

Research article

The tissue microarray data exchange specification: A community-based, open source tool for sharing tissue microarray data

Jules J Berman*¹, Mary E Edgerton² and Bruce A Friedman³

Address: ¹Cancer Diagnosis Program, National Cancer Institute, Bethesda, MD, USA, ²Department of Pathology, Vanderbilt University Medical Center, Nashville, TN, USA and ³Department of Pathology, University of Michigan, Ann Arbor, MI, USA

Email: Jules J Berman* - bermanj@mail.nih.gov; Mary E Edgerton - Mary.Edgerton@vanderbilt.edu; Bruce A Friedman - bfriedma@med.umich.edu

* Corresponding author

Published: 23 May 2003

Received: 26 February 2003

BMC Medical Informatics and Decision Making 2003, **3**:5

Accepted: 23 May 2003

This article is available from: <http://www.biomedcentral.com/1472-6947/3/5>

© 2003 Berman et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Tissue Microarrays (TMAs) allow researchers to examine hundreds of small tissue samples on a single glass slide. The information held in a single TMA slide may easily involve Gigabytes of data. To benefit from TMA technology, the scientific community needs an open source TMA data exchange specification that will convey all of the data in a TMA experiment in a format that is understandable to both humans and computers. A data exchange specification for TMAs allows researchers to submit their data to journals and to public data repositories and to share or merge data from different laboratories. In May 2001, the Association of Pathology Informatics (API) hosted the first in a series of four workshops, co-sponsored by the National Cancer Institute, to develop an open, community-supported TMA data exchange specification.

Methods: A draft tissue microarray data exchange specification was developed through workshop meetings. The first workshop confirmed community support for the effort and urged the creation of an open XML-based specification. This was to evolve in steps with approval for each step coming from the stakeholders in the user community during open workshops. By the fourth workshop, held October, 2002, a set of Common Data Elements (CDEs) was established as well as a basic strategy for organizing TMA data in self-describing XML documents.

Results: The TMA data exchange specification is a well-formed XML document with four required sections: 1) Header, containing the specification Dublin Core identifiers, 2) Block, describing the paraffin-embedded array of tissues, 3) Slide, describing the glass slides produced from the Block, and 4) Core, containing all data related to the individual tissue samples contained in the array. Eighty CDEs, conforming to the ISO-11179 specification for data elements constitute XML tags used in the TMA data exchange specification. A set of six simple semantic rules describe the complete data exchange specification. Anyone using the data exchange specification can validate their TMA files using a software implementation written in Perl and distributed as a supplemental file with this publication.

Conclusion: The TMA data exchange specification is now available in a draft form with community-approved Common Data Elements and a community-approved general file format and data structure. The specification can be freely used by the scientific community. Efforts sponsored by the Association for Pathology Informatics to refine the draft TMA data exchange specification are expected to continue for at least two more years. The interested public is invited to participate in these open efforts. Information on future workshops will be posted at <http://www.pathologyinformatics.org> (API web site).

Background

Tissue Microarrays (TMAs), first introduced in 1998, are collections of hundreds of tissue cores arrayed into a single paraffin histology block [1]. Each TMA block can be sectioned and mounted onto glass slides, producing hundreds of nearly-identical slides. TMAs permit investigators to use a single slide to conduct controlled studies on large cohorts of tissues, using a small amount of reagent. The source of tissue is only restricted by its availability in paraffin and ranges from cores of embedded cultured cells to tissues from any higher organism. In a typical TMA study, every TMA core is associated with a rich variety of data elements (image, tissue diagnosis, patient demographics or other biomaterial description, quantified experimental results). Under ideal circumstances, a single paraffin TMA block can be sectioned into nearly identical glass slides dispensed to many different laboratories. These laboratories may use different experimental protocols. They may capture data using different instruments, different databases, different data architectures, different data elements and immensely different formats. These laboratories could vastly increase the value of their experimental findings if they could merge their findings with those of the other laboratories that used the same TMA block. Unfortunately, the practice of merging TMA data sets obtained at different laboratories using different information systems is infrequently practiced. A key barrier to this process is the incompatibility of the individual data sets. There simply is no specification for exchanging TMA data. Without such a specification, TMA data files can not be shared or merged, and the full scientific value of this new technology is not realized.

The need for a data exchange standards is not unique to TMA experiments. The same issues have challenged scientists who use another array technology, gene expression arrays. The history of MIAME (minimum information about a microarray experiment) and MGED (microarray gene expression databases) Standards has been described [2-4]. XML was chosen as the formatting language for the effort to standardize gene expression data. The purpose of XML is to eliminate barriers to data exchange and permit the integration of data from heterogeneous sources [5,6]. Consequently, a TMA specification in XML would provide another level of biomedical data integration in which array data of many different types can be combined and analyzed.

In the past five years, XML has emerged as the document format used in almost all new data standards. XML achieves its functionality through the use of metadata. Metadata is data that describes things, including data elements. In XML documents, most metadata is marked by enclosing sets of angle brackets:

```
<birthdate>Jan 1, 2003</birthdate>
```

In this example, `birthdate` is the metadata that describes and flanks "Jan 1, 2003". Besides providing metadata to describe the data elements contained in the XML file, XML files can be self-described by metadata. Self-description is a simple but powerful concept. If a file describes itself (its subject, its creator, its semantics and all its data elements), it can exist as a completely independent data object unassociated with any software applications or database instrument. There is an abundance of freely available software implementations, written in a variety of open source programming languages, that permit users to validate and parse XML data files [5].

The metadata in XML files can be formalized as well-defined common data elements whose definition and usage semantics are explicitly described in an accessible, unique document. Most efforts that create an XML specification for data domains (such as gene expression data), will also produce a formal document for the metadata (XML tags) included in the specification. Anyone wishing to implement the specification would need to understand the metadata definitions and refer to the metadata definition document from their XML data files. The standard guidelines for creating metadata specifications is the ISO-11179 [7].

Methods

Organization and sponsoring agencies

The Association for Pathology Informatics (API) is an organization whose mission is to promote the field of pathology informatics as an academic and a clinical subspecialty of pathology. Further information on the API is available at their web site:

<http://www.pathologyinformatics.org>

The Technical Standards Committee of the API, recognizing the importance of TMA technology to pathology departments, organized a TMA workshop to discuss the subject of a TMA data exchange specification. A sponsor-

Table 1: Organization and Support

-
1. The TMA data exchange specification effort is sponsored by the Association for Pathology Informatics.
 2. Workshops were chaired by a designated member of the Technical Specifications Committee of the Association for Pathology Informatics
 3. The costs of the workshops were partially offset by a meeting support grant from the National Cancer Institute (R13 CA094453).
 4. The APIII and AIMCL organizers provided meeting rooms and scheduled meeting time to the TMA workshop and advertised the TMA workshops in their brochures and meeting announcements.
-

Table 2: TMA Data Exchange Specification Workshops

May 30, 2001. Held in conjunction with Automated Information Management in the Clinical Laboratory 20th Annual Symposium, Ann Arbor, Michigan. Chair of speaker session: Mark A Rubin. Speakers: David Rimm, Steve Bova, Matt Van de Rijn, Jules Berman. Chair of Data Exchange Workshop: Jules Berman. Workshop summary available at: <http://65.222.228.150/jjb/tmasum.htm>

Oct. 6, 2001. Held in conjunction with Advancing Pathology Informatics, Imaging and the Internet, Pittsburgh, PA and co-sponsored by The National Cancer Institute. Chair, Mary Edgerton. Speakers: Olli Kallioniemi, Chris Chute, Richard Lieberman, Paul Spellman. Chair of Data Exchange Workshop: Mary Edgerton. Workshop summary available at: http://65.222.228.150/jjb/tma10_6.htm

May 22, 2002. Held in conjunction with Automated Information Management in the Clinical Laboratory 20th Annual Symposium, Ann Arbor, Michigan and co-sponsored by the National Cancer Institute. Chair of Speaker session: Mark A. Rubin. Speakers: James Bacus, Angelo de Marzo, Peggy Porter, David Rimm and Guido Sauter. Chair of Data Exchange Workshop: Dr. Mary Edgerton.

October 4, 2002. Held in conjunction with Advancing Pathology Informatics, Imaging and the Internet, Pittsburgh, PA. Chair of speaker session: Mary Edgerton. Speakers: Steve Hewitt, Ulysses Balis. Chair of Data Exchange Workshop: Mary Edgerton.

ship listing is contained in Table 1. The complete listing of TMA Data exchange workshops is found in Table 2.

The first community discussion was held May 30, 2001 at the "Automated Information Management in the Clinical Laboratory" meeting, held at the University of Michigan and organized by Dr. Bruce Friedman. The complete listing of TMA Data exchange workshops is found in Table 2.

Workshop attendees endorsed an effort by the Technical Standards Committee of the API, to create a TMA data exchange specification. Consensus was reached on the following guiding principles, and has been documented in a [public domain] workshop summary, available at:

<http://65.222.228.150/jjb/tmasum.htm>

1. The standard should be free and non-proprietary.
2. The standard should be self-descriptive. Anyone reviewing a TMA file should be able to precisely determine how the data is organized by reading the data tags included in the file.
3. The standard should, when feasible, use publicly available common data elements linked to a web site that fully defines each common data element included in the standard (needed to support dataset-independent distributed network queries). This means that the committee that creates the TMA standard must work with other standards committees to ensure cross-database compatibility of common data elements

4. The standard should be generic (able to describe any laboratory's TMA data structure)
5. The standard should be extensible. This means that there will need to be a standards committee that can make changes in the standard over time and that can keep a documented history of modifications in the standard.
6. The standard should be easy to implement. It should be relatively easy for a programmer to translate any commercial TMA dataset into the TMA standard (and to reverse the process)
7. The standard should not be a requirement. The committee that creates the standard should take no measure to require laboratories to implement the standard. Those using the standard would be able to choose that data that is included in their shared datasets (e.g. they may choose to withhold or encrypt patient identifiers)
8. The standard should have community buy-in. Laboratories, commercial vendors, pathology organizations, government agencies, and other standards committees should all have the opportunity to comment on the standards.

Subsequent workshops affirmed the guidelines established in the first conference, but plans to develop a formal standard (approved by a Standards Organization) were abandoned in favor of developing a community-supported TMA specification. The specification would conform to pre-existing standards for creating XML vocabularies and well-formed XML documents. This strategy would produce a new specification as a standard metadata document [see Discussion]. The current draft specification complies with guidelines formulated during the first workshop, and the CDEs and TMA data structure conform with all subsequent recommendations from workshop participants. The workshops themselves were composed of representatives from academia, industry and government. About 75 people were present at each workshop.

Results

The data structure

Every TMA file is an XML file that is divided into 4 sections

1. A header section, with data elements that provides basic information about the file (creator, date created, etc.). The header elements are taken directly from the Dublin Core, a set of specification elements used in libraries throughout the world to index electronic information files <http://dublincore.org>.

Table 3: List of Common Data Elements (XML Tags)

histo
header
Title
Creator
Subject and Keywords
Description
Publisher
Contributor
Date
Resource Type
Format
Resource Identifier
Source
Language
Relation
Coverage
Rights Management
filename
tma
block
block_purpose
block_identifier
block_description
block_meshnumbers
block_related-blocks
block_creator
block_creation-date
block_protocol
block_number-of-cores
block_core-size
block_core-spacing
block_qa
block_status
block_array-hash
slide
slide_identifier
slide_cut-date
slide_storage
slide_section-thickness
slide_level
slide_sectioning-protocol
slide_test-date
slide_test-category
slide_test-protocol
slide_stain
slide_test-control
core
core_array-id
core_tissue-fixative
core_organism
core_demographics-record
core_anatomic-site
core_organ
core_tissue-origin
core_data-descriptor
core_data-descriptor_QC
core_clinical-info
core_clinical-info_id
core_clinical-info_availability
core_histo-repository
core_histo-repository_specimen-id
core_histo-repository_specimen-diagnosis

Table 3: List of Common Data Elements (XML Tags) (Continued)

core_histo-repository_donor-block
core_histo-repository_donor-block_drill-site
core_histo-repository_donor-block_drill-site_diagnosis
core_results
core_results_slide-id
core_results_method-of-reading-data
core_results_number-of-nuclei-counted
core_results_tissue-intensity
core_results_percent-tissue-staining
core_results_compartment-staining
core_results_staining-pattern
core_image
core_image_slide-id
core_image_identifier
core_image_format
core_image_checksum
core_image_magnification
core_image_pixel-size

2. A block section, with data elements that describe the TMA block (how many cores, how large are cores, how are the cores arrayed in the block, etc.)

3. A slide section, with data elements that describe the slides prepared from the TMA block (how are the slides stored, how are they identified, etc)

4. A core section with data elements that describe each of the cores in the TMA block (what case did the core come from, what block from the case was used to make the core, what drill-site in the block was used, what was the diagnosis of the drill-site, what clinical history is associated with the core, what demographic information is associated with the patient from whom the core was taken, etc.). This section is by far the longest section, with well-annotated data for every core in the TMA array.

The data elements

Common Data Elements (CDEs) are well-defined XML metadata tags that can be used to consistently describe data in different XML files. Eighty CDEs were created to describe the kinds of data contained in a TMA file. These data elements are listed in Table 3. Each data element is fully described as a set of features conforming to the ISO-11179 standard for meta-data [7]. The CDE descriptor file is included as an attachment to this article [see Additional file 1]. The most current descriptor file for the TMA CDEs can be found as a public document at:

http://65.222.228.150/jjb/tma_cde.htm

This web address is the Unique Resource Identifier (URI) for the TMA CDEs. XML files that use any of the TMA CDEs can refer to the URI, ensuring that anyone encoun-

tering the CDE will understand its intended meaning [5,6].

Two example CDEs from the descriptor file are shown. Each CDE is followed by a basic set of information as specified in ISO-11179 that fully describes the CDE. TMA files may refer directly to the URL as a namespace reference for the TMA CDEs.

slide_test_date

Identifier: slide_test_date

Version: 1.0

Registration Authority: Association for Pathology Informatics

Language: English (en)

Obligation: Optional

Datatype: Date (YYYY-MM-DD)

Maximum Occurrence: Unlimited

Definition: This is the date that the slide was stained; it should be expressed in integers as YYYY-MM-DD.

slide_test_category

Identifier: slide_test_category

Version: 1.0

Registration Authority: Association for Pathology Informatics

Language: English (en)

Obligation: Optional

Datatype: Character String

Maximum Occurrence: Unlimited

Definition: This is the type of special study applied to this slide, e.g. FISH,

Immunohistochemistry, in situ PCR, BLOT, regular stain, other.

Semantic rules for the TMA data exchange specification

Six semantic rules define the TMA Data Exchange Specification. The specification refers to CDEs listed in Table 3.

1. The TMA file must consist of well-formed XML.
2. Every TMA file must have histo as its root element.
3. Every TMA file must have header, tma, block, slide, and core element sections.
4. The tma element is nested under the root element, histo. The header and block elements are nested under the tma element. The slide and core elements are nested under the block element but are not nested within each other.
5. The header elements are the Dublin Core elements and provide general information about the file and the laboratory that produced the file. The header section must be the first TMA CDE nested under the 'tma' element.
6. Elements that begin with block_, slide_ or core_ are nested by the hierarchy contained in their name and separated by underscores.

This approach gives the TMA creators enormous flexibility while still providing a rich set of metadata and a uniform data structure.

1. The semantics of every TMA data exchange file can be entirely specified by six semantic rules that can be understood by non-programmers.
2. Users of the specification can add tags of their own creation and can even add arguments to the list of CDEs. Users are free to add any XML constructs they wish (DTDs, Schemas, Entities, non-parsed data, RDF references, attributes, etc.). Namespace prefixes are allowed.
3. The only TMA CDEs (XML tags) required in every TMA document are histo, header, tma, block, slide, and core. There are no data inclusion requirements, so a valid TMA file may consist exclusively of XML tags.
4. The six semantic rules are easy to model in validating software implementations. A Perl script (tmavalid.pl) for validating TMA data exchange files is included as a supplemental attachment with this publication.
5. The six semantic rules provide enough data structure and metadata to for TMA users to design understandable and parsable TMA documents.
6. The six semantic rules and the CDEs can be referred to from URIs within the TMA documents, so that TMA documents can be self-descriptive.

Examples of valid TMA data exchange files

Example 1 is the simplest and shortest possible TMA data exchange file.

Example 1.xml

```
<histo>

  <tma>

    <header>

  </header>

  <block>

    <slide>

  </slide>

  <core>

  </core>

  </block>

  </tma>

</histo>
```

The specification does not require that XML tags actually enclose data. Example 1, consisting only of tags, is a well-formed TMA file. Only the minimal required CDEs, are contained: histo, header, tma, block, slide and core. Note that 'histo' is the root element. The tma file nests under histo. A single TMA file may include several tma tags, allowing a collection of many different TMA data sets in a single document. The 'header' section nests under 'tma'. Typically, the 'header' section will contain the Dublin Core elements. The header section, when populated by all the Dublin Core header elements, will permit indexing services, libraries, publishers, and anyone examining the TMA document, to easily determine the basic identifying information about the file (who made the file, what is the file, when was the file made, where was it made, etc).

Example 2.xml

```
<histo>

  <my_own_tag>

</my_own_tag>

  <tma>

  <header>
```

```
<font>

</font>

</header>

<block>

  <slide>

</slide>

  <core>

  <secret-tag>

</secret-tag>

  </core>

  <core name = "liver">

</core>

  <slide>

</slide>

</block>

</tma>

</histo>
```

Example 2 is also valid. Notice that several of the tags do not match CDEs found in Table 3. This is permitted. Also, an attribute was added for 'core' a required TMA element. The attribute is not part of the TMA specification. The TMA validator ignores elements and attributes added by the user.

Example 3.xml (invalid structure)

```
<HISTO>

  <tma>

    <block>

      <header>

    </header>

    <slide>

  </slide>
```

```
<core>
</core>
</block>
</tma>
</HISTO>
```

In Example 3, there are two errors. The 'HISTO' element is in uppercase. As in all XML, elements tags are case sensitive, and the root element, histo, must appear as lowercase. Additionally, the 'header' element is nested under the block element. There are very few nesting rules in the TMA specification, but the specification requires a separate section for header and for block. The 'header' tag must be the first TMA CDE following the 'tma' tag. Users may add their own tags that precede the 'header' tag after the 'tma' tag. Non-CDE tags are tags that are not included in Table 3, and are simply ignored by the TMA validator. The block element may contain only block, slide and core CDEs.

Example4.xml

Example 4 illustrates use of the self-describing nesting hierarchy of CDEs taken from Table 3. In this example, values for several of the tags are added.

```
<histo>
  <tma>
    <header>
      <Title>All-Purpose Tissue Array</Title>
    </header>
    <block>
      <block_identifier>15</block_identifier>
      <slide>
        <slide_identifier>5534</slide_identifier>
      </slide>
    <core>
      <core_array-id>22</core_array-id>
      <core_histo-repository>Generic Tissue Bank
```

```
<core_histo-repository_donor-block>99945
  <core_histo-repository_donor-block_drill-site>78,90
    <core_histo-repository_donor-block_drill-site_diagnosis>bcc
  </core_histo-repository_donor-block_drill-site_diagnosis>
</core_histo-repository_donor-block_drill-site>
</core_histo-repository_donor-block>
</core_histo-repository>
</core>
</block>
</tma>
</histo>
```

In Example 4, there is a hierarchy of core CDEs, and the order of the hierarchy is specified by underscore characters within each CDE.

When run through the Perl validator script (validtma.pl), Example 4 produces the following results, consisting of a list of the used tags, a statement that the XML file parsed as a valid TMA file, and an authentication string for the file.

```
Begining to parse ex4.xml now
histo
tma
header
Title
block
block_identifier
slide
slide_identifier
core
```

core_array-id
 core_histo-repository
 core_histo-repository_donor-block
 core_histo-repository_donor-block_drill-site
 core_histo-repository_donor-block_drill-site_diagnosis

Finished. ex4.xml is a valid Tissue Microarray File.

The one-way hash of your file is
 bb0aa66660074cd3322ff339c7a45d47

TMA CDEs beginning with block, core, or slide and containing underscore characters contain a self-describing hierarchical XML nesting pattern.

core_histo-repository_donor-block_drill-site_diagnosis, nested under
core_histo-repository_donor-block_drill-site, nested under
core_histo-repository_donor-block, nested under
core_histo-repository, nested under
core

This means that once an hierarchical TMA CDE from Table 3 is chosen, the user is committed to include the ancestor CDEs. In the case of the CDE, core_repository_donor_block_drill-site_diagnosis, the ancestors shown above would need to be included in the TMA file with nesting as shown in Example 4. As always, the insertion of user-created elements is ignored by the validator, even when those elements interrupt a nested hierarchy. The Perl script for the TMA validator is included as an attachment file with this article [see Additional file 2].

Discussion

Data Exchange Specifications are written so that databases related to a specific data domain may be designed with common data and common data structures. Standards are now available for creating XML documents [6] and CDEs [7]. Data exchange specifications that conform to XML and CDE standards and receive support from the user community become powerful research devices. The availability of large numbers of TMA files conforming to the data exchange specification will permit the inter-laboratory comparison of TMA data and the integration of TMA data with data from other biological databases. Research-

ers will be able to submit their TMA data as supplemental files with their research publications so that reviewers and readers can examine the original research data. Because the specification provides a way to produce a self-describing file, it would be a simple matter to port the data from TMA files into virtually any commercial or open source database.

Specifications are only adopted when they fill the needs of a heterogeneous user community [8]. The community of TMA users includes: pathologists, research scientists, informaticians, commercial tissue repositories, and journal editors. In order to develop a set of standards that will appeal to all these groups, the TMA Data Exchange specification was developed in a series of open workshops. One of the early concerns was that advances in TMA technology would be stifled by a rigid standard containing a list of required Common Data Elements and a required data structure. A second concern was expressed by researchers who were using proprietary database implementations. These participants wanted the freedom to add [proprietary] data elements to TMA exchange documents without violating the specification. They also indicated that they required a loose data structure that could be easily re-constructed from their own databases.

The most common way of specifying properties of an XML document is through a DTD (Data Type Definition) or a Schema [5,6]. In fact, the workshops considered several versions of a DTD without obtaining approval at any of the workshops. The group's emphasis on flexibility and open design of the specification, particularly the requirement that users be allowed to add their own tags, made writing a DTD difficult. The spectrum of workshop participants, which included pathologists, imaging experts, and tissue bankers, was not particularly focused on the technicalities of XML. We fully expect that users will eventually move toward a DTD or schema to support the TMA specification.

The current draft of the TMA Data Exchange specification satisfies user requests for maximum flexibility. With freedom comes responsibility. It is quite possible to design a TMA data file that conforms to the TMA data exchange specification but lacks annotational detail. Of the eighty TMA CDEs in Table 3, only six are required. A TMA that lacks detailed identifying information for blocks, slides, cores and images may pass inspection by the validator, but it would be of little scientific value. Similarly, a TMA that lacks a full set of header information (Title, Creator, Subject and Keywords, Description, etc.) cannot be sensibly indexed or shared. Depending on comments from implementers, the specification may need to expand the number of required CDEs.

Those who wish to use the specification will need short programs that map their pre-existing database elements over to the equivalent data elements of the TMA Data Exchange specification. The programs will need to adhere to the general XML structure described in this paper. For elements with no corresponding listing in the specification, researchers will be expected to create their own metadata tags for their TMA files. It is hoped that TMA files will soon be available to the public as supplemental attachments to journal submissions. This means that researchers will need to exclude identifying information from shared TMA files. These considerations place additional burdens on researchers. The TMA Data Exchange Specification provides software programmers with a promising new area for development.

Conclusion

The TMA data exchange specification is now available in a draft form with community-approved Common Data Elements and community-approved general file format and structure. The specification can be freely used by the scientific community. It is designed to be independent of the source of the data, including the source of the tissue, the experimental protocol, the imaging modality, the data capture method, and the schema for internal storage. The metadata file of fully described TMA CDEs (*tma_cde.htm*) and a Perl implementation of the validator (*validtma.pl*) are attached as supplemental files with this article. Efforts sponsored by the Association for Pathology Informatics to establish a TMA data exchange specification are expected to continue for several more years. The interested public is invited to participate in these efforts.

Author Contributions

Dr. Berman wrote the first draft of the article and wrote the Perl validator for the current TMA specification draft, chairs the API Technical Standards Committee that launched the TMA data exchange specification and chaired the first TMA Data Exchange discussion held at the U of Michigan AIMCL meeting, May 30, 2002. Dr. Mary Edgerton wrote the first draft of TMA common data elements, helped design the TMA data structure. Dr. Edgerton was Principal Investigator for the National Cancer Institute grant R13 CA094453, supporting the TMA workshops, and chaired three of the four TMA data exchange specification workshops. Dr. Friedman is the current President of API and conceived the idea of a series of community-based workshops to further TMA informatics. Dr. Friedman helped plan, organize and arrange funding for the TMA informatics workshops. Dr. Friedman has reviewed and corrected all drafts of the article.

None of the authors have any financial interest in this article or in the TMA data exchange specification.

Additional material

Additional File 1

TMA Common Data Elements HTML file of each data element from Table 3 annotated with metadata descriptors (tma_cde.htm).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-3-5-S1.htm>]

Additional File 2

TMA Validator Perl Script for validating TMA files with the TMA Data Exchange Specification (validtma.pl).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-3-5-S2.pl>]

Acknowledgements

The TMA data exchange standards effort was co-sponsored by The Association for Pathology Informatics and by the National Cancer Institute. Some of the costs of meeting workshops were underwritten by the APIII and by the AIMCL meetings.

References

1. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G and Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens** *Nat Med* 1998, **4**:844-847.
2. Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C and Lepage M: **Design and implementation of microarray gene expression markup language (MAGE-ML)** *Genome Biol* 2002, **3**:Research0046.1-0046.9.
3. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA and Causton HC: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data** *Nat Genet* 2001, **29**:365-371.
4. Perou C: **Show me the data!** *Nat Gen* 2001, **29**:373. (inclusive)
5. White C, Quin L and Burman L: **Mastering XML: Premium Edition** Sybex, San Francisco 2001.
6. Ahmed K, Ayers D, Birbeck M, Cousins J, Dodds D, Lubell J, Nic M, Rivers-Moore D, Watt A, Worden R and Wrightson A: **Professional XML Meta Data** Wrox Press Ltd, Birmingham 2001.
7. Solbrig HR: **Metadata and the reintegration of clinical information: ISO 11179 MD Comput** 2000, **3**:25-28.
8. Linstone HA: **Multiple Perspectives for Decision Making: Bridging the Gap Between Analysis and Action** North Holland, New York 1984.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6947/3/5/prepub>