0. Table of content

0. Table of content ......................................................................................................................................... 1

1. Preface .................................................................................................................................................. 1

2. The management of knowledge and topic maps .................................................................................. 1

3. The XML Topic Maps namespace

   3.1 The composition of XML Topic Maps .......................................................................................... 4
   3.2 The XTM 1.0 specification ........................................................................................................... 5
   3.3 Creation and management of XML Topic Maps ........................................................................... 6

4. Topic maps for archival applications ................................................................................................... 8

   4.1 Application possibilities for electronic record-keeping ................................................................. 8
   4.2 Exemplary application .................................................................................................................. 9

5. Conclusion ............................................................................................................................................. 12

6. Sources .................................................................................................................................................. 12

1. Preface

The XML-technology is a giant leap forwards in the field of electronic record-keeping. But it isn’t the all-embracing solution for long term preservation of electronic records in a good, well organized and accessible state. Through DTD’s, XML Schemas and stylesheets, archivists dispose of a mechanism to present the information and knowledge which is being held in records in an explicit and durable way, but these technologies do not suffice on their own to present knowledge about the archives, their context and their content. After all, documents exist in many different types, formats and structures. The registration of knowledge on archives is, together with the need for a collective search mechanism, an important aspect in the logical or intellectual management of electronic records.

Recently, a new technology to describe archives on meta-level is being used: topic maps. A topic map serves as an application and device independent layer or map in which sources of information are linked to each other. Topic maps are used in an attempt to register knowledge for the future. In a topic map, knowledge is represented by structured relations between subjects and by clearly defining the nature of the relationship. However, the link with XML is obvious. One particular version of topic maps is an XML-namespace and uses the XML-grammar. This version of topic maps is called XML Topic Maps (XTM). Topic Maps connect very closely to the preservation of electronic records and are ideal for usage in combination with XML-documents. Moreover, Topic Maps are standardised.

This article is dedicated to XTM. The general introduction to topic maps is followed by a more specific review of this new tagging mechanism for knowledge. Subsequently the applicability of XML Topic Maps in electronic record-keeping is looked at. Finally a practical example of a topic map is given.

2. The management of knowledge and topic maps
The goal of knowledge management is to administer knowledge and make it more accessible so the user can find information swiftly and efficiently. It is important to put yourself in the place of the user and to assume that you do not know in advance in which document the wanted information is registered. A user can be interested for instance, in all the information pertaining to a certain topic, he can search for associated information or he wants to consult documents with a specific characteristic. An additional challenge is the distribution of knowledge a certain organisation: paper files, intranet applications, electronic documents, e-mail systems, databases, etc. After all, it’s necessary to provide an access to the wanted information. This means that a heterogeneous group of sources of information needs to be made accessible. Traditional management of knowledge tools such as indexes, wordlists and thesauri no longer suffice in the goal-oriented search for knowledge or in the providing of knowledge on the sources of information. The solution to this problem was found in the development of representation languages for knowledge such as topic maps. Other representation languages besides topic maps exist, but topic maps have the advantage of being an open standard.

Topic maps originated in the early 1990’s and have evolved into conceptual layers in which the knowledge on certain subjects is listed and described in a formal way. The subjects who are a part of topic maps and who are described within the topic map paradigm as “topic”, can be any given (“thing”) from reality: a book, a person, an object, an idea, an emotion, etc. Within the management of knowledge topics usually mean sources of information and documents. At that time a topic map is a coordinated layer in which knowledge on and from documents is registered. The topic map produces the knowledge that is contained in an entirety of documents. In its most simple form, a topic map can be compared to a subject index of an (electronic) publication. The index is a list of key words (topics), their occurrences (the pagenumber of the book in which the information is mentioned) and the mutual associations (the links that are being produced by “see references”). In more advanced applications, the relations are more important and a topic map provides access to a entirety of documents.

Although topic maps were initially meant to replace the classical search instruments, very soon it became clear that topic maps also offer possibilities to record knowledge in an electronic way and be processed by computers. Computers can only process sequences of characters and not unstructured or natural information.
Therefore it’s desirable to record electronically as much information as possible. But even then there is a big gap between reality and electronically registered characters which can only be overcome with knowledge of the underlying concepts. In the day-to-day life, this knowledge is often implicitly present with users of the information, but it’s important to register this knowledge explicitly with regards to record-keeping in the long run. With topic maps it’s possible to record the concepts which link the information to the knowledge. The first purpose of topic maps is not to provide the information itself, but rather the meaning and the connection between the information. Due to this, knowledge is registered for the future and is made processable for computers.

To represent knowledge in a topic map, not only are the topics indicated, the topics are also linked to each other and their association is described. Knowledge is translated into relations on a logical level in topic maps. For instance, a topic map can describe that the person Filip Boudrez is the author of the article on XML Topic Maps for digital archiving. The topics Filip Boudrez and XML Topic Maps are described (respectively ‘person’ and ‘article’) and linked. The two topics are described and the nature of their connection is reported. (‘is the author of’). On top of this, every topic can be linked to sources with more information on the topic (occurrences or ‘characteristics’). By describing topics and their relation, faster and more accurate search actions are possible because relations and their nature can also be used as a search condition. The indication of occurrences of topics also allows information to be found with the aid of certain characteristics. Two or more topics can be joined together (merging) so that all information on a certain subject can be found.

Topic maps go further in providing information and knowledge than, for instance, a classic index in which information is mainly provided through authority records (compare to search engines such as google). A number of typical full text problems are avoided with topic maps (homonyms, synonyms, indexing of non-textual documents, etc.) An other important difference to a full text-index is that the index data or metadata doesn’t need to appear in the records themselves. It is also interesting that a topic map provides different search- or navigation entries to retrieve the sought-after information. The user can gather the information he wants through several different search procedures and is therefore no longer obliged to use the logic of the inventor of the finding aid system.

Knowledge about the sources of information in topic maps is kept outside the information source itself. The topic map can be considered as a separate and coordinated macro-layer. After all, the topic map isn’t recorded inside the information source itself. This offers certain advantages. Firstly, multiple and even an unlimited number of topic maps can apply to the same information sources. Secondly, topic maps are autonomous and can be changed without the need to alter the sources of information. This also applies to the navigation routes within the existing topic maps. Thirdly, you don’t depend on the internal structure of the information sources to find information1. Finally, two or more topic maps can be joined together. This permits a modular composition of topic maps and an integration of information from several different sources of information.

The topic map paradigm was established by ISO as an official standard in January 2000. Within this standard, a uniform standard recording method was captured. The ISO standard uses SGML as a recording method for topic maps and their characteristics and HyTime2 for the links. When the official standard was established, the independent consortium TopicMaps.org put together a group of authors with the task to draft up XML Topic Maps (XTM). XTM should enable the web-implementation of Topic Maps and should be usable in combination with XML. XTM was published in 2001 and the XTM Document Type Definition was taken up as an annex in the ISO-Standard in October 2001. XTM uses XML instead of SGML as recording method and URL’s (Xlink) instead of HyTime for the reference3. XTM is in the public domain and is an XML-namespace which

---

1 Should all information sources be present in XML, then the goals of topic maps would be feasible to a certain degree on the basis of XML sources. The XML hierarchy and the internal links are usable to make documents accessible. Not all the sources of information are present in XML, so that transformations to XML are expensive. Besides, XML sources are limited, due to the statical and hierarchical navigation-system which furthermore is difficult controllable and scalable.

2 HyTime was registered in 1992 as ISO-standard. HyTime was developed to add multimedia- and hyperlink-functionalities to SGML. (ISO/IEC-10744(1992): Information technology -- Hypermedia/Time-based Structuring Language (HyTime))

3 http://www.topicmaps.org/xtm/
delivers the structure and the grammar for the description of knowledge on topics. With XTM-DTD one disposes of a standardised recording method for exchangeable, representative knowledge on information. Nowadays XTM is the most occurring topic map application and is supported by most of the topic map applications. SGML Topic Maps are rare and hardly ever have any software support. In May 2001, they started with the standardisation of a query language for topic maps within the core of the ISO. (TMQL: Topic Maps Query Language).

The choice for XTM instead of SGML topic maps or other knowledge representation languages is obvious. Other knowledge representation languages are difficult or not at all exchangeable and aren’t application independent. Topic maps do meet all these criteria and can be approached from several domains and platforms. In comparison to SGML-Topic Maps, XML Topic Maps are more easy to use and offer the possibility for webimplementation. An XML Topic Map can be made accessible through a browser so that navigation on the basis of links, is possible. The use of XTM in combination with electronic records in XML is of course also very important.

XML Topic Maps show some resemblances to the Resource Description Framework (RDF) of the W3C. Both of these concepts fit in with the initiative to build a semantic web and are useful in the information- and knowledge management. However, RDF is more specialised in the sources of information themselfs and less in the knowledge structures which surpass the level of the sources. With RDF for instance, metadata is attributed to documents. Topic maps and RDF are more complementary than concurrential. It’s possible that RDF and topic maps aren’t just applicable on the same documents or data, but can also exchange information with each other. Several initiatives are running at the moment to further integrate topic maps and RDF.

3. The XML Topic Maps namespace

3.1 The composition of XML Topic Maps

The cornerstones for topic maps are topics. Normally, their is a 1 on 1 relation between subjects or sources of information and topics. A topic has three characteristics: a name, an occurrence and a role. A topic has an internal ID, a basicname (the representation name), an optional variation name (an alternative for a basicname meant for specific use) and optional parameters. The ID and the basicname of a topic can be identical, but this isn’t necessary. The topics are linked to relevant internal or external sources with information on the subject. (f.i. image, video, book, webpage, etc.). These sources of information form together with the classification, the <occurrences> (characteristics') of the topics. External sources are described through <resourceRefs> and internal sources with <resourceData>. It’s being determined through the tag <instanceOf> to which type or class (<TopicType>) a topic belongs. The topic ‘e-mail’ for instance belongs to the class ‘electronic record’.

The topics are the buildingstones of the topic map, but the associations are the real purposes of topic maps. The knowledge is formally being registered in associations by relating topics to each other, by describing their relation and by describing the role of topics in the relation. The relations between topics are called “associations” and are indicated with the XML-element <association>. Within the association the type of association is firstly defined. Subsequently it’s being indicated which topics are members of the association and what their precise role is whithin the association.

Besides the topic-elements and the relations, the third piler of a topic map is the context in which the knowledge is valid. The context that determines in which perspective, which language, which version, etc. the topic map is true, is being indicated with the <scope> element.

---

5 [http://www.w3.org/RDF/](http://www.w3.org/RDF/)
6 The authority records in a classical are the topics, the pagenumbers, the occurrences and the view-references.
3.2 The XTM 1.0 specification

The XML Topic Map specification is an XML-namespace in which 19 elements are established which can be used when composing an XTM-document\(^7\).

**XTM DTD 1.0**

```xml
<!DOCTYPE topicMap [ @id ID #IMPLIED
xmlns CDATA #FIXED 'http://www.topicmaps.org/xtm/1.0/
xmlns:xlink CDATA #FIXED 'http://www.w3.org/1999/xlink'
xml:base CDATA #IMPLIED>

<!ELEMENT topicMap ( topic | association | mergeMap )*>

<!ATTLIST topicMap
id ID #IMPLIED
xmlns CDATA #FIXED 'http://www.topicmaps.org/xtm/1.0/
xmlns:xlink CDATA #FIXED 'http://www.w3.org/1999/xlink'
xml:base CDATA #IMPLIED>

<!ELEMENT topic ( instanceOf*, subjectIdentity?, ( baseName | occurrence )* )>

<!ATTLIST topic id ID #REQUIRED
instanceOf ( topicRef | subjectIndicatorRef )>

<!ELEMENT subjectIdentity ( resourceRef?, ( topicRef | subjectIndicatorRef )* )>

<!ATTLIST subjectIdentity id ID #IMPLIED
baseName ( scope?, baseNameString, variant* )>

<!ATTLIST baseName id ID #IMPLIED
baseNameString ( #PCDATA )>

<!ATTLIST variant ( parameters, variantName?, variant* )>

<!ATTLIST variantName resourceRef | resourceData )>
```

\(^7\) The full XTM specification is available on [http://www.topicmaps.org/xtm/](http://www.topicmaps.org/xtm/)
3.3 Creation and management of XML Topic Maps

Topic maps can be static as well as dynamic. The knowledge structure and the content is fixed with statical topic maps. Dynamical topic maps are build up continually, and their knowledge structure and content is fed with knowledge from the documents it maps out.

Basically, topic maps are nothing more than XML-files. Because of this, topic maps are application- and appliance independent, expandable and adjustable. XML Topic Maps can be composed manually in a texteditor. On the other hand, XML Topic Maps can be generated automatically and dynamically on the basis of XML-documents and databases. This requires more sophisticated tools.

When composing a Topic Map, it’s best to start from a classification of topic maps and the nature of their mutual connection. The first essential step is distilling the concepts and the semantics of information in
documents. Subsequently the topics are classified and further described. The final step is the recording of the connections between the topics.

However, it is not the intent for users to search for their information in the XML-version of a topic maps. It’s possible to put the information in topic maps at the user’s disposal in several different ways. It’s possible to generate an index, a wordlist, thesaurus, table of content etc. dynamically on the basis of the topic map. The end result will be delivered to the user as an HTML-page. The user can be offered navigable overviews of categories, topics, scopes and relations. At the end of the search, the user receives a detailed reproduction of all items with regards to one topic (image 3). Topic Maps can also be depicted graphically. The user receives a schematic depiction of the topics and their relations, joined to the links which lead to (related) documents or sources of information on the topics themselves (image 2).

Image 2: Graphical depiction of a topic map in a web browser. The topics function as junction through which the user can navigate to the searched and related information.

(© mondeo-publishing.com).

Image 3: HTML-output with the depiction of a topic on the basis of a topic map. From the topic (‘Puccini’) are depicted: the different designations, internal and external information, the characteristics and the relations.

(© Ontopia-Omnigator)
4. **Topic maps for archival applications**

The overall applications of topic maps are legion. Up till now, topic maps are most frequently used to find information in series of documents, but topic maps can be used just as well to deliver the structure and content of websites. Topic maps are an ideal instrument to integrate information that is scattered over different sources. Topic maps offer a solution to organise and navigate through large quantities of information which is or isn’t available on line. Topic maps are also implemented in searchengines.

4.1 **Application possibilities for electronic record-keeping.**

Topic maps offer several interesting perspectives for record-keeping purposes. The importance of topic maps should be situated mainly in the field of logical record-keeping management. It’s obvious that topic maps are used to describe and make records accessible on meta level. In an archival context it’s best to assume that the documents which were archived, are static and that there is a need for a dynamic and coordinated finding aid mechanism. Topic maps are used here as an independent knowledge layer on top of the information layer and they function as a navigation tool to organize, to retrieve and give access. With topic maps, information can be controlled in several formats, on several operating systems and on several locations.

Topic maps deliver the knowledge that is necessary to interpret archived documents. In combination with the metadata of records, topic maps deliver the necessary semantic and contextual data of related documents which are kept mostly in an heterogene environment. Fonds or records can be linked to a creator or to certain business processes through a topic map. Whether the records are paper and/or electronic, e-mails or databases, is of no importance. Related records can be connected to each other.

Topic maps offer the possibility to provide access to the archives on the basis of knowledge in the form of connections between records (elements) mutual and their respective context. This way it is possible for the user to search for information in a conceptual framework which consists of context and relations. This offers more possibilities than making it accessible on the basis of information or on the data from the records themselves. For electronic records this means, among others, that one isn’t tied to the providing of access on the basis of hierarchic folders in which electronic information is ordered in a classical way and/or in combination with a fulltext search. A topic map isn’t just hierarchical build, but it offers the possibility of jumping through different hierarchic levels using semantic links. This way a topic map offers a synthesis and an abstraction from the managed information in a more meaningful way. As a consequence, a topic map is also known as a virtual knowledge layer. A topic map offers more navigation possibilities so that a user has more chance of finding the information he searches for. After all, topic maps are linked to each other through associations, common characteristics and/or sources of information. Besides links, the user also receives information on the documents, so he can assess which links are of interest to him.

Topic maps can be created and managed in several different ways. One interesting possibility is compiling a topic map and keeping it up to date for all related documents and records, and to integrate (through merging) the different topic maps on archival level. On the top of the pyramid stands a topic map which charts all the documents that are preserved in an archive. Such a topic map can, for instance, be the basis of the internetwebsite that gives an overview of the records of an archive. This topic map can be consulted as an HTML-page and is at the disposal of the users who are searching for specific sources. Another possibility is linking the documents to records schedules and classification schemes. This application is

---

8 The structure of the website is the same as that of the topic map. The content of the website is partly derived from the topic map, and partly from the occurrences of the topics.
intended more for the administrative services of the organisation and is best integrated in the intranet/website of the archival service.

Concrete application examples of topic maps within the archival context are (not) known or online available. NARA (National Archives and Records Administration) is planning to use the topic maps technology in their archival strategy for electronic records to link knowledge (concepts and relations) and information (data attributes in XML-documents) to each other\(^9\). The lack of functional application examples needn’t surprise. Topic maps are a reasonable young standard. Furthermore, topic maps are very flexible, and this asks for flexible tools. Therefore, the development of the necessary software is not an easy assignment, so that only a limited amount of tools is available at the moment. The topic maps possibilities are anxiously awaited in the XML-, metadata and archival community.

It’s conceivable that the young age of this technology could dissuade archivist from implementing topic maps as an archival purpose. But, it can be argued that the standarisation guarantees some stability and that the essence is the knowledge which is recorded in XML-documents. For the moment it offers the best guarantees available in the world of electronic durability.

### 4.2 Exemplary application

An exemplary application in XTM is given beneath. In the following topic map knowledge about the “population register” is described.

XML Topic Map example

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE topicMap SYSTEM "xtm_topic_maps.dtd">
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/
xmlns:xlink="http://www.w3.org/1999/xlink">
  <topic id="record ">
    <baseName>record</baseNameString>
  </topic>
  <topic id="electronic_record">
    <instanceOf><topicRef xlink:href="#record 
    <baseName>electronic record</baseNameString>
  </topic>
  <topic id="database">
    <instanceOf><topicRef xlink:href="#electronic record 
    <baseName>database</baseNameString>
  </topic>
  <topic id="population registry">
    <instanceOf><topicRef xlink:href="#database 
    <baseName>Data from the population registry concerning persons deregistered before 1983</baseNameString>
  </topic>
  <occurrence>
    <resourceData>Archival number: CDROM SA 908</resourceData>
  </occurrence>
  <occurrence>
    <resourceRef xlink:href="http://www.antwerpen.be/david/cases.htm"
  </occurrence>
</topic>
  </topic>
  <topic id="electoral register2000">
    <instanceOf><topicRef xlink:href="#database 
    <baseName>electoral register of the community-, district- and province counsel, 2000</baseNameString>
  </topic>
  <occurrence>
    <resourceData>Archival number: CDROM SA 907</resourceData>
  </occurrence>
  <occurrence>
</topic>
</topicMap>
```

The user can find a lot of information on the record “population register” on the basis of such a topic map. The population register:

- is a record, more specifically an electronic record with as description: “data from the population register concerning persons deregistered before 1983.”
- is archive of the Department Civel Affairs-Service population, milition and elections
- its metadata is recorded in the file metadata_beam.xml
- is created during the realization of the population registration
- used to be a database
- is archived as an XML-file
- has the archivalnumber CDROM SA 907
- is related to the electoral register
- is documented on http://www.antwerpen.be/david/cases.htm

Important information on the record “population register” and its context is described in such a topic map.
Image 4: A possible application of topic maps. The archives overview on the website of the city archive of Antwerp, can be based on a topic map. By consulting the topic map, the user finds the description, the context and the relations of the population register.

5. Conclusion

Topic maps offer two interesting possibilities for archival services. On the one hand the archivist can record knowledge on records in an electronic, exchangeable and platform independent way for the future in a structural way in a topic map. This way information from several sources can be related to each other and knowledge structures and underlying connections can be build and exchanged. On the other hand electronic sources can be put at everyone's disposal through topic maps. In comparison to a number of classic finding aid instruments, topic maps offer an additional value, so that the searched after information can be found efficiently by the user. Topic maps can be an efficient instrument for archives, especially when they are used for the management as well as for making records accessible.

6. Sources

http://www.ornl.gov/sgml/sc34/doocument/0058.htm
http://easytopicmaps.com
http://www.topicmaps.org
http://www.isotopicmaps.org/
R. MOORE, Knowledge-Based Persistent Archives, 2001.