XLDM: AN XLINK-BASED MULTIDIMENSIONAL METAMODEL

Paulo Caetano da Silva
Federal University of Pernambuco, Brazil
Mateus Silqueira Hickson Cruz
Ceu System - São Paulo, Brazil
Valéria Cesário Times
Federal University of Pernambuco, Brazil

_____

ABSTRACT

The growth of data available on the Internet and the improvement of ways to handle them consist of an important issue while designing a data model. In this context, XML provides the necessary formalism to establish a standard to represent and exchange data. Since the technologies of data warehouse are often used for data analysis, it is necessary to define a cube model data to XML. However, data representation in XML may generate syntactic, semantic and structural heterogeneity problems on XML documents, which are not considered by related approaches. To solve these problems, it is required the definition of a data schema. This paper proposes a metamodel to specify XML document cubes, based on relationships between elements and XML documents. This approach solves the XML data heterogeneity problems by taking advantages of data schema definition and relationships defined by XLink. The methodology used provides formal rules to define the concepts proposed. Following this formalism is then instantiated using XML Schema and XLink. It also presents a case study in the medical field and a comparison with XBRL Dimensions and a financial and multidimensional data model which uses XLink.

Keywords: XLDM, XML, XLink, XBRL, Multidimensional Data Metamodel

_____

Address for correspondence / *Endereço para correspondência*

Paulo Caetano da Silva, Graduated in Chemical Engineering at Bahia Federal University (1985). Master's degree in computer networks at Salvador University – UNIFACS (2003). PhD at Pernambuco Federal University, in Computer Sciences, in database – XML area (2010). Currently, he is a professor at Salvador University - UNIFACS, in the master course of Systems and Computation, he is also an analyst at Brazil Central Bank. He has experience in the area of computer sciences, focused on Software Engineering, Database and XML, actuating, mainly, in the following themes: XBRL, OLAP for XML, information systems, web and financial information.E-mail: paulo.caetano@bcb.gov.br

Mateus Silqueira Hickson Cruz, Graduated in Computer Science at Bahia Federal University (2009). Has interest in the following themes: databases, mobile applications development, XML and XBRL.E-mail: mateus@ceusystem.com.br

*Valéria Cesário Times,*Graduated at Pernambuco Catholic University (1991), master's degree in Computer Sciences at Pernambuco Federal University (1994) and PhD on Computer Science - Leeds Metropolitan University (1999). Currently is an adjunct professor I at Pernambuco Federal University. She has experience in the area of Computer Sciences, focused on Information Systems, actuating mainly in the following themes: data warehouse, geographic data base, geographic information services, geographic information systems and OLAP tools.E-mail: vct@cin.ufpe

1.   INTRODUCTION

Data are usually available in several formats. XML (eXtensible Markup Language) is used to integrate them in order to achieve efficient data interchange and handling. XML, being an extensible meta-language, allows new markup languages to be defined for specific domains. Due to its extensibility, XML is used for heterogeneous data sources integration. This makes XML documents a rich source of information for the organizational decision maker. Similarly, the use of Data Warehouses systems (Kimball & Ross, 2002) allows the identification of tendencies and standards in order to better conduct the companies' businesses. However, the integrated use of these technologies is still under development. In order to transform XML into a technology that helps the decision making process, it is integrated its use with Data Warehouses systems to unify them.

Applications and technologies derived from XML use XLink (XML Linking Language) (XML Linking, 2001) as an alternative for representing the semantics and the structure of information, expressing relations between concepts which are usually defined in a schema based on XML Schema (XML Schema, 2004). XBRL (eXtensible Business Reporting Language) (XBRL International, 2008) is, among the works that represent data semantics usage of XLink, an international standard to represent and publish financial reports that use extended links for modeling financial concepts (e.g. arithmetic operations between accounting facts).

On the other hand, XML presents some problems due to its flexibility in data representation, known as heterogeneities: (i) semantics, in which similar information is represented by different names (e.g. *enterprise* and *company*) or dissimilar information is represented by the same name (e.g. *virus* in the informatics area and in the medical area); (ii) syntax, where the semantically equal content is represented in several ways, for example, in different languages or in several measurement units (e.g. meters and feet); and (iii) structure, where the data are organized in several structures (e.g. in different kinds of hierarchies, attributes or elements) (Näppilä, Järvelin & Niemi, 2008). This flexibility in representation is important, though it makes the usage of XML data a complex task. XLink has been used to represent semantic and structural information, expressing relations among concepts that are normally defined through XML Schema. From the combined use of XML Schema and XLink, this paper represents a multidimensional metamodel, entitled XLDM *(XLink Multidimensional Data Metamodel)*, which solves the heterogeneity questions in XML and specifies data cube models for several semi-structured data applications.

Among the related papers´ analysis, the existence of a model that solves heterogeneity problems in XML and makes its application possible in multiple human knowledge domains has not been verified. Thus, the development of a metamodel based on XML documents, and relationships to solve these questions, consists of the motivation for this paper. Elements, attributes and relationships were defined for the XLDM specification, allowing a greater expressivity and enabling its applicability in different domains. The formalization of XLDM was created to allow the multidimensional data schema to be based on XML Schema's and XLink's definitions.

This paper is organized as follows: Section 2 discusses the main proposals for defining XML data models, including a description of the most important dimensional

models and other approaches of which definitions have affected the development of our work. Contributions are presented in Section 3, which includes the XLDM specification and its formalization. In Section 4 a case study for the medical area is to be found. A comparison against a multidimensional metamodel that uses XLink for the financial area is given in Section 5, showing the proposed solution wideness, and finally, conclusions are presented in Section 6.

2.   DATA WAREHOUSE FOR XML DATA

A Data Warehouse system architecture for complex data is proposed by Boussaid, Messaoud, Choquet and Anthoard (2006) using XML documents. Based on the database view concept, Baril and Bellahséne (2003) present architecture for XML data integration and a formalism for DW specification. A data model is defined, using DTD (Document Type Definitions), to represent each view, aiming at semi-structured organized data. Then, it is proposed a Data Warehouse based on the views. Trujillo, Luján-Mora and Song (2004) use UML class diagrams to represent Data Warehouse systems in a conceptual level. From the definition of a DTD, which represents the same multidimensional model specified by the class diagram, XML documents are generated for data exchange. Pokorny (2001) describes how to represent a star model (Kimball & Ross, 2002) in XML, proposing the XML-Star schema and using DTD to explicit dimension hierarchies. One dimension was modeled as a DTD sequence logically associated, resembling the referential integrity in a relational database. The dimensional structures were not defined in the XML schema, leaving to the software applications the data multidimensionality understanding. Golfarelli, Rizzi and Vrdoljak (2001) discuss a multidimensional model represented in attribute trees. They use XML Schema to express the multidimensional model through the relation between sub-elements. Nassis, Rajugan, Dillon and Rahayu (2004) propose an object-oriented approach to develop a conceptual model for DW, named XML Document Warehouses (XDW). They also define dimensions by using XML and UML packages diagrams, in order to contribute to the hierarchic conceptual views construction. An XML repository, named xFACT, was built from the integration of object oriented concepts with XML Schema. Jensen, Moller and Pedersen (2001) present architecture in which the data in XML and relational formats are the information sources. The data schema is represented in UML diagram classes, and, subsequently, mapped into a relational structure. This process is defined by the authors as a logical data integration. It is, then, possible to use OLAP query tools, assuring time gain, since there is no physical data integration. Pedersen, Riis and Pedersen (2001) describe the benefits of combining data handled from OLAP and XML tools. Examples of queries to OLAP cubes, that may have had their dimensions complemented with new information acquired in XML files, are also presented. This addition is made using links from the OLAP cube to XML files. The liking process between OLAP and additional data in XML builds a logical integration, avoiding the reprocessing effort by the inclusion of external XML data.

Hümmer, Bauer and Harde (2003) define XCube, a Data Warehouse metamodel for XML documents, in which data schemas, based on XML Schema, were used to dimension, fact and cube representations. This approach has the advantages of a standardized environment, making the document to be reused more easily (e.g. dimension documents) for different domains. It also allows the integration with Web Services, as well as the insertion of comments in almost all elements of the documents, in different languages and with terms of specific areas of human knowledge. XCube

solves part of the heterogeneity problems in XML: (i) document structure, which is defined in schemes based on XML Schema; (ii) syntactic difference of content, which is partially solved through the units attribute to represent the same information in different units. Nevertheless, the treatment of elements with similar contents, written in different languages, is not discussed; and (iii) the semantic heterogeneity is not approached in XCube.

Hernández-Roz and Wallis (2006) specified XBRL Dimensions to model hypercubes of financial data. This model is based on a vocabulary definition, specified through XML Schema, and relationships based on XLink. These relationships express XML elements' hierarchical structure, dimensions and its members, thus it deals with the structural heterogeneity issue. Another type of relationship found in XBRL allows the creation of labels in several languages for each vocabulary element, solving the semantic heterogeneity problem. Syntactic heterogeneity is solved through the use of a *unit* attribute to define data measurement unit and through the definition of identical labels for elements that represent the same information written in different languages. However, this solution is limited to financial data representation.

Even though these works discuss Data Warehouse models for XML, they differ, for several reasons, from the data metamodel definition specified in this paper. In Boussaid et al. (2006), a Data Warehouse system based on XML architecture is presented but it lacks a detailed multidimensional data model. Pokorny (2001) and Golfarelli et al. (2001) specify the multidimensional model through software applications. Papers from Baril et al. (2003), Gottlob, Koch and Pichler (2003), Pokorny (2001) and Trujillo et al. (2004) are based on DTD or on the object-oriented paradigm. Finally, there are proposals of logical integration with the relational model: Baril et al. (2003), Jensen et al. (2001), and Pedersen et al. (2001). Besides, none of these papers have considered the use of XLink to define dimensional structures for XML.

The works that can be compared to the data metamodel presented here are XCube and XBRL Dimensions. Although the first one has not considered the use of XLink, it represented the cube, the dimensions and the facts, using distinct schemas. The second one, which is the only paper from the evaluated ones that uses XLink for the definition of the multidimensional model, is a restricted solution to a specific domain.

Gotlob et al. (2001) define a data model based on XML documents and a set of binary relations to propose an algorithm that evaluates XPath (XPath, 2007) expressions and optimizes the queries on these documents regarding time and storage space needed to perform them. XML document is described as a non-classified tree, i.e., a tree with an arbitrary number of children, ordered and labeled, in which each child node is ordered, and each node has a label. The document tree is represented by a set of binary relations, of which axes are the ones from XPath language (e.g. self, child, parent, descendent). As a result of that, the defineddata model allows the navigation in XML documents, performed by XPath language, which is the core mechanism for XML nodes addressing of other technologies, such as XQuery (XQuery, 2007) and XPointer (Grosso, 2003). Motivated by the use of XML applications, Barceló & Libkin (2005), Libkin & Neven (2003) and Libkin (2006) analyze query languages for XML trees, based on the same document definition given before, and present a group of definitions to handle XML data. These authors refer to several other proposals that consider naturally modeled XML data as non-classified trees, and also conceptualize an XML document. Boussaid et al. (2006) propose a technology that specifies data warehouses for star and snowflake

models logical definition which the data model is based on a mathematic formalization performed through XML Schema.

The concepts presented by Barceló et al. (2005), Boussaid et al. (2006), Gottlob et al. (2003), Libkin et al. (2003) and Libkin et al. (2006) define the XML document, navigation functions and a formalization for data warehouse typical models, i.e., star schema and snowflake schema. These definitions are taken into account in this paper for the multidimensional metamodel based on XML Schema and XLink. They were extended to include the relations among two or more documents through links. New definitions are given to represent the existence of relationships between XML documents.

In this paper, the proposal: (i) has a metamodel that solves the heterogeneity problems in XML data; (ii) uses linkbases, sets of links, for the definition of relationships among XML document elements, in order to specify the possible data cubes that can be used; (iii) is a metamodel that can be used in a variety of domains. In the next section, our data cube metamodel, which considers the use of XML, XML Schema and XLink technologies, is presented.

3.   A MULTIDIMENSIONAL METAMODEL BASED ON LINKS

This section shows a multidimensional metamodel for applications that uses XML as source of information. Initially, mathematical definitions are given, which allow a non-ambiguous metamodel specification (see Section 3.1). Then the set of XMLdocuments, based on XML Schema and XLink, that compose the XLDM specifications proposed here are discussed in Section 3.2.

3.1   Formal Definitions for XLDM

Definition 1: A rooted tree, ordered and non-classified, is a tree with unlimited child quantity, in which each node is given a unique label, and that is an N* element (i.e. a finite string of natural numbers). Then, a rooted, ordered, labelled and non-classified tree T is defined as $(D, <_{pre})$ in which:

1. The element $\varepsilon \in D$ (an empty string) is the root;

2. D is a set of nodes, named tree domain, which is a subset of N^{*}, such $g \in D$, implies $b \in D$, if, and only if, $b <_{pre} g$. The relation $<_{pre}$ defines the order of the document, i.e., it is the pre-fixed relation with the elements of D, being $b <_{pre} g$ if, and only if, the only way from the root to g goes through b.

Besides the relationship between nodes, in a given XML document, the relationships between nodes of two different documents are also defined. That is why the inclusion of Rσ is necessary in the XML document definition, which is a data structure that can be defined as follows:

Definition 2: An XML document *d* is represented as a 5-tuple $(T, \beta, \lambda, R\chi, R\sigma)$, in which:

1. $T = (D, <_{pre})$ is the rooted, ordered, labeled and non-classified tree;

2. β is a set of tags (XML elements);

3. $\lambda: D \rightarrow \beta$ is a function that assigns a node in *T* on each XML tag;

4. $R\chi$ is a set of binary relations on β, e.g. parent, child and sibling;

5. $R\sigma$ is a set of binary relations on $\beta' \times \beta''$, where β' and β'' are groups of tags from distinct XML documents, d' and d'', respectively, with $d' \neq d''$, $d' \subset D$ and $d'' \subset D$.

Definitions 1 and 2 are exemplified in Figure 1, assuming that the trees shown in this figure represent the documents d' and d''. The nodes ε and b are elements which have a binary relationship $\chi_{\varepsilon-b} \in R\chi$, such as *parent-child relationship*, and among the documents there is a relationship $\sigma_{b-a'} \in R\sigma$, established by the elements b and a'.

Definition 3: Taking into account that d is an XML document $(T, \beta, \lambda, R\chi, R\sigma)$, in a relation to $\chi \in R\chi$, and $\rho(\beta)$ as a subset of β, the function $f_\chi: \rho(\beta) \to P(\beta)$ is defined, with $f_\chi(X) = \{y \in \beta \mid \exists x \in X,$ so that $(x,y) \in \chi\}$. Then the relation name could be overridden, such as f_{child}.
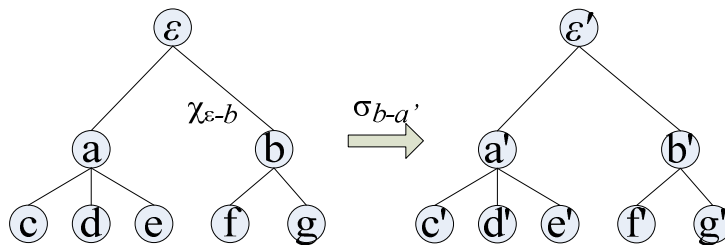
Figure 1. Relationships between the documents d´ and d´´ and among their elements.

Definition 4: Being d' and d'' two XML documents $(T', \beta', \lambda', R\chi', R\sigma')$ and $(T'', \beta'', \lambda'', R\chi'', R\sigma'')$, respectively, for a relation $\sigma \in R\sigma'$ in d', the function $g_\sigma: P(\beta') \to P(\beta'')$ is defined by $g_\sigma(X) = \{y \in \beta'' \mid \exists x \in X,$ so that $(x,y) \in \sigma\}$.

An XML document consists of element structures, which contain sub-elements and attributes. The attributes are added to the elements in opening declarations (tag). Between an opening and a closing tag, there may be any number of sub-elements. The attributes can be used to make reference among elements or between elements and other XML documents. According to these properties, the following definitions are used to represent the data cube metamodel proposed in this paper.

Definition 5: Being (F,S) a data warehouse (DW) schema, where F is a set of facts having m measures, $\{F.M_q, 1 \leq q \leq m\}$, and a set of independent dimensions r, $S = \{S_s, 1 \leq s \leq r\}$, where each S_s contains a group of I_j domains, $\{S_s.I_j, 1 \leq j \leq i\}$ and each I_j contains a group of n members, $\{I_j.N_p, 1 \leq p \leq n\}$. The (F,S) schema is composed of schemas and link bases documents:

1. F defines a set of fact tags;
2. S defines a set of dimension tags;
3. I defines a set of domain tags of a dimension;
4. N defines a set of member tags of a domain;
5. H defines a set of hypercube tags;
6. M defines a set of measures for a fact;
7. L is a set of link bases documents, represented as a 5-tuple $(T, \beta, \lambda, R\chi, R\sigma)$;
8. $\forall s \in \{1,...,r\}$, S_s defines elements associated to facts $f \in F$;

9. $\forall s \in \{1,...,r\}$ and $\forall i \in \{1,..., i_j\}$, $S_s.I_j$ defines relationships between dimensions and domains.

10. $\forall i \in \{1,...,i_j\}$ and $\forall n \in \{1,..., n_p\}$, $I_j.N_p$ defines relationships between domains and members.

As the definition of XLink allows types of relationships between elements, by the use of the attribute *xlink:arcrole*, this is used for the definition, considered below, of relationships between elements representing members, domains, dimensions, facts and cubes.

Definition 6: Considering $l \in L$, a linkbase$(T,\beta,\lambda, R\chi, R\sigma)$, the following relations are defined in l:

1. domain-member: $\forall n \in N$, $f_\chi(n) = \{i \in I \mid \exists n \in N$, so that $(n,i) \in \chi\}$;

2. dimension-domain: $\forall i \in I$, $f_\chi(i) = \{s \in S \mid \exists i \in I$, so that $(i,s) \in \chi\}$;

3. hypercube-dimension: $\forall s \in S$, $f_\chi(s) = \{h \in H \mid \exists s \in S$, so that $(s,h) \in \chi\}$;

4. *all* and *not-all*: $\forall n \in N$, $f_\chi(n) = \{f \in F \mid \exists n \in N$, so that$(n,f) \in \chi\}$;

Figure 2 shows how the relationships defined by the Definition 6 can be established. Each circle represents a node, which corresponds to an XML element. The lines that connect the nodes represent the possible relationship types that can exist between the elements. The *domain-member* relationship connects elements that are part of a domain. The *dimension-domain* relationship links the domain to the dimension. The *hypercube-dimension* relationship connects the fact to the dimension. Lastly, the relationships *all* and *not-all* state, for a given fact, whether all domain members are part of the hypercube or not. These relationships are established in the XLDM metamodel through link bases.
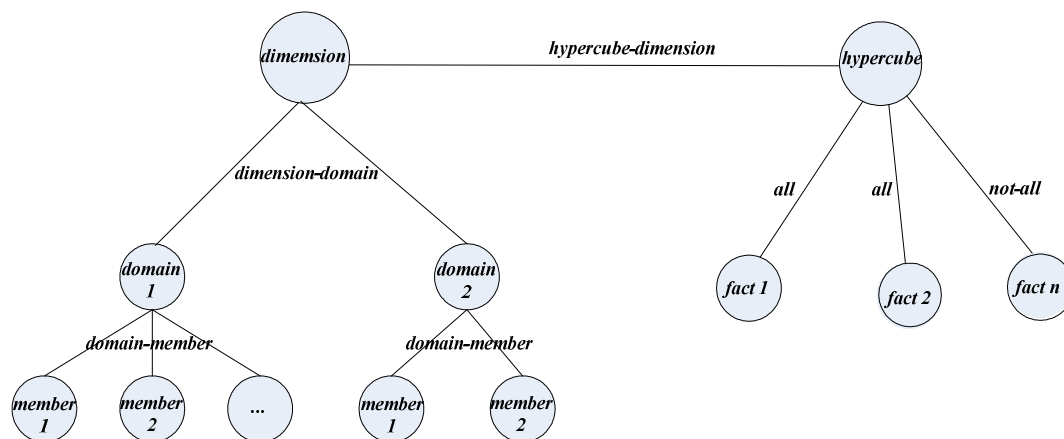
Figure 2. Relationships between the trees that represent the documents d´ and d´´.

Definition 7: Allow d to be an XML document $(T,\beta,\lambda, R\chi, R\sigma)$. A hypercube is defined as $\{\forall h \in H \mid H \subseteq d\}$ and $\{\forall f \in F \mid f \subseteq d\}$, and there is a function $f_\chi(f) = \{h \in H \mid \exists f \in F$, such that $(f,h) \in \chi\}$;

The XML formalism allows the insertion of sub-elements multi-levels in an XML element and the establishment of relationships, through XLink, which define hierarchies among elements. Thus, the definition of the *domain-member* relationship allows the building of hierarchies in a dimension. For example, one dimension *country* can have a *domain-member* relationship with the element *brazil*, then, this one can

have the same kind of relationship with other elements, e.g. b*razil and south*, b*razil and northeast.* Based on the definitions discussed in this section, documents that compose the XLDM metamodel specification were created. Next, the XLDM document specifications are discussed.

3.2   **XLDM Documents**

It is noticed that, due to the inherent XML technology flexibility, different data cube metamodels for data warehouses based on XML can be specified. For this reason, XML data heterogeneity problems are made evident. However, the use of XLink and XML Schema can solve such problems through the specification of dimensions, facts and cubes. The proposed multidimensional metamodel is based on the definitions presented in section 3.1. To do so, *instance-schema.xsd* and *linkbase-schema.xsd* documents have been specified based on these definitions and are available at http://www.cin.ufpe.br/~pcs3/XLDM/Spec.

An XML database that uses XLink is made of schemas, linkbases and XML instances, i.e.; XML documents with the data. The schemas specify the elements that represent the facts, the dimensions, the dimension members and the cubes. The linkbases define the relationships between members, dimensions and facts, establishing combinations of cubes that can exist. In the instance, the facts occur and, combined with dimension members, determine a data cube.

The XML instance document, which may contain one or more cubes, has a structural dimensionin which the contexts are presented with the dimension members. There is also a non-dimensional structure, with the measures of the facts.Figure 3 shows the UML components diagram (Unified Modeling Language, 2005) for the XLDM metamodel proposed in this section. This figure illustrates the data organization according to this model. Based on XML Schema, the vocabulary, i.e. a set of elements to be used in the XML instance, is specified. The relationships among the instance elements and between them and other resources are expressed in linkbases. The specified data types are common to a variety of domains. This was done in order to broaden the model applicability. However, it is possible to create types for a specific domain.

The attributes and elements declarations, which can be found in XLDM instances, are made in *instance-schema.xsd*. An element declaration of particular importance, the instance root element, is shown in Listing 1. The presence of this element and its children in the instance is based on Definition 1, 2 and 3. Initially, the element naming is performed. In XLDM, two alternatives are given to this element identification, in order to name it according to the domain to which it is being applied: (1) changing its declaration in the *instance-schema.xsd* document; (2) without changing its declaration in the *instance-schema.xsd* document, a label can be created for this element, specifying it in the link base Label and creating a relationship between two documents ($R\sigma$, Definitions 2 and 4). Thus, the instance *xldm* element has a label, for domain application, specified on the link base label. Next, there is the element description. After this, the declarations of the references to schemas, link bases, roles and arcroles are performed. To do so, *schemaRef*, *linkbaseRef*, *roleRef* and *arcrole Re f*elements were specified to allow the establishment of relationships among documents ($R\sigma$) as shown in Definition 2 and 4. A significant characteristic is the obligatoriness of two link bases (entitled Definition and Label, which will be discussed later). Thus, the

number of minimum occurrences in the *linkbaseRef* element is defined as two. Finally, the elements that can occur in the instance are declared, such as *item* and *tuple*. *Fact Schema* and *Hypercube Schema* are created from the *instance-schema.xsd* definitions, which define, respectively, the elements that represent the facts and the dimensions members. Regarding organization purposes, these schemas can be specified in the same document or in distinct documents. In the instance documents, it is mandatory the presence of *contextRef* and *unitRef*attributes in the elements that represent the facts.They make reference to the dimensional context and the fact unit, represented by the elements *context* and *unit*, found in the multidimensional structure. The declaration of these two attributes and elements establishes the binary relations ($R\chi$) described in Definitions 2 and 3 and its use is illustrated at the extract of the XLDM instance document show in Listing 11.
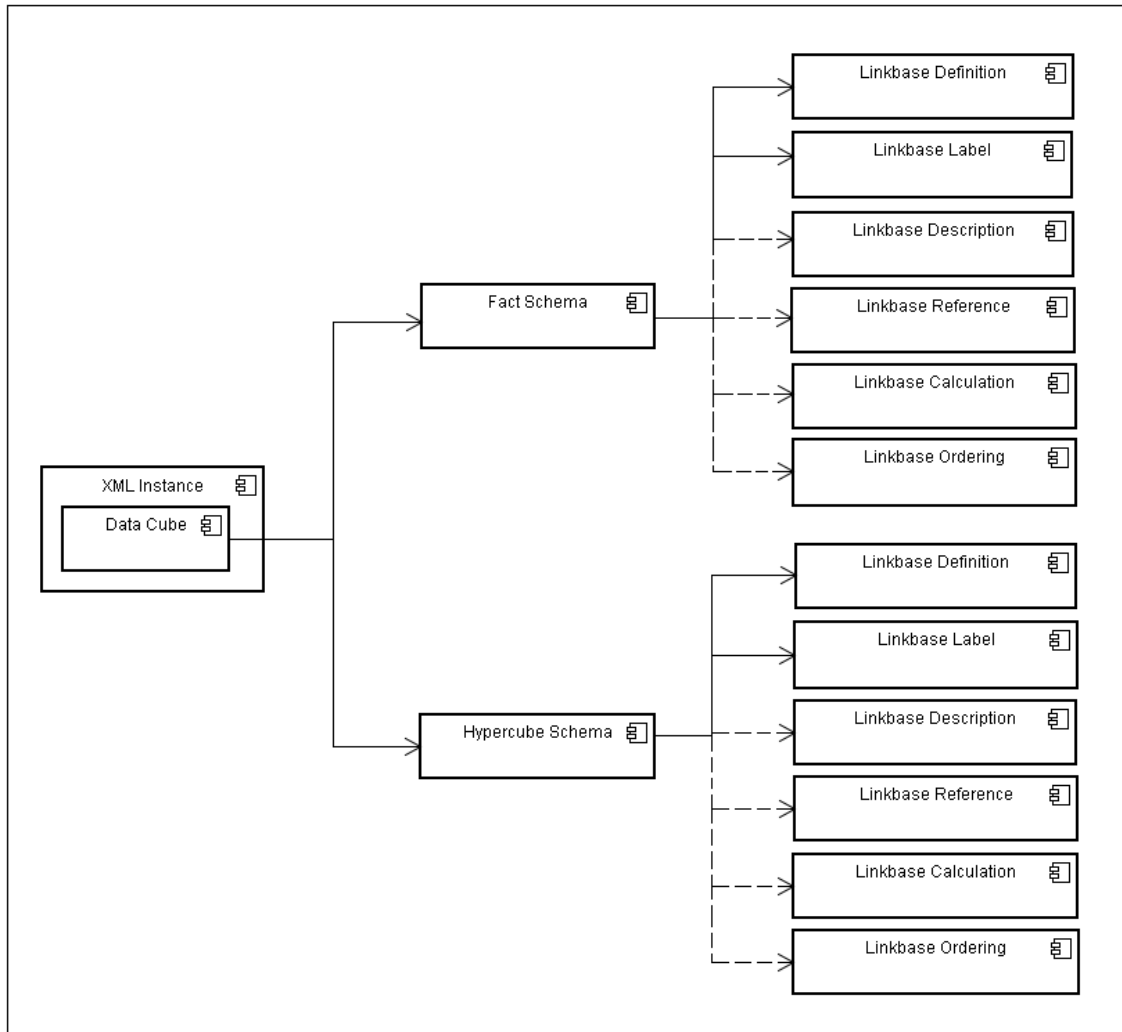
Figure 3.The XLDM Multidimensional Metamodel.

Listing 1. Definition of an XLDM Root Element

```
<element name="xldm">
      . . .
<complexType>
<sequence>
<element ref="link:schemaRef" minOccurs="1"
maxOccurs="unbounded"/>
<element ref="link:linkbaseRef" minOccurs="2"
maxOccurs="unbounded"/>
<element ref="link:roleRef" minOccurs="0"
maxOccurs="unbounded"/>
<element ref="link:arcroleRef" minOccurs="0"
maxOccurs="unbounded"/>
<choice minOccurs="0" maxOccurs="unbounded">
<element ref="xdmi:item"/>
<element ref="xdmi:tuple"/>
<element ref="xdmi:context"/>
<element ref="xdmi:unit"/>
<element ref="link:footnoteLink"/>
</choice></sequence>
  </complexType></element>
```

Linkbases are defined to conform to the relationships that can be present in a great variety of domains. The definitions occur by adding roles specification for the *arcrole* attribute, besides the inclusion of elements and attributes. Listing 2 shows the arcrole *multiplication-item* specification and, in Listing 3, the Description linkbase

definitionis illustrated. XLDM proposed linkbases willbe discussed next and their uses are illustrated in Section 4.

Listing 2. The multiplication-item Arcrole definition

```
<arcroleType id="multiplication-item"
cyclesAllowed="undirected"
arcroleURI="http://www.example.br/arcrole/multiplication-item">
<definition>Target (a primary item declaration) is a multiplication
factor which composes the value of the source (a primary item
declaration)</definition>
<usedOn>calculationArc</usedOn>
</arcroleType>
```

Listing 3 The Description Linkbase definition.

```
<element name="descriptionlLink"
substitutionGroup="xl:extended">
<annotation>
<documentation>
descriptionLinkelementdefinition.
</documentation>
</annotation>
<complexType>
<complexContent>
<restriction base="xl:extendedType">
<choiceminOccurs="0" maxOccurs="unbounded">
<elementref="xl:title"/>
<elementref="link:documentation"/>
<elementref="link:loc"/>
<element ref="link:descriptionArc"/>
</choice>
                        . . .
</complexContent>
</complexType>
</element>
```

1. The Definition linkbase, considered mandatory, allows the creation of the hierarchical element structure. This solves the problem of structural heterogeneity. For the structural aspect, the relationship definition among dimensions, members and cubes is performed by the Definitionlinkbase. The relationship is expressed by the arcrole *domain-member* lists and the possible members of a domain, which is associated to the dimension through the arcrole *dimension-domain*. The cross product between the dimensions and the facts, in order to establish the possible cubes to be used in the instance, is defined by the arcrole *hypercube-dimension*. To include the measures in the cube, the arcrole *all* is to be used and to exclude the member of a domain in a cube specification, the arcrole*notAll* is also used. These relations, based on the Definitions 5, 6 e 7, determine the relationship among members, dimensions, facts, cubes and possible values for the attribute *arcrole*, so that a data multidimensional model may be created. These relationships are shown in Figure 2. Besides these relationships, illustrated in Listing 6, other relationships are also specified. Their definition occurs with the following arcroles: (a) *main*, which defines a relationship between a concept and another as main, e.g. in a model for a disease treatment, a medical procedure is defined as the main one for a disease treatment; (b) *secondary*, which defines a relationship as secondary. In the disease treatment example, there may be the main procedure and the secondary ones. Then, the attribute *order* indicates the order in which the secondary procedures should be performed; (c) *substitution*, which determines the possibility of substitution of a concept for another one, e.g., a medical procedure can be replaced by another one. The attribute *order* indicates the order in which the concepts can be replaced, e.g., it defines the order in which the procedures replace the main one;

2.   The Label linkbase allows the use of different labels for the same element, which can be specified in different languages through the attribute *xml:lang*. This is mandatory, so that the semantic and syntactic heterogeneities can be avoided;

3.   The Ordering linkbase is an optional linkbase defined to determine not only the elements' presentation order in the instance, but also their processing order, which can differ from the presentation. For the definition of links aiming at specifying the presentation orders, an extended link element *presentationLink* is used. For processing purposes, the element *processingLink* is used;

4.   The Description linkbase is another optional linkbase introduced in this metamodel in order to supply textual description to a relationship. For example, the relationship between a disease and its description is represented by these linkbase arcs. Placing the descriptions in a different linkbase contributes to the model modularity;

5.   The Reference linkbase is also optional and it is used to define elements that represent references;

6.   The Calculation linkbase expresses arithmetic relations. It was defined so that, besides the sum operation, the arithmetic operations of multiplication, division, exponentiation and n-th root are specified. To do so, values are defined for the *arcrole* attribute. Listing 4 shows the possible arcroles for this linkbase, which depends on the arithmetic operation. Attributes are also defined, with proper domains, for each kind of operation. For example, the attribute *weight* changes its domain based on the operation. For sums, the domain varies from -1 to 1, what means that its value is completely or partially used in the addition, which results in the parent element value. This attribute domain for multiplication is the set of real numbers. Regarding the exponentiation and n-th root operations, there are the attributes *exponent* and *index*, of which domain is the set of natural numbers. For division, only the *arcrole* values are used in the numerator and denominator specification.

By using XLDM, the heterogeneity questions of the XML data, mentioned in section 1, are solved as follows: (i) in semantics, the Label linkbase establishes one or more names for an element defined in the schema. Therefore, a unique element can have several names, and distinct elements in different domains can have the same name. This allows an application to perform the processing through the element itself or through its label; (ii) in syntactic, the Label linkbaseallows the definition of names in different languages for the same element and the attribute *unit* allows to inform the unit referring to the measure; and (iii) in the structural, the schema defines the elements, their attributes, and child elements. The Definition linkbase specifies the hierarchy among the elements, determining the XML document structure.

Listing 4. Attribute arcrole in Calculation Linkbase.

Attribute arcrole Name	Meaning

`<link:calculationArcxlink:type="arc"`
`  xlink:arcrole="http://www.example.br/arcrole/summation-item"`
`xlink:from="A"xlink:to="B" weight="1.0"/>`
	The total value of concept B contributes for the formation of concept A.

`<link:calculationArcxlink:type="arc"`
`  xlink:arcrole="http://www.example.br/arcrole/multiplication-item"`
`xlink:from="A"xlink:to="B" weight="3.0"/>`
	Three times the value of B is a multiplication factor to compose the value of A.

`<link:calculationArcxlink:type="arc"` `  xlink:arcrole="http://www.example.br/arcrole/numerator-item"`	The value of B is the numerator of the

|	
`xlink:from="A"xlink:to="B"/>`	division that compose the value of A.

`<link:calculationArcxlink:type="arc"`
`  xlink:arcrole="http://www.examplo.br/arcrole/denominator-`
`item"`
`xlink:from="A"xlink:to="C"/>`
	The value of C is the denominator of the division that composes the value of A.

`<link:calculationArcxlink:type="arc"`
`  xlink:arcrole="http://www.example.br/arcrole/exponentiation-`
`item"`
`xlink:from="A"xlink:to="B" expoent="3.0"/>`
	The value of A corresponds to the value of the third exponent of B.

`<link:calculationArcxlink:type="arc"`
`  xlink:arcrole="http://www.example.br/arcrole/nthroot-item"`
`xlink:from="A"xlink:to="B" index="2.0"/>`
	The value of A corresponds to the n-th root of B.

4.   CASE STUDY

In order to demonstrate the applicability of the proposed metamodel, a case study is presented to apply XLDM to medical data represented in XML documents. At www.cin.ufpe.br/~pcs3/XLDM, a different example, dealing with sales, is also available.

4.1   XLDM Application

Figure 4 shows UML components diagram for the data cube model used in this case study. The *TreatmentCube* hypercube has four dimensions: *Patient, Procedure, Medication and Disease*. They relate to the hypercube through the arcroles *hypercube-dimension*. There is also the relationship of the cube with the *Dosage* measure, made with the use of the arcrole *all*. The schema created is shown in Listing 5. It contains the definition of the data cube *TreatmentCube*, the dimension *PatientDimension*, the domain *PatientDomain,* and a member of this domain (*Patient1*). Finally, the specification of the *Dosage* measure is made. The use of the attribute *abstract* in some elements indicates that these are used only for structural organization purposes and it is not possible to be used in the instance.

Listing 5. Schema Document

```
<element id="TreatmentCube"
name="TreatmentCube"type="xldmi:stringItemType"abstract="true"
substitutionGroup="xldmdt:hypercubeItem"/>
<element id="PatientDimension"
abstract="true"name="PatientDimension"type="xldmi:stringItemType"
substitutionGroup="xldmdt:dimensionItem"/>
<element id="PatientDomain" name="PatientDomain"
type="xldmi:stringItemType"abstract="true"
substitutionGroup="xldmi:item"/>
<element id="Patient1" name="Patient1" type="xldmi:stringItemType"
substitutionGroup="xldmi:item"/>
<element id="Dosage" name="Dosage"
type="xldmi:stringItemType"substitutionGroup="xldmi:item"/>
```
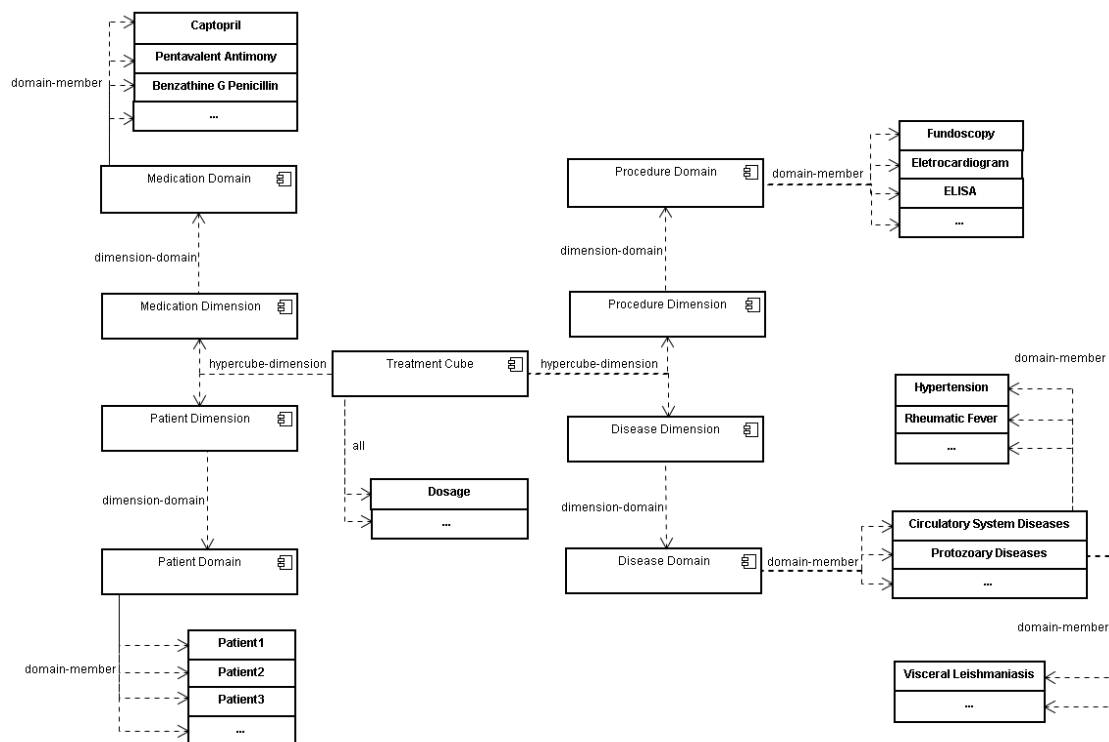
Figure 4.The Data Model for the Treatment Cube.

Listing 6 illustrates some elements of the Definition linkbase, in which the definitions of the hierarchic relations occur. The arcrole all is used for the relationship between the TreatmentCube cube and the Dosage measure. The dimensions are linked to the cube through the arcrole hypercube-dimension. The dimensions also relate to its domains through the arcrole dimension-domain. Finally, the arcrole domain-member provides the representation of the hierarchies in the dimensions.

Listing 6. Definition Linkbase

```
<definitionLinkxlink:type="extended"
xlink:role="http://www.example.br/xldm/treatment">
<locxlink:type="locator" xlink:label="lbl_TreatmentCube"
xlink:href="treatment.xsd#TreatmentCube"/>
<locxlink:type="locator" xlink:label="lbl_Dosage"
xlink:href="treatment.xsd#lbl_Dosage" />
<definitionArcxlink:type="arc" xlink:arcrole=http://www.example.br/arcrole/all
xlink:from="lbl_Dosage" ink:to="lbl_TreatmentCube"/>
<locxlink:type="locator" xlink:label="lbl_PatientDimension"
xlink:href="treatment.xsd#lbl_PatientDimension" />
<definitionArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/hypercube-
dimension"
xlink:from="lbl_TreatmentCube" xlink:to="lbl_PatientDimension"/>
<locxlink:type="locator" xlink:label="lbl_PatientDomain"
xlink:href="treatment.xsd#lbl_PatientDomain"/>
<definitionArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/dimension-
domain"
    xlink:from="lbl_PatientDimension" xlink:to="lbl_PatientDomain"/>
<locxlink:type="locator" xlink:label="lbl_Patient1"
xlink:href="treatment.xsd#lbl_Patient1" />
<definitionArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/domain-
member"
    xlink:from="lbl_PatientDomain" xlink:to="lbl_Patient1"/>
</definitionLink>
```

The Label linkbase, used to create labels for the members, can be seen in Listing 7. In this case, the label specifications are used to represent the medicine

commercial names and supply the ICD-10 code, an international standard for diseases. In this example illustrated in Listing 7, the attribute *lang* indicates that the drug name is specified in the English language. These relations use the arcrole *concept-label*.

Listing 7. Label Linkbase

```
<labelLinkxlink:type="extended"  xlink:role="http://www.example.br/xldm/treatment">
<locxlink:type="locator" link:label="lbl_Captopril"
xlink:href="treatment.xsd#Captopril" />
<label xlink:type="resource" xml:lang="en" xlink:label="lbl_Capitopril_Comercial"
xlink:role="http://www.example.br/role/label">
Capoten
</label>
<labelArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/concept-label"
   xlink:from="lbl_Captopril" xlink:to="lbl_Capitropil_Comercial"/>
</labelLink>
```

For the medical area, it is necessary to specify an order for performing certain procedures during the disease treatment. In the proposed data model, it is possible to express this by ordering representation using the Orderinglinkbase. This linkbase can be seen in Listing 8, which uses a processingLink element to define that the first element that should be processed is the cube, followed by the dimension, the domain and the member. This is done by the numerical value of the attribute *order*.

Listing 8. Ordering Linkbase

```
<processingLinkxlink:type="extended"
xlink:role="http://www.examplE.br/xldm/treatment">
<locxlink:type="locator"  xlink:label="lbl_TreatmentCube"
xlink:href="treatment.xsd#lbl_TreatmentCube" />
<locxlink:type="locator"  xlink:label="lbl_PatientDimension"
xlink:href="treatment.xsd#lbl_PatientDimension" />
<processingArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/parent-
child"
   xlink:from="lbl_TreatmentCube" xlink:to="lbl_PatientDimension" order="1" />
<locxlink:type="locator" xlink:label="lbl_PatientDomain"
xlink:href="treatment.xsd#lbl_PatientDomain" />
<processingArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/parent-
child"
   xlink:from="lbl_PatientDimension" xlink:to="lbl_PatientDomain" order="2"/>
<locxlink:type="locator" xlink:label="lbl_Patient1"
xlink:href="treatment.xsd#lbl_Patient1" />
<processingArcxlink:type="arc" xlink:arcrole="http://www.example.br/arcrole/parent-
child"
   xlink:from="lbl_PatientDomain" xlink:to="lbl_Patient1" order="3"/>
</processingLink>
```

The Description linkbase is used to provide textual descriptions related to a certain concept. Listing 9 shows the description made for an element that represents the disease Rheumatic Fever.

Listing 9. Description Linkbase

```
<descriptionLinkxlink:type="extended"
xlink:role="http://www.example.br/xldm/treatment">
  <locxlink:type="locator" xlink:label="lbl_Rheumaticfever"
  xlink:href="treatment.xsd#RheumaticFever"/>
<description xlink:type="resource"
xlink:role="http://www.example.br/role/description"
xlink:label="desc_RheumaticFever">
     Rheumatic fever is an inflammatorydisease that may develop two to three weeks
     after a Group A streptococcal infection such as strep throat or scarlet fever).
     It is believed to be caused by antibodycross-reactivity and can involve the
     heart, joints, skin, and brain. Acute rheumatic fever commonly appears in
     children ages 5 through 15, with only 20% of first time attacks occurring in
     adults
</description>
</descriptionLink>
```

Listing 10 shows the composition of a medicine, making evident the relationship between the formula components and the medicine by using the Calculation linkbase. This linkbase explains that the medicine Hydralazine is composed by 5% of Sodium Nitroprussiate and 25% of Isotonic Glucose Solution.

Listing 10. Calculation Linkbase

```
<calculationLinkxlink:type="extended" xlink:role="http://www.example.br/treatment">
  <locxlink:type="locator" xlink:label="lbl_Hydralazine"
xlink:href="treatment.xsd#Hydralazine" />
  <locxlink:type="locator" xlink:label="lbl_SodiumNitroprussiate"
  xlink:href="treatment.xsd#SodiumNitroprussiate" />
  <calculationArcxlink:type="arc"
xlink:arcrole="http://www.example.br/arcrole/summation-item"
  xlink:from="lbl_Hydralazine" xlink:to="lbl_SodiumNitroprussiate" weight="0.05"/>
  <locxlink:type="locator" xlink:label="lbl_IsotonicGlucoseSolution"
xlink:href="treatment.xsd#IsotonicGlucoseSolution"/>
  <calculationArcxlink:type="arc"
xlink:arcrole="http://www.example.br/arcrole/summation-item"
  xlink:from="lbl_Hydralazine" xlink:to="lbl_IsotonicGlucoseSolution" weight="0.25"/>
</calculationLink>
```

An extract of the instance document is illustrated in Listing 11, showing a data cube, where the context Patient1_Hypertension is defined. In this context, the dimension members PatientDimension, ProcedureDimension, MedicationDimension and DiseaseDimension are given. The temporal view of the fact is established by the element period, with sub-elements for the starting and ending date for which the fact is valid. A temporal hierarchy can be established by other sub-elements, e.g. elements for year, semester. The element *unit* defines the measure unit. In the example, the unit referring to the fact is milligrams per day, i.e. for *Patient1*, which is undertaking the fundoscopy treatment, during the period of January 1st, 2007 to December 31st, 2007, for the disease hypertension, the daily dose of the medicine Captopril is 50mg.

Listing 11. XLDM Instance

```
<context id="Patient1_Hypertension">
<entity>
<segment>
<xldmdi:explicitMember
         dimension="treat:PatientDimension">treat:Patient1</xldmdi:explicitMember>
<xldmdi:explicitMember

dimension="treat:ProcedureDimension">treat:Fundoscopy</xldmdi:explicitMember>
<xldmdi:explicitMember
dimension="treat:MedicationDimension">treat:Captopril</xldmdi:explicitMember>
<xldmdi:explicitMember
dimension="treat:DiseaseDimension">treat:Hypertension</xldmdi:explicitMember>
</segment>
```

```
</entity>
<period><startDate>2007-01-01</startDate>
<endDate>2007-12-31</endDate></period>
</context>
<unit id="mg_day"><divide>
<unitNumerator>mg</unitNumerator>
<unitDenominator>day</unitDenominator></divide>
</unit>
<treat:DosagecontextRef="Patient1_Hypertension" unitRef="mg_day">50</treat:Dosage>
```

5.   A COMPARATIVE ANALYSIS BETWEEN XLDM ANDXBRL DIMENSIONS

Since XLDM and XBRL Dimensions use the same technologies, XML Schema and XLink, it is important to highlight the differences between them. This section provides a comparison between an application of XMDL to financial indexes and XBRL Dimensions. Besides having a broader applicability than XBRL Dimensions, XLDM has also agreater expressivity and its use is easier. An example to show that the use of XLDM, in the financial field, can have more expressivity than XBRL is given by the use of the proposed arcroles for the Calculation linkbase. Listing 12 shows the specification of a financial index named *ExposureRiskIndex*. It is formed by the relationship between *ExposureValue* and *CreditRiskCapitalRequirements*. This element was extracted from the XBRL taxonomy of project COREP (Boixo& Flores, 2005), an initiative of CEBS (Committee of European Banking Supervisors) to provide a framework of financial reports for some institutions from the European Union. It is not possible to express this kind of relationship in XBRL Dimensions. It happens because there is no relationship for division operation on XBRL Dimensions.

In addition to that, byusing the Description linkbase, it is possible to give a description to a concept. In Listing 13, there is a part of the Description linkbase, describing the Exposure Risk Index concept. As a result, the XLDM contribution, in the financial field, extends the possibilities offered by XBRL Dimensions. Consequently, besides the proposed generalization, a wider expressivity is assured. A taxonomy for some financial indexes are available at www.cin.ufpe.br/~pcs3/XLDM/FinancialDataModel.

Listing 12.Financial Index Specification

```
<calculationLinkxlink:type="extended" xlink:role="http://www.example.br/financial">
  <locxlink:type="locator" xlink:label="lbl_ExposureValue"
xlink:href="financial.xsd#ExposureValue"/>
  <locxlink:type="locator" xlink:label="lbl_ExposureRiskIndex"
xlink:href="financial.xsd# ExposureRiskIndex"/>
<calculationArcxlink:type="arc"
xlink:arcrole="http://www.example.br/arcrole/numerator-item"
xlink:from="lbl_ ExposureRiskIndex" xlink:to="lbl_ExposureValue"/>
  <locxlink:type="locator" xlink:label="lbl_CreditRiskCapitalRequirements"
  xlink:href="financial.xsd#CreditRiskCapitalRequirements"/>
<calculationArcxlink:type="arc"
xlink:arcrole="http://www.example.br/arcrole/denominator-item"
xlink:from="lbl_ExposureRiskIndex"    xlink:to="lbl_CreditRiskCapitalRequirements"/>
</calculationLink>
```

Listing 13. Debt Composition Index Description

```
<descriptionLinkxlink:type="extended"
xlink:role="http://www.example.br/financial">
<locxlink:type="locator"xlink:label="lbl_ExposureRiskIndex"
xlink:href="financial.xsd#ExposureRiskIndex"/>
<description xlink:type="resource" xlink:role="http://www.example.br/role/description"
xlink:label="desc_ExposureRiskIndex">
   <desc:Description>Ratio between Exposure Value andCredit Risk Capital
Requirements.
```

```
</desc:Description>
</description>
<descriptionArcxlink:type="arc" xlink:arcrole=http://www.example.br/arcrole/concept-
description"
xlink:from="lbl_ExposureRiskIndex" xlink:to="desc_ExposureRiskIndex"/>
</descriptionLink>
```

A XLDM application can be found in the LMDQL (Link-based and Multidimensional Query Language) (Silva & Times, 2009). LMDQL is a language that has operators for multidimensional analysis of data represented in XML documents interconnected through XLink. LMDQL has an operator, *OperatorDefinition,* which allows the users to create new operators, through mathematical relation sexpressed in the linkbase Calculation. In operator's creation through *OperatorDefinition*, are generatedthe*linkbase Calculation, Definition*and *Label*and the operators scheme (XML *Schema*). In such cases, the solution for the problems of data heterogeneity in XML is also reached. The operator specification is represented on *linkbase Calculation* to define the arithmetic relation that constitutes the new operator. LMDQL language processor was incorporated into the OLAP server *Mondrian* (Mondrian, 2008) so that the analytical queries could be performed based on this languagein XML documents. After this implementation, the creation and use of the operator, that represents the index, *ExposureRiskIndex* were possible. Figure 5 illustrates this operator creation by LMDQL operator*OperatorDefinition*and Figure 6 demonstrates its use. Figure 7 shows *linkbase Calculation*, stored on SGBD DB2 Express C (IBM – DB2 Express C, 2006), generated by the operator *OperatorDefinition*, which represents the index created.

6.   CONCLUSION

The metamodel proposed in this article solves the heterogeneity problems of XML data through the specification of data schemas, by usingXML Schema, and the relationships between them, through linkbases. The specification of this metamodel allows its use in different domains, because they have linkbases that determine common relationships with several knowledge areas, e.g. ordering relation, hierarchy, element naming, description and reference. For arithmetic relationships, the Calculation linkbase comprises all kinds of basic arithmetic operations, making various mathematic expressions possible. An important characteristic is that the metamodel, being based on XLink, can be extended for representing relationships that are not predicted. For this reason, XLDM makes the development of processing XML data tools, that use XLink (Silva & Times, 2009), (Silva, Santos and Times, 2010), easier. This paper presents the formalism for XML cubes based on XLink, thus, allowing the specification of the proposed metamodel based on a set of XLDM documents definitions.

The case study shows its applicability in the medical and financial area. Another application of the metamodel XLDM was made in the field of sale organization, which can be seen in http://www.cin.ufpe.br/~pcs3/XLDM/foodMartXML. For further proposals, it is intended to use this metamodel in other domains, and to validate it in other contexts with the LMDQL language, developed for the analytical processing of XML data that uses XLink. A CASE tool that uses XLDM to define data models for specific domains can also be seen as another indication of future proposals of work.

XLDM: Na Xlink-Based Multidimensional Metamodel

Operator Definition

```
with operatordefinition('[Measures].[ExposureValue]/[Measures].
[CreditRiskCapitalRequirements]',ExposureRiskIndex)
```

Create Operator Definition

List of Operator Definition

```
#ExposureRiskIndex# --> member [Measures].[ExposureRiskIndex] as '[Measures].[ExposureValue]/[Measures].[CreditRiskCapitalRequirements]'
#AcumulativeExposure# --> member [Measures].[AcumulativeExposure] as '1.00*[Measures].[ExposureValue]+1.00*[Measures].[CreditRiskCapitalRequirements]'
```

Figure 5.Index creation from the operator *OperatorDefinition*.

Ad-Hoc MDX Queries

```
select {[Measures].[CreditRiskCapitalRequirements],[Measures].
[ExposureValue],#ExposureRiskIndex#} ON COLUMNS,
 {([Exposure Class])} ON ROWS
from [Capital Requirements SA]
```

Process MDX query   Database: ○ Relational ● XML

Results:

	[Measures].[CreditRiskCapitalRequirements]	[Measures].[ExposureValue]	[Measures].[ExposureRiskIndex]
[Exposure Class].[All Exposure Classs]	480.00	234.52	0.49

Execution Time: 859 ms.
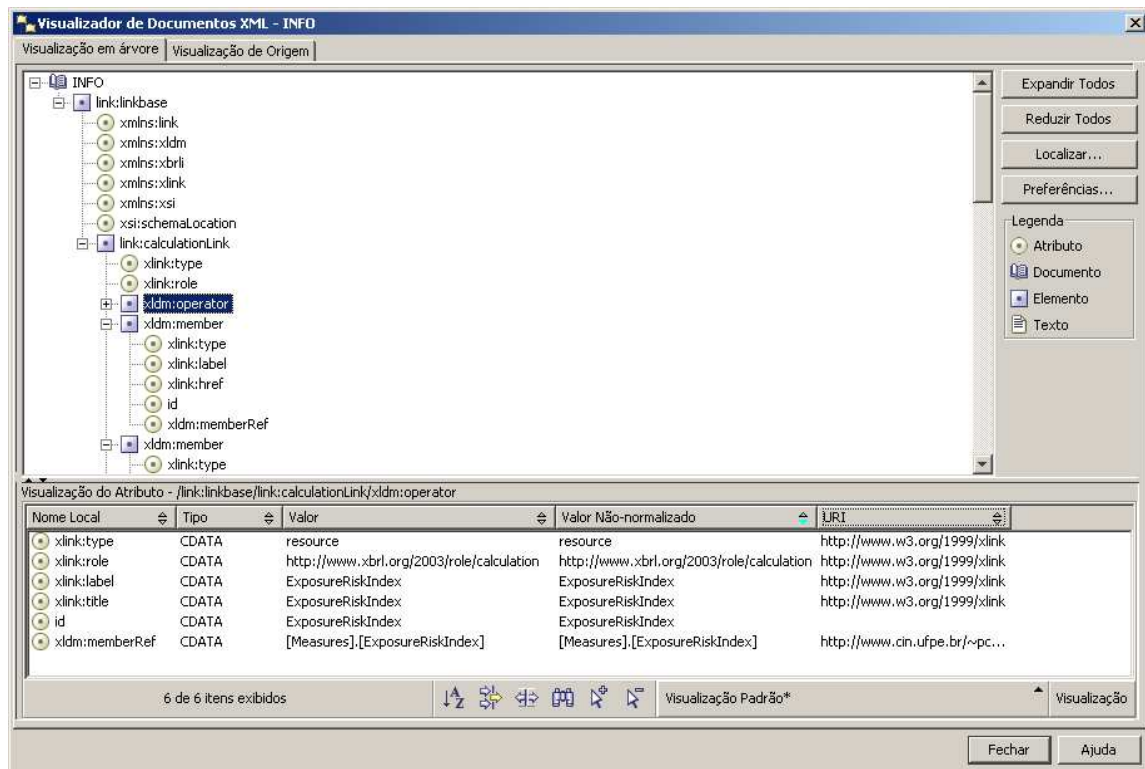
Figure 6.Use of the operator *ExposureRiskIndex*.

Figure 7.Linkbase Calculation, created by the LMDQL operator *OperatorDefinition*.

REFERENCES

Barceló, P. and Libkin, L. (2005).Temporal Logics over Unranked Trees.*Proceedings of the 20th Annual Symposium on Logic in Computer Science.*

Baril, X., Bellahséne (2003).*Z.: Designing and Managing an XML Warehouse. XML Data Management: Native XML and XML-Enabled Database Systems*. Addison Wesley Professional 455–474.

Boussaid, O., Messaoud, R. B., Choquet, R. and Anthoard, S. (2006). X-Warehousing: An XML-Based Approach for Warehousing Complex Data. *East-European Conference on Advances in Databases and Information Systems (ADBIS 06).*

Boixo, I.; Flores, F. (2005). New Technical and Normative Challenges for XBRL: Multidimensionality in the COREP Taxonomy. *The International Journal of Digital Accounting Research*. v. 5, n. 9, p. 79-104. ISSN: 1577-8517.

Golfarelli, M., Rizzi, S., Vrdoljak, B. (2001). Data Warehouse Design from XML Sources.*Proceedings of the 4th ACM International Workshop on Data Warehousing and OLAP* (DOLAP 2001), Atlanta, Georgia, USA, ACM Press 40–47.

Gottlob, G., Koch, C. and Pichler, R. (2003).XPath query evaluation: improving time and space efficiency.*19th International Conference on Data Engineering*.

Grosso, P. (2003).XPointer Framework W3C Recommendation. Retrieved from http://www.w3.org/TR/2003/REC-xptr-framework-20030325/

Hernández-Ros, I. and Wallis, H. (2006).XBRL Dimensions 1.0. Retrieved from www.xbrl.org/Specification/XDT-REC-2006-09-18.htm

Hümmer, W., Bauer, A., Harde, G. (2003). XCube – XML for Data Warehouses.*Proc.The 6th ACM Intl Workshop on Data Warehousing and OLAP*, p. 33–40.

IBM – DB2 Express C (2006). Retrieved from http://www-01.ibm.com/software/data/db2/express

Jensen, M. R., Moller, T. H. and Pedersen, T. B. (2001). Specifying OLAP Cubes On XML Data.*Technical Report 01-5003*. Department of Computer Science, Alborg University.

Kimball, R., Ross, M.(2002).The DataWarehouse Toolkit.John Wiley and Sons.

Libkin, L., Neven, F. (2003).Logical Definability and Query Languages over Unranked Trees.*LICS 2003*. Canada, Ottawa. IEEE Computer Society, p. 178-187.

Libkin, L. (2006). Logics for Unranked Trees: An Overview.*Logical Methods in Computer Science*, Vol. 2 (3:2) 2006, p. 1–31.

Mondrian (2008). Retrieved from http://mondrian.pentaho.org

Näppilä, T., Järvelin, K., Niemi, T. (2008). A tool for data cube construction from structurally heterogeneous XML documents.*Journal of the American Society for Information Science and Technology (JASIST)*, Vol. 59, Issue 3, p. 435-449.

Nassis, V., Rajugan, R., Dillon, T. S. and Rahayu, W. (2004).Conceptual Design of XML Document Warehouses.*Data Warehousing and Knowledge Discovery, 6th International Conference, DaWaK 2004*, p. 1–14.

Pedersen, D., Riis, K. and Pedersen, T. B. (2001). XML – Extended OLAP Querying.*Technical Report 02-5001*.Department of Computer Science, Alborg University.

Pokorny, J. (2001). Modeling Stars Using XML.*The 4th ACM Workshop on Data Warehousing and OLAP (DOLAP01)*, p. 24–31. USA, Atlanta.

Silva, P. C., Times, V. C. (2009). XPath+: A Tool for Linked XML Documents Navigation.*XSym 2009 - Sixth International XML Database Symposium at VLDB'09*.France, Lyon.

Silva, P.C., Times, V. C. (2009), LMDQL: Link-based and Multidimensional Query Language.*DOLAP 2009 - ACM Twelfth International Workshop on Data Warehousing and OLAP*.China, Hong Kong.

Silva, P.C., Santos, M. M., Times, V. C. (2010). XLPATH: XML Linking Path Language.*IADIS WWW/Internet 2010 (ICWI 2010) Conference*.

Trujillo, J., Luján-Mora, S., Song, I. (2004).Applying UML and XML for Designing and Interchanging Information for Data Warehouses and OLAP Applications.*Journal of Database Management* 15(1) 41–72.

Unified Modeling Language (2005). Retrieved from http://www.uml.org

XBRL International (2008). Retrieved from http://www.xbrl.org

XQuery 1.0: (2007). Retrieved fromhttp://www.w3.org/TR/xquery

XML Linking (2001). Retrieved fromhttp://www.w3.org/TR/xlink

XPath language (2007). Retrieved fromhttp://www.w3c.org/tr/xpath20

XML Schema (2004). Retrieved fromhttp://www.w3.org/TR/xmlschema-1