

Research Article

Using XML to encode TMA DES metadata

Oliver Lyttleton, Alexander Wright¹, Darren Treanor¹, Paul Lewis

Institute of Life Science, School of Medicine, Swansea University, SA2 8PP, ¹Department of Pathology and Tumour Biology, Leeds Institute of Molecular Medicine, University of Leeds, Leeds, U.K.

E-mail: *Oliver Lyttleton - o.lyttleton@swansea.ac.uk

*Corresponding author

Received: 17 December 10

Accepted: 27 June 11

Published: 24 August 11

This article may be cited as:

Lyttleton O, Wright A, Treanor D, Lewis P. Using XML to encode TMA DES metadata. *J Pathol Inform* 2011;2:40.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2011/2/1/40/84233>

Copyright: © 2011 Lyttleton O. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: The Tissue Microarray Data Exchange Specification (TMA DES) is an XML specification for encoding TMA experiment data. While TMA DES data is encoded in XML, the files that describe its syntax, structure, and semantics are not. The DTD format is used to describe the syntax and structure of TMA DES, and the ISO 11179 format is used to define the semantics of TMA DES. However, XML Schema can be used in place of DTDs, and another XML encoded format, RDF, can be used in place of ISO 11179. Encoding all TMA DES data and metadata in XML would simplify the development and usage of programs which validate and parse TMA DES data. XML Schema has advantages over DTDs such as support for data types, and a more powerful means of specifying constraints on data values. An advantage of RDF encoded in XML over ISO 11179 is that XML defines rules for encoding data, whereas ISO 11179 does not. **Materials and Methods:** We created an XML Schema version of the TMA DES DTD. We wrote a program that converted ISO 11179 definitions to RDF encoded in XML, and used it to convert the TMA DES ISO 11179 definitions to RDF. **Results:** We validated a sample TMA DES XML file that was supplied with the publication that originally specified TMA DES using our XML Schema. We successfully validated the RDF produced by our ISO 11179 converter with the W3C RDF validation service. **Conclusions:** All TMA DES data could be encoded using XML, which simplifies its processing. XML Schema allows datatypes and valid value ranges to be specified for CDEs, which enables a wider range of error checking to be performed using XML Schemas than could be performed using DTDs.

Key words: CDEs, DTD, statistical analysis, tissue microarray, TMA DES, XML

Access this article online

Website:

www.jpathinformatics.org

DOI: 10.4103/2153-3539.84233

Quick Response Code:



BACKGROUND

Tissue microarray (TMA) is a high throughput technology that allows hundreds of tissue samples to be placed in a recipient paraffin histology block.^[1] Tissue microarray blocks (TMAs) enable rapid molecular analysis of tissue to provide new diagnostic and prognostic biomarkers as

well as potential therapeutic targets in disease. TMAs are composed of a number of cores arranged in a grid. Each tissue core may be from a different *source block*, and the recipient TMA block can be sectioned to provide thin sections that are subsequently mounted onto glass slides for testing and analysis. As a result, whole tissue cross-sections that expend the tissue block are not required to

be used by a single research study.

The importance of using the XML standard^[2] to encode pathology data has been described previously.^[3] The Tissue Microarray Data Exchange Specification (TMA DES) is an XML specification for encoding TMA experiment data in a machine readable format that is also human readable.^[4] The TMA DES data format has been used by a variety of systems, and allows any persons or system using it to share their TMA data. It was created by the Technical Standards Committee of the Association for Pathology Informatics. TMA DES defines Common Data Elements (CDEs) that form a basic vocabulary for describing TMA data, and have commonly agreed upon syntax and semantics. This enables TMA DES data to be exported from computer applications such as Xperanto-TMA^[5] and TAMEE^[6] and shared among researchers. An extract from a modified version of a TMA DES file included with the paper that described the TMA DES DTD^[7] is shown in Figure 1. The XML in this extract defines CDE names in pairs of tags (e.g. “<Creator>...</Creator>”), and values for those CDEs are contained within those tags (e.g. “Joe Bloggs”). The CDEs in Figure 1 represent the header of a TMA DES file, which contains information such as a title assigned to the TMA and the creator of the TMA. The “header” CDE can contain other CDEs, with information such as a description of the TMA, or the source of the TMA, but only the “Title” and “Creator” CDEs are shown in Figure 1.

While TMA DES data is encoded in XML, the rules defining the syntax and structure of TMA DES data are described in a “Document Type Definition” (DTD) format.^[7] Figure 2 shows an example DTD definition for the “tma” CDE.

```

<tma>
<header>
<Title>TMA Blocks Available for Research from OSU ACSR Mid-
Region</Title>
<Creator>Joe Bloggs</Creator>
<header>
.....
.....
.....
</tma>
    
```

Figure 1: Extract from a modified TMA DES XML file^[7]

```

Identifier: tma
Version: 1.0
Registration Authority: Association for Pathology Informatics
Language: English (en)
Obligation: Required
Datatype: Empty
Maximum Occurrence: Unlimited
Definition: Ancestor element for block and header elements.
Comment: This tag is necessary, because the file could, in theory,
contain more than one tma data set, and you'd need to keep them
apart in that case. You don't need an identifier for each tma,
though, because a tma is identified by its block.
    
```

Figure 3: ISO 11179 description of the TMA DES “tma” CDE

Another standard called ISO 11179^[8] is used to describe the semantics of TMA DES CDEs.^[4] An example of an ISO 11179 description of the “tma” CDE is shown in Figure 3. This entry specifies the Registration Authority that has defined the CDE, the language defining the CDE, whether the CDE is required in a TMA DES document, the datatype of the CDE (e.g. is it textual, numeric, empty), the number of times it can occur in a TMA DES document, and a definition field that describes the semantic meaning of the CDE.

The multiple formats used for encoding data by TMA DES require separate applications for verifying syntactical and structural correctness, and for processing the data they contain, as illustrated in Figure 4. For example, an XML parser (a program used to verify that an XML document is compliant with the rules of the XML syntax, and to extract information from XML documents) could be used to validate and process TMA DES XML data, but it would not recognize a DTD file as a valid document, nor would it be able to extract information from a DTD file. The use of multiple encoding formats, instead of a single encoding format, adds to the burden of developing applications that read and write TMA DES data. In addition, there exist alternative XML-based formats for describing data which in some respects are superior to those used by TMA DES, and there are convincing arguments for adopting their use in place of those currently used by TMA DES. A notable effort to define a schema for TMA data is defined in.^[16] This paper defines an OWL schema containing concepts specific to TMA experiments, and illustrates the benefits of using OWL to describe TMA data. A limitation of XML is that there must be some agreement on the semantic meanings of CDEs before they are useful for

```

<!ELEMENT tma (header+, block+)>
    
```

Figure 2: DTD definition for the tma element

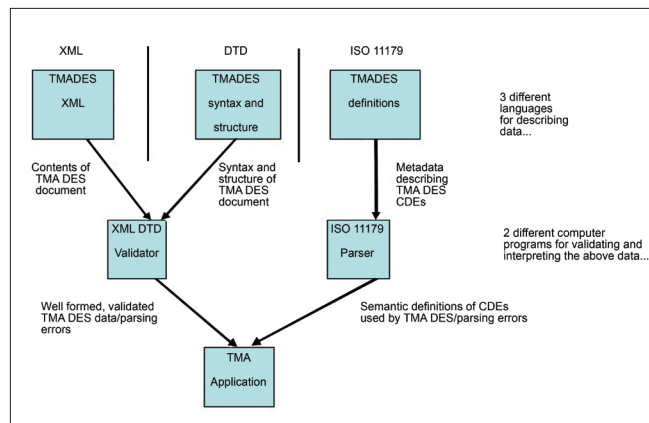


Figure 4: TMA DES data requires separate parsers for XML, DTD, and ISO 11179 documents

sharing data. OWL, on the other hand, is a format that can use CDEs defined in existing ontologies. The TMA OWL schema^[9] illustrates how existing OWL ontologies are used by the authors when producing TMA data. OWL also enables complicated relationships between CDEs to be defined. However, this OWL schema does not contain all the concepts that TMA DES does. It also does not address the issues with TMA DES that arise due to the lack of an encoding format for the existing TMA DES ISO11179 definitions. In addition, while OWL allows data to be described, it does not provide mechanisms for defining the syntax/structure of data, or valid ranges of values for data.

We propose using XML to encode all TMA DES data. Two XML based standards, XML Schema^[10] and the Resource Description Framework (RDF)^[11] can be used in place of DTDs and ISO 11179 respectively.

Advantages of XML Schema over DTDs

XML Schema has a number of advantages over DTDs. Firstly, unlike DTDs it allows datatypes for CDEs to be specified. For example, the TMA DES CDEs in Figure 5 represent the date a block was created, a textual description of the block, and the number of cores in the block.

These data are of different types (date, textual and numeric respectively). Unlike DTDs, XML Schemas enable data to be checked for invalid values by allowing datatypes to be specified for CDEs. For example, a program may require that the value of the `block_number-of-cores` CDE in Figure 5 be numeric (e.g. “26” as above), not textual (e.g. “twenty six”).

XML Schemas are also more flexible than DTDs when specifying valid ranges of values for CDEs. For example, the following constraints on values could be imposed on CDEs using XML Schema:

- `Block_core-size` must be a real number with one decimal place
- `Block_number-of-cores` must be an integer between 0 and 100 inclusive
- `Core_GENDER` must be a single character with the value “M” or “F”
- `Block_identifier` must be composed of two characters, followed by 2 single digit integers, then a hyphen, then 3 single digit integers (e.g. “TA00-050”, “BD10-450”).

The TMA DES specification defines valid ranges of values for CDEs in the definition attribute of entries in an ISO 11179 file. For example, the definition attribute for the `core_organism` CDE is:

Organism name at species level for organism whose tissue is represented in the donor block, Comment: URI for taxonomy.dat is <ftp://ftp.ebi.ac.uk/pub/databases/taxonomy/taxonomy.dat> The correct entry for human tissue is “9606 human”

```
<block_creation-date>2000-1-1</block_creation-date>
<block_description>heart tissue from 7 Autopsy
patients</block_description>
<block_number-of-cores>26</block_number-of-cores>
```

Figure 5: CDEs from TMA DES XML data

Computer applications may not understand plain English definitions for CDEs such as the above. TMA DES does not provide a commonly accepted mechanism for enabling parsers to automatically check that the values of CDEs comply with their ISO 11179 definitions. These definitions serve as guidelines for producers of data, rather than a mechanism for automatically checking that data is formatted correctly. In contrast, XML parsers can verify that the value of CDEs satisfy constraints that are expressed in XML Schema documents.

Advantages of RDF Schema over ISO 11179

RDF is commonly encoded in XML, whereas ISO 11179 does not specify how metadata should be encoded. There may be inconsistencies in how parsers interpret ISO 11179 data. For example, a parser may identify CDE names as a string of characters, followed by a colon, and terminated with a carriage return. This parser would identify three CDEs in the sample text below (“Maximum Occurrence”, “Definition,” and “characters”) whereas in actual fact only “Maximum Occurrence” and “Definition” are valid CDEs.

Maximum occurrence: Unlimited

Definition: This is the slide identifier issued by the distributor; it is a character string that permits the following characters: “[0-9][a-z][A-Z].-_:”

Another example of a potential problem occurs in the ISO 11179 definition of the TMA DES “`block_protocol`” CDE.^[12] This definition is shown below:

Datatype: Character string maximum occurrence: Unlimited

This line contains definitions for two attributes of the CDE (“Datatype” and “Maximum Occurrence”). Each attribute for all other CDEs defined in this document are on their own line, therefore we can assume that this is a typographical error and the line should be split in two, with separate lines defining the “Datatype” and “Maximum Occurrence” attributes. While it may appear obvious that this definition contains two attributes of a CDE, it is not obvious to machines (that process this data). For example, a parser we wrote to convert ISO 11179 data to RDF determined that the datatype of the `block_protocol` CDE was “Character String Maximum Occurrence: Unlimited”.

Unlike ISO 11179, RDF is part of the stack of protocols that compose the “Semantic Web.”^[13] Berman

emphasizes the importance of XML providing logic and meaning to data, allowing it to be part of the “Semantic Web.”^[3] The use of RDF to encode TMA DES metadata could potentially enable applications to process TMA DES data with greater autonomy.

We have implemented an XML Schema that describes the syntax and structure of TMA DES. We have also produced an RDF file that provides the same semantic definitions that are specified in the TMA DES ISO 11179 file.

MATERIALS AND METHODS

Creating an XML Schema version of the TMA DES DTD

There were two main tasks in creating an XML Schema for TMA DES:

1. The structure and syntax rules specified in the TMA DES DTD are replicated in an XML Schema.
2. The datatypes and constraints on values for CDEs defined in the TMA DES ISO 11179 file are defined in this XML Schema.

In order to create an XML Schema version of the TMA DES DTD, an online tool^[14] was used to convert the TMA DES DTD into an XML Schema. An example of a CDE, “core_histo-repository,” from the resulting XML Schema is shown in Figure 6.

The XML Schema extract, Figure 6, specifies the same information as the definition for the core_histo-repository CDE in the TMA DES DTD. The XML Schema document produced by the DTD to XML Schema converter (Supplementary File 1) contains definitions for all the TMA DES CDEs.

After replicating the TMA DES syntax and structure rules in an XML Schema document, the TMA DES ISO 11179 file was consulted in order to determine the datatypes that the TMA DES XML Schema document should specify for CDEs. The following is a list of the CDE datatypes used in the ISO 11179 file:

- Character String
- Date (YYYY-MM-DD)
- Integer
- Real Number
- Real Numbers
- Character String representing taxonomy.dat identifier number followed by an allowable taxonomy.dat name for the identifier number
- Decimal number
- Decimal number ranging from 0 to 100
- Empty

Some of these datatypes are mapped to their XML Schema equivalents [Table 1].

Other datatypes contained restrictions on their values that were also defined in our XML Schema (e.g. “Decimal

number ranging from 0 to 100”). We modified the CDE definitions in the XML Schema file so that they had the appropriate datatypes. For example, those which were defined as being “Real” or “Decimal” type CDEs in the ISO 11179 file were defined as being of type “xs:float” in the XML Schema. The TMA DES XML Schema was also modified where appropriate so that it specified constraints on the values of CDEs that were specified in the TMA DES ISO 11179 file. Figure 7 shows an XML Schema extract that specifies a constraint on the value of the “core_organism” CDE. This CDE definition uses a regular expression to specify that values for this CDE must consist of a sequence of digits, followed by any number of alphanumeric characters (“<xs:pattern value=”[0-9]*.*”/>”). XML Schema allows other constraints such as valid ranges of numeric values and particular values for textual data to be specified.

Creating an RDF version of the TMA DES ISO 11179 file

We wrote a program called “ISO11179toRDF” that parsed the TMA DES ISO 11179 file and created an RDF version of this file. This program is a bash script written on a computer running a Unix operating system (Supplementary File 2). Some alterations had to be made to the TMA DES ISO 11179 file in order to ensure that all attributes were described in a consistent fashion. ISO 11179 does

```
<xs:element name="core_histo-repository">
  <xs:complexType mixed="true">
    <xs:choice>
      <xs:element ref="core_histo-repository_specimen-id" />
      <xs:element ref="core_histo-repository_specimen-diagnosis" />
      <xs:element ref="core_histo-repository_donor-block" />
    </xs:choice>
  </xs:complexType>
</xs:element>
```

Figure 6: XML Schema definition of a TMA DES CDE

```
<xs:element name="core_organism" type="core_organism_type"/>
<xs:simpleType name="core_organism_type">
  <xs:restriction base="xs:string">
    <xs:pattern value="[0-9]*.*"/>
  </xs:restriction>
</xs:simpleType>
```

Figure 7: XML Schema definition of the core_organism CDE

Table 1: Mapping ISO 11179 datatypes to equivalent XML schema datatypes

ISO 11179 datatype	XML schema datatype
Character string	Xs:string
Date (YYYY-MM-DD)	Xs:date
Integer	Xs:integer
Real number(s)	Xs:float
Decimal number	Xs:float

not specify syntax for encoding data. If the data contained in this file were encoded in XML, it would not have been necessary to manually alter the file so that the program could parse it, illustrating the advantage of our approach of encoding this data using XML. These alterations included:

- Removing HTML links at the beginning of the document which linked to the individual CDEs.
- Removing carriage returns from property values.
- Ensure that there was only one property definition per line (e.g. in the entry for the “block_protocol” CDE, split the line “Datatype: Character StringMaximum Occurrence: Unlimited” into two lines, “Datatype: Character String” and “Maximum Occurrence: Unlimited”).

The program was run from the command prompt on a machine running Unix as shown below:

```
ISO11179toRDF TMADES_CDEs.txt >> TMADES_CDEs.rdf
```

The program takes the name of a file containing ISO 11179 definitions as input (in this case, “TMADES_CDEs.txt”). The output from the ISO11179toRDF application was an RDF version of the ISO11179 definitions. This output could be redirected to a file, as in the example above, using the “>>” operator followed by the name of a file (“TMADES_CDEs.rdf”). Figure 8 shows an extract from the RDF file generated in this manner which describes the “histo” CDE.

The RDF generated by ISO 11179 to RDF contains some redundant information. The “Obligation”, “Datatype” and “MaximumOccurrence” CDEs contain data that is also specified in the TMA DES XML Schema document. For the sake of completeness, however, we have included them in our RDF file (Supplementary File 3).

RESULTS

Validating TMA DES data with our TMA DES XML schema

We performed an experiment where our TMA DES XML Schema document was used to confirm that existing TMA DES data were well formed and valid. The TA00-050.XML file representing TMA DES data^[7] was successfully tested for conformance to the syntax and rules in this XML Schema document using an online XML schema validator.^[15] We also deliberately introduced errors into the TA00-050 XML file, and attempted to validate it with both the XML Schema validator and an XML DTD validator^[16] in order to compare their error checking capabilities. Table 2 shows the errors that were introduced, along with the messages generated by the two validators. While the first three errors were detected by both the XML Schema and the DTD validator, the last two errors were only detected when validation took place against the XML Schema document.

Parsing RDF data generated from TMA DES ISO

```
<rdf:Description rdf:about=" histo ">
  <cd:Version>1.0</cd:Version>
  <cd:RegistrationAuthority>Association for Pathology
  Informatics</cd:RegistrationAuthority>
  <cd:Language>en</cd:Language>
  <cd:Obligation>Required</cd:Obligation>
  <cd:Datatype>Empty</cd:Datatype>
  <cd:MaximumOccurrence>One</cd:MaximumOccurrence>
  <cd:Definition>Root element of the Tissue Microarray
  document</cd:Definition>
  <cd:Comment></cd:Comment>
</rdf:Description>
```

Figure 8: RDF description of the “histo” CDE

Table 2: Errors introduced into TMA DES XML data, and error messages returned by XML Schema and DTD parsers when the XML was validated against them

Type of error	Error introduced	Output from XML schema validator	Output from XML DTD validator
Mis-spelled CDE names	“block_purpose” was changed to “block_purposeee”	Invalid content was found starting with element “block_purposeee”	Element type “block_purposeee” must be declared
Mis-matched XML tags	“<block_creator>Yi, Hun</block_creator>” changed to “<block_creator>Yi, Hun</histo>”	The element type “block_creator” must be terminated by the matching end-tag “</block_creator>”	The element type “block_creator” must be terminated by the matching end-tag “</block_creator>”
Incorrect use of brackets	“<block_core-size>2.0 </block_core-size>” changed to “<block_core-size>>2.0 <</block_core-size>”	The content of elements must consist of well-formed character data or markup	The content of elements must consist of well-formed character data or markup
Incorrect datatypes	<block_number-of-cores>Twenty six</block_number-of-cores>	“Twenty six” is not a valid value for “integer”	Error not detected
Incorrectly formed data	<core_results_percent-tissue-staining>80</core_results_percent-tissue-staining> changed to <core_results_percent-tissue-staining>800</core_results_percent-tissue-staining>	Value “800” is not facet-valid with respect to maxInclusive “1.0E2” for type “percent”	Error not detected

Table 3: Triples generated by W3C RDF parser from TMA DES RDF file

Subject	Predicate	Object
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/Version	"1.0"
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/RegistrationAuthority	"Association for Pathology Informatics"
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/Language	"en"
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/Obligation	"Required"
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/Datatype	"Empty"
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/MaximumOccurrence	"One"
http://www.w3.org/RDF/Validator/run/ histo	http://www.w3schools.com/rdf/Definition	"Root element of the Tissue Microarray document"

11179 data

When an RDF parser processes an RDF file, it should be able to extract the object or concept which is being described, the attribute names of that object/concept, and the values of these attributes (this combination of data is known as a “triple”). Statements such as the following can thus be created by RDF parsers:

- The cd:Version of the Histo object is “Version 1.0”
- The cd:RegistrationAuthority of the Histo object is “Association for Pathology Informatics”
- The cd:Language of the Histo object is “en”

After executing ISO11179toRDF, an RDF file with definitions for the 80 TMA DES CDEs was created. This TMA DES RDF file was successfully parsed using the W3C RDF Validation Service.^[17] Examples of the triples generated by this service are shown in Table 3, and the complete output from the W3C RDF parser is in Supplementary File 3.

CONCLUSIONS

TMA DES specifies a syntax which allows researchers to share their TMA data. TMA DES facilitates collaborative research and the dissemination of research results involving TMA data. The introduction of the TMA DES has greatly facilitated the recording, storage, and sharing of TMA data. One current problem however is that three formats are used for encoding TMA DES data (XML, DTD and ISO 11179). We have shown that all TMA DES data could be encoded using XML, which simplifies processing. XML Schema allows datatypes and valid values to be specified for CDEs, which enables a wider range of error checking to be performed using XML Schemas than could be performed using DTDs. The use of RDF encoded in XML makes it easier to share this data. XML specifies how data is encoded, which enables computer applications to parse it.

We produced an XML Schema version of the TMA DES DTD, and used this to validate the TA00-050. XML file that was included as a supplementary file in the study on the TMA DES DTD published by Nohle *et al.*^[7] and

demonstrated the superior error checking capabilities of our schema to the existing TMA DES DTD. The development of new tools which validate TMA DES data using XML Schemas is an opportunity for future work.

We created a program which converted data in the TMA DES ISO 11179 file to an RDF XML format. The output from this program was successfully parsed by the W3C RDF Validation Service, with no human assistance required. In contrast, the ISO 11179 file required manual pre-processing in order that it could be parsed correctly by our program, illustrating the benefits of encoding TMA DES metadata in XML. RDF is also part of the stack of protocols that compose the “Semantic Web,” which envisions applications using metadata to make decisions about how data is processed without human guidance. An avenue for future investigation is the development of “software agents” that can aggregate RDF descriptions of TMA DES data from multiple sources, and “reason” and perform operations on that data in an autonomous fashion.

We have encoded the syntax and semantic rules for TMA DES in a single language, XML, and have illustrated how a wider range of error checking can be performed on TMA DES data using XML schemas instead of DTDs, and how the use of RDF encoded in XML for defining semantic rules for TMA DES CDEs allows these rules to be interpreted by applications, allowing semantic data to be easily shared. Pathologists could benefit from these advantages by using TMA DES processing applications which use XML Schemas and RDF.

REFERENCES

1. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, *et al.* Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4:844-7.
2. Bray T, Paoli J, Sperberg-McQueen CM, Maler E, Yergeau F. Extensible Markup Language (XML) 1.0. W3C Recommendation; 2008. Available from: <http://www.w3.org/TR/2008/REC-xml-20081126/> [Last accessed 2008 Nov 08].
3. Berman J. Pathology data integration with eXtensible Markup Language. *Hum Pathol* 2005;36:139-45.
4. Berman JJ, Edgerton ME, Friedman BA. The tissue microarray data exchange

- specification: A community-based, open source tool for sharing tissue microarray data. *BMC Med Inform Decis Mak* 2003;3:5. Available from: <http://www.biomedcentral.com/1472-6947/3/5> [Last accessed on 2010 Dec 03].
5. Xperanto-TMA. Available from: <http://www.xperanto.snubi.org/TMA/>. [Last accessed on 2010 Dec 03].
 6. Thallinger G, Baumgartner K, Pirklbauer M, Uray M, Pauritsch E, Mehes G, et al. TAMEE: data management and analysis for tissue microarrays. *BMC Bioinformatics* 2007;8:81.
 7. Nohle DG, Ayers LW. The tissue microarray data exchange specification: A document type definition to validate and enhance XML data. *BMC Med Inform Decis Mak* 2005;5:12. Available from: <http://www.biomedcentral.com/1472-6947/5/12> [Last accessed on 2010 Dec 03].
 8. Solbrig HR. Metadata and the reintegration of clinical information: ISO 11179. *MD. Comput* 2000;3:25-8.
 9. Kang HP, Borromeo CD, Berman JJ, Becich MJ. The tissue microarray OWL schema: An open-source tool for sharing tissue microarray data. *J Pathol Inform* 2010;1. pii:9.
 10. Thompson H, Beech D, Maloney M, Mendelsohn N. XML Schema Part 1: Structures. W3C Recommendation Available from: <http://www.w3.org/TR/xmlschema-1/> [Last accessed on 2004 Nov 28].
 11. Lassila O, Swick R. Resource Description Framework (RDF) Model and Syntax Specification. Available from: <http://www.w3.org/TR/REC-rdf-syntax/> [Last accessed on 2010 Dec 03].
 12. Supplementary file for "The tissue microarray data exchange specification: A community-based, open source tool for sharing tissue microarray data". Available from: <http://www.biomedcentral.com/content/supplementary/1472-6947-3-5-s1.htm>. [Last accessed on 2010 Dec 03].
 13. Berners-Lee T, Hendler J, Lassila O. The semantic web. *Sci Am* 2001 284:34-43.
 14. DTD, XML Schema and XML document conversion software tool. Available from: http://www.hitsw.com/xml_utilities/ [Last accessed on 2010 Dec 03].
 15. XML Schema validator, Available from: <http://tools.decisionsoft.com/schemaValidate/> [Last accessed on 2010 Dec 03].
 16. XML DTD validator, Available from: <http://www.xmlvalidation.com/>. [Last accessed on 2010 Dec 03].
 17. W3C Validation service, Available from: <http://www.w3.org/RDF/Validator/> [Last accessed on 2010 Dec 03].