the intersection of applicable characteristics.

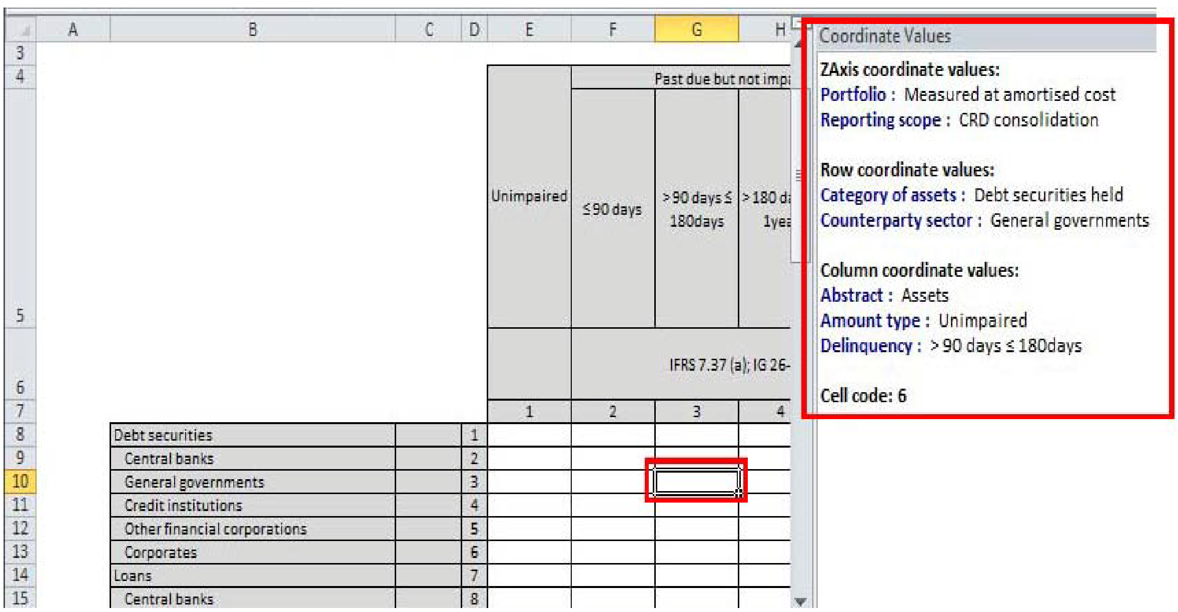

Figure 13: Mapping between original reporting framework cell and the corresponding data point (example)

Other advantage of the data point automation in this tool is the mechanical translation from a Data Point Model into an XBRL taxonomy. The process would be now greatly automated.

[6] See ExGen tool at www.openfiling.info/?page_id=67 website.

[7] See "XBRL proof of concept release" (2011-06-12) at www.eurofiling.info/finrepTaxonomy/taxonomy2012

6.4.- Data Point Model origins: BSI-MIR Statistics

The example is the two first lines of the table UEM.1 “Summary Balance Sheet (Business in Spain)”, their definition according to the Data Point Model structure, and the detail of Main category for the first line.

ASSETS	Total		Residents in Spain				Residents in other EMU states				Rest of the world	
			Euro		Foreign currencies		Euro		Foreign currencies		All currencies	
1. Cash (a)	0001		0081		0121							
2. Loans	0005		0085		0125		0205		0245		0285	

#	Table	Cell code	Base item label	Base item code	Main category	Counterparty	Geographical area	Currency	Time interval	Amount type	Amount interval	Related to minimum reserve	Entity code	Loan type by purpose	Transfer	Collateral/Guarantee	Percentage interval
1	UEM-1	0001	Assets	mi6	x	x	x	x	-	x	-	-	-	-	-	-	-
2	UEM-1	0081	Assets	mi6	x	x	x	x	-	x	-	-	-	-	-	-	-
3	UEM-1	0121	Assets	mi6	x	x	x	x	-	x	-	-	-	-	-	-	-
4	UEM-1	0005	Assets	mi6	x	x	x	x	x	x	-	-	-	-	-	-	-
5	UEM-1	0085	Assets	mi6	x	x	x	x	x	x	-	-	-	-	-	-	-
6	UEM-1	0125	Assets	mi6	x	x	x	x	x	x	-	-	-	-	-	-	-
7	UEM-1	0205	Assets	mi6	x	x	x	x	x	x	-	-	-	-	-	-	-
8	UEM-1	0245	Assets	mi6	x	x	x	x	x	x	-	-	-	-	-	-	-
9	UEM-1	0285	Assets	mi6	x	x	x	x	x	x	-	-	-	-	-	-	-

#	Table	Cell code	Base item label	Base item code	Main category	Assets	Liabilities	Liabilities and equity	Managed assets, custody asset	Cash on hand
						AS	LI	LE	MA	x7
1	UEM-1	0001	Assets	mi6	x	x				x
2	UEM-1	0081	Assets	mi6	x	x				x
3	UEM-1	0121	Assets	mi6	x	x				x
4	UEM-1	0005	Assets	mi6	x	x				

DataPoints	Base items	Item type	Name	ID	Period	Data type
6	Assets	Primary item	mi6	base_mi6	instant	monetary
6	Changes in assets	Primary item	md6	base_md6	duration	monetary

Figure 14: Reporting framework and Data Point Model for BSI-MIR statistics (fragment)

The cell 0001 (Cash) is defined as Assets (named “mi6”, see bottom) with a combination of members of (Main category, Counterpart, Geographic area, and Amount type). In the detail, the Main Category member is (Cash on hand).

Just below, the cell 0005 (Loans) shares the same disaggregations than the cell 0001. The Main category is not Cash on hand, but Loans (not represented due lack of space). The time interval breakdown is now need, as the Loans have a maturity period.

The base item Assets (named mi6) is different of the base item Changes in assets (named md6). The reason is more related with XBRL restrictions (each base item must be defined either stock (instant) or duration (flow)) than with statistical purposes. Hence the final language (XBRL) restrictions still "contaminate" with such restrictions in the "high level" modeling methodology (Data Point Model).

This example[8] is probably the first use of the Data Point Model, at least widely published. This reporting framework is actually used in Spain for the collection of statistics (Balance Sheet Items and Monetary Interest Rates, BSI-MIR) required by the European System of Central Banks. This Data Point Model was developed by the Bank of Spain, and later contributed to the Eurofiling initiative.

6.5.- Data Point Model embedded in COREP reporting framework

In this is example[9] only basic spreadsheet functionality is used, with no need for any specific tool. In this proof of concept artifact, the original template Credit Risk, Standard Approach is showing supervisory information (see left side: "credit counterparty" or "total exposures"). The template is a spreadsheet, as evident in the numbering in rows and columns. The spreadsheet has been prepared (by using basic spreadsheet functionality) to replace original labels ("credit counterparty" or "total exposures") with data point short-names (see right side: "SCC" or "254" linking with the data dictionary), while maintaining exactly the original format.

In this proof of concept is demonstrated who code each supervisory label with a code linked with the Data Point Model. In such a way, the regulatory expert would maintain the original template view, while the IT expert simply can ask for the short-names of the base items, dimensions and members. The IT experts fell generally more comfortable with unique short-names than with reporting labels. In this example, while defining the Data Point Model, the regulatory experts discovered slightly different labels or descriptions for the same element, that only can create confusion and errors among no experts.

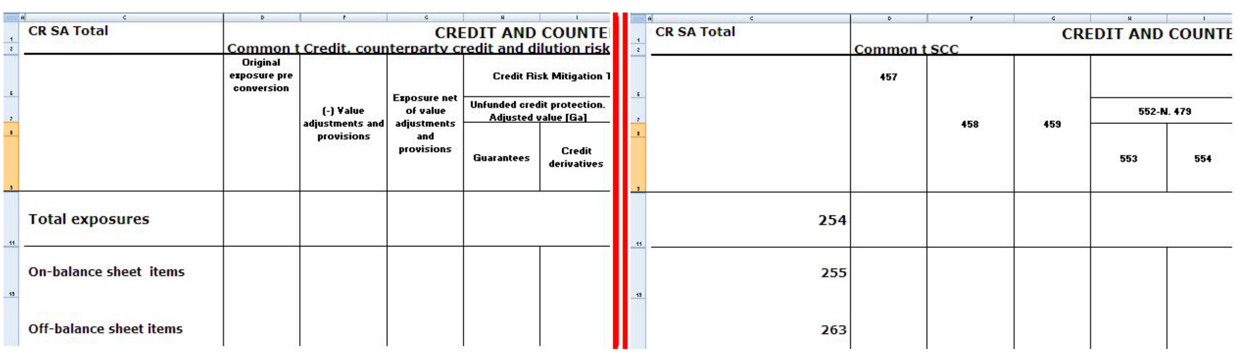

Figure 15: Example of Data Point Model embedded into COREP reporting framework (fragment)

[8] See "BSI-MIR Taxonomies" (2010-01-27) at www.eurofiling.info/bsi-mirTaxonomies/taxonomy

[9] See "Ongoing intermediate release" (2010-12-31) at www.eurofiling.info/corepTaxonomy/taxonomy2012

6.6.- Basel II, Financial, Statistics & Solvency II and the Data Point Model

The Data Point Model would be used (in theory) to reconcile different regulatory frameworks. At least some basic breakdowns are common irrespective the regulatory jurisdiction. Definition of countries, currencies, time intervals and so on are independent of the financial regulators. In a further step, it would be worth investigating if different regulatory frameworks share additional elements. The reduction of complexity (and very likely costs) is clear: a reduction of heterogeneity facilitates economies of scale for the benefit of standardization.

The Joint Expert Group on Reconciliation of credit institutions' statistical and supervisory reporting requirements (JEGR), established jointly by the European System of Central Banks' (ESCB) Statistics Committee and Financial Stability Committee together with the European Banking Authority has published on March 2012 a database and bridging manual[10] for Monetary and Financial Institutions MFI balance sheet and interest rate statistics and EBA guidelines on FINREP and COREP/Large Exposures.

An internal Data Point Model exercise made by the Eurofiling team shows below how much potentially compatible are the breakdowns of COREP/FINREP/BSI-MIR/SolvencyII. Commonality is found in 8 dimensions: Main categories, Amount type, Currency, Collateral, Geography, Sector, Entity code and Time intervals. COREP shares Risk type and Impairment with FINREP, Approach with SolvencyII and Percentages with BSI-MIR.

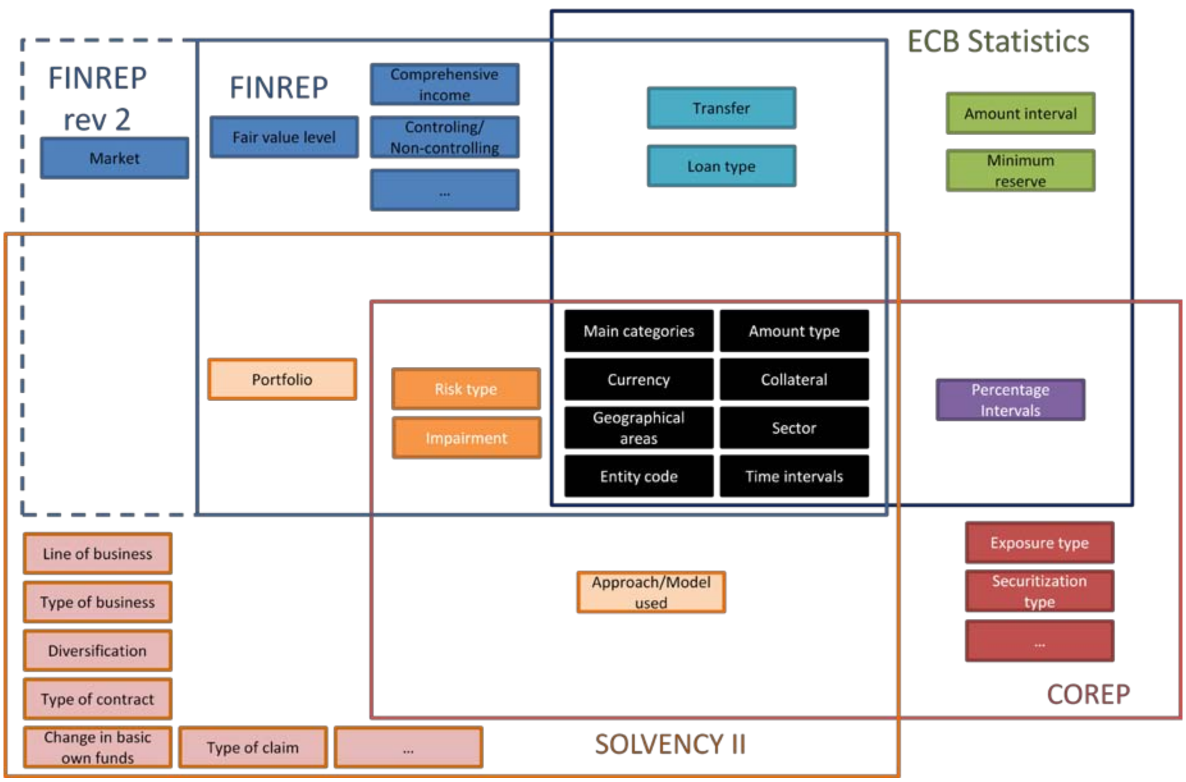

Figure 16: Dimensional reconciliation among COREP/FINREP/Solvency II/ECB Statistics (Ochocki, 2011)

---

[10] Available at www.ecb.int/press/pr/date/2012/html/pr120323.en.html

7.- Why the Data Point Model? Conclusions

The traditional problem is that the accountants have problems understanding data exchange formats (such as XBRL) and the IT experts are very reluctant to learn about accounting. Something in the middle as *lingua franca* is needed.

The data exchange formats are oriented for IT transmission, to be primarily managed by computers. To overload the reporting language with semantic and rendering features is possible, but neither quick nor cheap. And, in any case, a semantic layer such as the Data Point Model is always required.

The regulatory frameworks are not necessarily examples of data formalism. In the case of regional agreements such as European COREP/FINREP, they are the result of many meetings and transactions to accommodate different national requirements. The rationale about the specific inclusion/exclusion of each cells is not always fully tracked, as there are not a formal method enough detailed to tracking at cell level. Maybe that after a sufficient number of meetings and the usual rotation of participants, sometimes the original rationale would be lost (Stephenson, 1967). The Data Point Model may also be of help here as a formal method for tracking agreements at cell level.

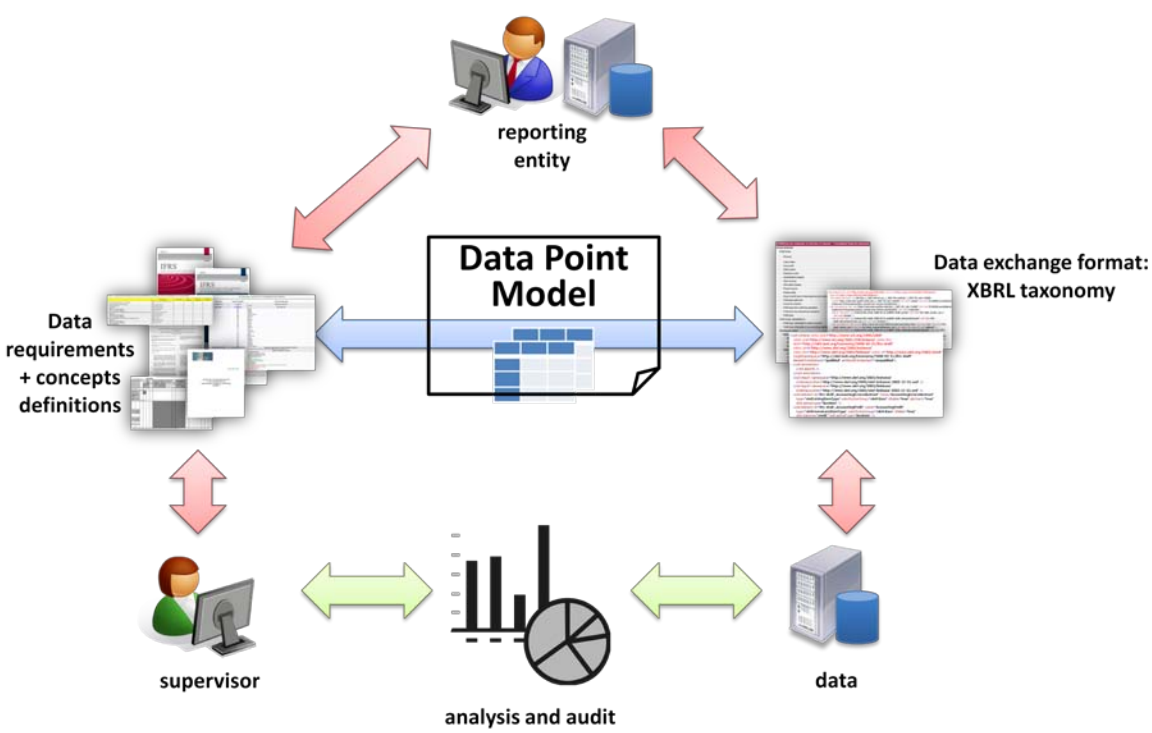

Figure 17: Data Point Model as metamodel between regulatory framework and data exchange format

The Data Point Model is very useful for understanding and even debugging the reporting framework. A kind of “four eyes principle” is applied by the team in charge of developing the Data Point Model.

First at all, the team must be basically composed by accountants. The Data Point Model is a re-codification of the regulatory framework, and only the experts in regulation understand it. In the current practice, the Supervisor that promulgates the regulatory framework also publishes the Data Point Model. Therefore, the supervisors developing the Data Point Model would eventually reach situations like duplicated data points (the same cell is used in different tables), members with different meanings in different data points (inconsistent breakdowns, disaggregations or characteristics), unusual members of members used in a single cell (peculiarities to be confirmed), insufficient information in the reporting framework to create some data points (source of errors), and so on.

The Data Point Model helps IT experts to create XBRL taxonomies, databases and IT reporting systems. The specific breakdowns, disaggregations or characteristics required by the supervisor are well documented. The tag of each cell is now a set or coordinates in the information system.

As the number of Base Items and Members is much smaller than the number of cells, their definitions would be more exhaustive and consistent. Defining tag by tag, most of the information is duplicated or, even worse, slightly different in each jurisdiction. By using the members consistently, each breakdown, disaggregation or characteristic is defined once and only once, and then consistently applied in all the data points in which it appears.

However, nothing is perfect.

The use of the Data Point Model or, more properly, the XBRL taxonomies generated following exactly the Data Point Model approach, has been criticized by several experts. *Data Point Modeling uses XBRL dimensions to express all characteristics (so called hidden dimensions). In highly dimensional taxonomies, each concept is exploded along all its dimensions (hidden or not) and the "logical concept" does not appear any more.* (Jarry, 2012). Other very experienced XBRL practitioners claim that this data centric (*highly dimensional*) destroys the immediate correspondence between the reported information in XBRL and the original reporting framework. This is true. Going back from a dimensional expression to the original cell (or cells) in the reporting framework is, by far, more complex than can be read in a hierarchical LC3 tag like the US GAAP in the example of XBRL International.

Further research into the formal modeling for large regulatory frameworks is required. The Data Point Model is an advance, but not exempt from controversy.

While the task of modeling reporting frameworks is IT-related, the experts in charge of creating formal models for regulatory frameworks (as the Data Point Model) are the qualified accountants. Tagging is basically an accountancy issue.

So, there is a big opportunity for academic research in this area of accounting.

REFERENCES

AHN, L. *et al.* (2008) "reCAPTCHA: Human-Based Character Recognition via Web Security Measures" SCIENCE Vol 321, 12 September 2008: 1465-1468 http://www.cs.cmu.edu/~biglou/reCAPTCHA_Science.pdf

BENARI, G. *et al* (2012) "EIOPA aligns Solvency II XBRL taxonomy with EBA" available at http://solvencyiiwire.com/eiopa-aligns-solvency-ii-xbrl-taxonomy-with-eba/9555

BOIXO, I. FLORES, F. (2005). "New Technical and Normative Challenges for XBRL: Multidimensionality in the COREP Taxonomy", International Journal of Digital Accounting Research, vol.5, n.9: 79-104 http://dx.doi.org/10.4192/1577-8517-v5_3

BOIXO, I., SCHMEHL, K. (2010). "Collaborative Development of IT Supervisory Frameworks", XBRL Global, 1, 2, 37-44. http://www.eurofiling.info/documents/XBRLglobalVol2SB.pdf

BONSÓN, E., ESCOBAR, T., FLORES, F. (2009): "Towards an ontology-based network for banking supervision", Online Information Review, vol.33, n.5: 943-955

BOVEE, M., ETTREDGE, M., SRIVASTAVA, R.P., VASARHELYI, M.A. (2002): "Does the Year 2000 XBRL Taxonomy Accommodate Current Business Financial Reporting Practice?", Journal of Information Systems, vol.16, n.2: 165-182

CANGIALOSI, L. (2011) "XBRL Rulemaking Request" Letter to SEC from the Committee on Corporate Reporting (CCR) of Financial Executives International (FEI) http://www.financialexecutives.org/KenticoCMS/Knowledge-Center/Comment-Letters/FEI---CCR-XBRL-Rulemaking-Request.aspx Accessed in April 2012

CHOMSKY, N. (1956) "Three Models for the Description of Language". IRE Transactions on Information Theory 2 (3): 113–124

COHEN, E.E. (2004): "CAP Forum on E-Business: Compromise or Customize: XBRL's Paradoxical Power", Canadian Accounting Perspectives, vol.3, n.2: 187-206

DEL VECCHIO, V., DI GIOVANNI, F., PAMBIANCO, S. (2007) "Unified model for statistical data representation and processing" Accessed in April 2012 http://www.bancaditalia.it/statistiche/quadro_norma_metodo/modell_SIS/matrixmod.pdf

EBA (2011) "EBA consultation paper on draft ITS on supervisory reporting requirements" available at: http://www.eba.europa.eu/News--Communications/Year/2011/EBA-consultation-paper-on-draft-ITS-on-supervisory.aspx

EIOPA (2012) "Ex-ante Publicity for DPM Negotiated Procurement Procedure" https://eiopa.europa.eu/fileadmin/tx_dam/files/Procurement/calls_for_expression_of_interest/Ex-ante_Publicity_for_Negociated_DPM_Procurement_Procedure.pdf Accessed in April 2012

GRAY, G.L., MILLER, D.W. (2009): "XBRL: Solving Real-World Problems", International Journal of Disclosure and Governance, vol.6, n.3: 207-223

HAMSCHER, W. *et al* (2005) "Financial Reporting Taxonomy Architecture" http://www.xbrl.org/technical/guidance/FRTA-RECOMMENDATION-2005-04-25.htm

HMRC (2011) “XBRL - when to tag, how to tag, what to tag” http://www.hmrc.gov.uk/ct/ct-online/file-return/online-xbrltag.pdf Accessed in April 2012

JARRY, E. (2012) “EIOPA methodology for architecting the taxonomies” http://archive.xbrl.org/24th/sites/24thconference.xbrl.org/files/TSMT10EricJarry.pdf

MATHERNE L. (2012). “Defining the interoperability issue” speech to the XBRL Member Assembly held on 19 March 2012, Abu Dhabi. http://www.xbrl.org/sites/xbrl.org/files/imce/Defining%20the%20interoperability%20issue%20-%20Louis%20Matherne.pdf Accessed in April 2012

NEL, G.F., STEENKAMP, L.P. (2008): "An explanatory study of chartered accountants' awareness and understanding of XBRL", Meditari Accountancy Research, vol.16, n.1: 79-93

OCHOCKI, B. *et al.* (2011) “Data modeling and ExcelXBRLGen” Accessed in April 2012 http://www.openfiling.info/wp-content/upLoads/data/EurofilingWebinar20110713.pdf

PACIOLI, L. (1494) “Tractatus Particularis de Computis et Scripturis” Summa de Aritmetica, Geometria, Proportioni et Proportionalita. Venetia.

PELLIZZARI, G. (2009) “Implementing XBRL for a cross-border banking group in Europe” http://www.eurofiling.info/10th_workshop/presentations/GPellizzariImplementingXBRLCross-borderBankingGroupEurope.ppt Accessed in April 2012

RAMIN, K.P., PRATHER, D.A. (2003): "Building an XBRL IAS/IFRS Taxonomy", The CPA Journal, vol.73, n.5

RODRIGUEZ, C., GUTIERREZ, L. (2010) “COREP: Data Point Model 2010” http://www.eurofiling.info/corepTaxonomy/20100705%20-%20COREP%20%20Data%20Point%20Model.ppt Accessed in April 2012

RUIZ, A., FIEDURA, R. *et at*. (2012) “Data Point Modeller EXGEN. Excel Generator” www.openfiling.info/?page_id=67 Accessed in April 2012

SEC (2011) “Summary of XBRL Information for Phase 3 Filers” http://www.sec.gov/spotlight/xbrl/xbrlsummaryinfophase3-051011.shtml Accessed in April 2012

STEPHENSON, G. R. (1967). Cultural acquisition of a specific learned response among rhesus monkeys. In: Starek, D., Schneider, R., and Kuhn, H. J. (eds.), Progress in Primatology, Stuttgart: Fischer, pp. 279-288.

THOMAS, A. (2012) “Training CPAs in XBRL tagging”. International Business Reporting (iBR). vol 02 issue 01, 12-13

VALENTINETTI, D. (2011) “XBRL e Sistema Informativo di Bilancio”. Doctoral dissertation. Universitá degli Studi “G. d’Annunzio”

WELLER, A. (2012) “Eurofilings - Modeling taxonomies using DPM” http://archive.xbrl.org/24th/sites/24thconference.xbrl.org/files/TSMT9AndreasWeller.pdf

XII (2012) “XBRL International Inc. Financial Statements” http://www.xbrl.org/financial-statements-0 Accessed in April 2012