An XML Cytometry Standard Based on DICOM

Robert C. Leif[a]*

[a]XML_Med, a Division of Newport Instruments, 5648 Toyon Road, San Diego, CA, USA 92115-1022

ABSTRACT

Introduction: The International Society for the Advancement of Cytometry (ISAC) Data Standards Task Force (DSTF) is developing a new Advanced Cytometry Specification (ACS). DICOM has developed and is extending a pathology extension. The work of both groups is complementary with some overlap. Interoperation would benefit both groups and permit each to benefit from the other's expertise. Methods: The design and implementation of the CytometryML version of the ACS schemas have been based on each schema describing one object (modularity), iterative (spiral) development, inheritance, and reuse of data-types and their definitions from DICOM, Flow Cytometry Standard, and other standards. Results: These schemas have been validated with two tools and XML pages were generated from highest level schemas. Binary image data and its associated metadata are stored together in a zip file based container. A schema for a table of contents, which is one of the metadata files of this container, has recently been developed and reported upon. The binary image data is placed in one file in the container; and the metadata associated with an image in another. The schema for the image metadata file includes elements that are based on the DICOM design. This image schema includes descriptions of the acquisition context, image (including information on compression), specimen, slide, transmission medium, major optical parts, optical elements in one or more optical channels, detectors, and pixel format. The image schema describes both conventional camera systems and scanning or confocal systems.

Keywords: ISAC, DICOM, CytometryML, XML, schema, image, metadata, channel

1. INTRODUCTION

Medical-scientific standards' developers from different groups should cooperate and collaborate because no one of our groups has sufficient domain knowledge to encompass the continuum of research, laboratory, and clinical informatics. Technology transfer from the researcher to the clinician can be facilitated and expedited when both employ an overlapping, interoperable software base. Therefore, it would be beneficial to all groups to maximize the interoperability of standards produced by various scientific and medical societies, such as the International Society for the Advancement of Cytometry (ISAC) Advanced Cytometry Standard (ACS)[1,2], with both Digital Imaging and Communications in Medicine (DICOM)[3,4] and HL7 Version 3[5]. One of the best ways of achieving this goal is to employ common data-types including their descriptions. Because of its ubiquity, XML is the obvious syntax to employ for describing measurements and the data generated by medical-scientific instruments. It should be cautioned that because of the present very large investment in medical software and the need for high-reliability, it is both impractical and potentially hazardous to make rapid changes in software technology. Laboratory information systems and related entities, such as picture archiving systems, are complex and are based on existing standards, such as DICOM, which have a long lifetime, and span the domain knowledge of multiple professional groups. Thus, a plan based on reuse of existing designs and evolution rather than revolution is required.

In the case of cytometry or quantitative cell analysis, there is a strong overlap between the interests of pathologists and cytometry scientist-engineers. Pathologists use the technology, which is often invented by the cytometrists. In general, there is a difference in emphasis between the two groups in that pathologists have a greater interest in morphology and cytometrists have a greater interest in molecular markers. However, the interests of both groups are an overlapping continuum. This difference is demonstrated by the difference of content of the work products of the ISAC Data Standards Task Force (DSTF), such as the ACS[2], and the content of DICOM (ftp://medical.nema.org/medical/dicom/2008/). Much of the content of the ACS[2], concerns the detailed instrumentation, sample processing, and algorithms for the quantitation

*rleif@rleif.com; 1 619 582-0437; newportinstruments.com

of fluorescence; whereas, DICOM optical imaging is directed to measurements of three or fewer colors and clinical practice including the concept of patients and a clinical environment. The subject of this paper is the development of a cytometry standard for images that can, at least potentially, interoperate with DICOM and provide a model for the long-term evolution of DICOM into an XML based standard. Since the ISAC DSTF has made flow cytometry its major priority, the work described in this paper should be considered preliminary.

The capacity to reuse existing standards including DICOM and translate them into XML schemas has previously been demonstrated[6]. Consensus at ISAC has been achieved that there should be a combined flow and image cytometry standard[7]. The technology developed for CytometryML is based on the realization that DICOM can be viewed as an informal ontology. The DICOM standard[3] provides a reasonably rigorous and exhaustive organization of the medical imaging knowledge domain in a hierarchical format and contains much of the relevant entities and their relations, as well as including other standards to extend its base. DICOM often provides sufficient documentation to obviate or, at least, minimize the need for the creation of a separate text description of data-types.

CytometryML and the ISAC ACS differ from that of the Open Microscopy Environment (OME)[8,9], which has produced very useful and impressive research tools. In that, the ACS requirements[7] will 1) lead to a standard that encompasses both digital microscopy and flow cytometry; 2) be significantly based on the reuse of DICOM and other clinical data standards and 3) the ACS has been designed to be used in instruments that conform to existing regulatory requirements.

Recently a version of the ACS container has been proposed[10], The ACS container is a zip file which includes either the files that describe or pointers (URIs) to a cytometry measurement. These contents include the table-of-contents, the binary data containing file(s) that contain the image and/or list-mode (waveform) cytometry data, as well as other metadata files that describe the: image, list-mode data, instrument, specimen, data manipulation, data analysis, structured reports[11] and other types of documents. The major schemas include: a table of contents and descriptions of the files that contain the binary data (image and/or list-mode), series, instrument (microscope or flow-cytometer), compensation, gating method, analysis method, as well as the experiment overview, sample and specimen details, results, report (structured report), and other file(s). A list-mode file, which is the traditional output of a flow cytometer, contains the content of a multiparameter waveform[12]. Each of the major schemas listed in the table of contents is in turn supported by other (helper) schemas. DICOM tags and value representations have been included as attributes in the DICOM based data-types of the CytometryML schemas.

2. **METHODS**

The design of the schemas was modular[13,14], iterative, and essentially followed the spiral method of development[15]. The existing CytometryML schemas were updated to conform to ISAC DSTF suggestions and consensus items. Other changes were made in accordance with the design of DICOM working groups 26 and 27. A simple object based design methodology was employed; and the schema were treated as if they were the descriptions of the data-types in the classes or packages of object-oriented languages. The size of the individual schemas was kept manageable by budding off individual data-types that had become large and/or complex and their components into a new schema. Inheritance was possible by extension for data-types in both their parent schemas and other schemas that imported the parent schemas. Unfortunately, the present version of the XML schema standard does not permit restricting a complexType that is imported from one schema to another[16].

In order to maximize simplicity, the schemas were and are being constructed to be an acyclic graph. No schema was imported by any that it imports. Import diagrams was created from the individual CytometryML schemas by transferring one of them from Altova XMLSpy to Altova Schema Agent. The vector image created by the Altova Schema Agent was then stored as a Windows enhanced metafile (EMF) file. This EMF file was extensively formatted in VISIO (http://office.microsoft.com/en-us/visio/default.aspx).

Whenever possible, data-types were reused from existing standards[1]. These standards included: Digital Imaging and Communications in Medicine (DICOM)[3], ISAC Flow Cytometry Standard[17] (FCS), ECMA-International (http://www.ecmainternational.org/), Human Resources Consortium (http://www.hr-xml.org), and the Open Microscopy Environment (OME) (http://www.ome-xml.org/) Data-types were also reused from designs, which were or are under

development, such as those of DICOM WG 26 and WG 27, or the ISAC DSTF. Data-types previously present in CytometryML were also reused. Data-types were created only when they did not exist in another standard; or the existing data-type(s) was inappropriate. The search for an existing data-type began with the DICOM Data Dictionary, Part 6, which was searched for strings that partially matched the description of the required data-type. Once the data-type was found, its tag was used to search DICOM Information Object Definitions, Part 3[3]. The table that described the attributes of the Information Object Definition (IOD) was used as the basis of an XML schema and the definitions of the attributes were reused as documentation elements in these schemas. Numerical types were reused from ECMA-International[18] because of their clarity and independence from the choice of programming language. Int32 is intuitively obvious. Double could refer to some size of integer, such as 32 bits or double precision float (Float64).

The CytometryML schemas were developed using the XML schema definition language, XSDL[19,20], and were validated with both Stylus Studio (http://www.stylusstudio.com) and XMLSpy (http://www.altova.com). These schemas are primarily derived from DICOM datatypes with XSDL documentation elements that included references to the descriptions of the datatypes in the DICOM standard[3] and datatypes that could be part of an extension of DICOM. The visibility of the source of datatypes and elements has and is being maximized by importing rather than including all schemas. A schema's xlmns attribute value including its prefix was reused when the schema was imported into other schemas. The numerous, ubiquitous xs: prefixes have been eliminated by not including a prefix in the value of the xlmns attribute for schema and instead including a prefix based on the name of the schema for the xlmns attribute that makes the schema's elements visible unto itself.

3. RESULTS

3.1 DICOM Tags and Value Representations

The DICOM Tag is the unique value that identifies DICOM data elements. The simplest and most efficient way to include Tags in XSDL is as attributes[21]. XML attributes are a simpleType. simpleTypes describe data-types that consist of a single entity, such as a string, a number, or an enumeration. The declaration of a Tag_Type is shown in Code Fragment 1.

Code Fragment 1 (Schema)

```
1<simpleType name="Tag_Type">
2 <restriction base="string">
3    <pattern value="[0-9a-fA-F]{4},[0-9a-fA-F]{4}"/>
  </restriction>
</simpleType>
```

Element 1 assigns the name Tag_Type to a simpleType, which will be in the form of a string (element 2). This string is restricted by the pattern (element 3) to two pairs of 4 hexadecimal characters, which are separated by a comma. The letters a-f and A-F represent 10-16. The range of characters is shown within the square brackets [ ] and the number of characters, 4, is shown within the braces or curly brackets { }. The two strings with </ in front of them respectively end the restriction and end the simpleType elements.

Since the Value Representation (VR) identifies the type of a data-type, it is an enumeration, which consists of two letter abbreviations (Code Fragment 2).

Code Fragment 2 (Schema)

```
1<simpleType name="VR_Type" id="VR_Type">
2 <restriction base="string">
3    <enumeration value="AE"/> <!--Application Entity-->
4    <enumeration value="AS"/> <!--Age String-->
5    <enumeration value="AT"/> <!--Attribute Tag-->
6    <enumeration value="CS"/> <!--Code String-->
```

```
7     <enumeration value="TM"/> <!--Time-->
<!--.....................................................-->
23    <enumeration value="UI"/> <!--Unique Identifier (UID)-->
24    <enumeration value="UL"/> <!--Unsigned Long-->
25    <enumeration value="US"/> <!--Unsigned Short-->
26    <enumeration value="UT"/> <!--Unlimited Text-->
   </restriction>
</simpleType>
```

Each of the VR values has been enumerated together with a comment at the right, which consists of the VR's name. Elements 8 through 22 have been omitted.

3.2 Example of the Creation of a DICOM Data-Type in XSDL

The creation of a data-type for the DICOM Code Value type involved 2 steps. The first was the creation of the underlying data-type, which is a 16 character string. The token type is a restriction on the XML string type.

Code Fragment 3 (Schema)

```
1 <simpleType name="Bd_16_Type">
2  <restriction base="token">
3     <minLength value="1"/>
4     <maxLength value="16"/>
   </restriction>
</simpleType>
```

The Bd_16_Type (element 1) is a bounded string with a lower bound of 1 character (element 3) and an upper bound of 16 characters (element 4). This means that if a value of this data-type has to be entered by a user, the user must entire at least one character. A subject for future discussion is the best choice for a lower bound; the user could be forced to enter more than 1 character. The standard DICOM attribute Short_String_Type was created by restriction Bd_16_Type, which was effectively a renaming.

Code Fragment 4 (Schema)

```
1<annotation>
2  <documentation>(0008,0100) Code Value SH value multiplicity = 1.
PS 3.3, 8.1 "CODE VALUE: The Code Value (0008,0100) is an identifier that is unambiguous within the Coding Scheme
denoted by Coding Scheme Designator 0008,0102) and Coding Scheme Version (0008,0103).

Note: The Code Value is typically not a natural language string, e.g. “T-04000”.
   </documentation>
</annotation>
3<complexType name="Code_Value_Type">
4  <simpleContent>
5     <extension base="dicom:Bd_16_Type">
6        <attribute name="Tag" type="dicom:Tag_Type" fixed="0008,0100"/>
7        <attribute name="VR" type="dicom:VR_Type" fixed="SH"/>
      </extension>
   </simpleContent>
</complexType>
```

Code Fragment 4 is an example of how a specific DICOM data-type is encoded together with its Tag and VR. It begins with the XSDL pair of elements (1 and 2) that are used to document code. The first line of element 2 was copied out of PS 3.6-2008, 6 Registry of DICOM data elements. The rest was copied from PS 3.3, 8.1 CODE VALUE. A complexType

(element 3) was required because the data-type includes 3 parts. The actual data-type (element 5) and the DICOM Tag and VR, respectively elements 6 and 7. The Tag and VR can be represented as attributes because they are both based on simple types. The use of an attribute permits the use of a simpler syntax than that of an element[19,20]. presently the use of the Tag and VR attributes has been limited to the schemas. These attributes have been omitted from the XML pages because they decreased the readability of the XML pages. The prefix dicom: in element 5 indicates that the data-type is present in a schema that has dicom as the prefix in its XML name space declaration (xmlns:).

This approach has the advantages of: minimizing the costs of and effort for design, coding, documentation, and testing; reducing the probability of regulatory problems and errors. It also maintains consistency with DICOM and facilitates the interoperation of XML and DICOM by providing two keys, the Tag and VR, that can be used to facilitate translation.

3.3 Example Image_Context

The Image_Context_Type described below in Code Fragment 5 is a complexType that describes the contextual information, which can change for each image (instance). The information that is unchanged needs to be provided only once, as XML documents, in the series container or pointed to by a Uniform Resource Identifier (URI) in the series file.

Code Fragment 5 (Schema)

```
1<complexType name="Image_Context_Type">
2  <sequence>
3     <element name="Acquisition_Context"
        type="acquisition:Acquisition_Context_Type"/>
4     <element name="Image_Specific_File"
        type="formats:Image_Specific_File_Type"/>
5     <element name="Application" type="specimen:Application_Type"/>
6     <element name="Slide" type="slide:Slide_Info_Type"/>
7     <element name="Transmission_Medium"
        type="trans:Transmission_Medium_Info_Type" minOccurs="0"/>
8     <element name="Optical_Parts" type="context:Optical_Parts_Type"/>
9     <element name="Channel_Reference"
        type="channels:Channel_Reference_Type"/>
10    <element name="Detector_Info" type="context:Detector_Info_Type"/>
11    <element name="Pixels" type="pixel:Pixel_Type"/>
   </sequence>
</complexType>
```

The Acquisition_Context (Element 3) is located in the acquisition schema (not shown). The Acquisition_Context element includes the following items: 1) The Acquisition Number, which is sequentially changed (incremented) for each run of the instrument independently of the measurement being performed. The augmentation of the Acquisition Number by the inclusion of a DICOM unique identifier (UID) is open for discussion. 2) The Instance_Number is an integer that is increase by 1 for every set of binary data included in a series. 3) The Instance_Role can be the traditional member of a series, a single measurement that is not a member of a series, a control, a variable, which is a means to permit storage of a single XML page at the level of a series and permit a single element or attribute that is based on a simpleType, such as concentration or time, to have different values for different instances[10], and other, which can be anything; 4) The URI of the series XML document for the instance has also been included. The Instance_Role is presently a subject under discussion by the ISAC DSTF.

The information that is contained within many conventional image files is unavailable without a program that will translate this information into XML. One way to circumvent this problem is to provide this information specific for each image as part of a standardized XML document, such as the Image_Specific_File (Element 4 of Code Fragment 5. (Schema) and Element 1 of Code Fragment 6. (XML), which is shown below.

Code Fragment 6 (XML)

```
1  <Image_Specific_File>
2    <formats:File required_to_process="true" File_Name="TIFF_Image.xml"
       File_URI="http://www.newportinstruments.com/cytometryml/cytometryml.htm"
       Format="TIFF">
3      <Instance_Creation_Date_Time>2008-12-27T09:30:47.0Z
       </Instance_Creation_Date_Time>
4      <UID Value="1.9.749.95558.1.1.1.1.1"/>
5      <Originality>
6        <files:Originality>Original</files:Originality>
       </Originality>
7      <Creator>
8        <files:Person_Name>
9          <about:PreferredGivenName>Robert</about:PreferredGivenName>
10         <about:MiddleName>C</about:MiddleName>
11         <about:FamilyName>Leif</about:FamilyName>
12         <about:qualification>Ph.D.<about:qualification/>
13         <about:Generation/>
         </files:Person_Name>
       </Creator>
14     <Description>This is the type of file produced by many laboratories for
       publication. It normally would not be the original data, which would be
       uncompressed</Description>
     </formats:File>
15   <formats:TIFF_Image_File File_Extension="TIFF" File_Content="Mixed"/>
16   <formats:Compression>
17     <formats:Lossy_Image_Compression>
18       <formats:Lossy_Image_Compression_Ratio>5
         </formats:Lossy_Image_Compression_Ratio>
19       <formats:Lossy_Image_Compression_Method Image_Quality="10">JPEG_Lossy
         </formats:Lossy_Image_Compression_Method>
       </formats:Lossy_Image_Compression>
     </formats:Compression>
20   <formats:Image_Componants Examination_Characteristic="Primary"
       Binning_Secondary="2">
21     <Viewing>Mono</Viewing>
22     <Rows>512</Rows>
23     <Columns>512</Columns>
24     <Photometric_Interpretation>
25       <Multispectral>Multi-Spectral</Multispectral>
       </Photometric_Interpretation>
26     <Planar_Configuration>color-by-pixel</Planar_Configuration>
     </formats:Image_Componants>
   </Image_Specific_File>
```

The Image_Specific_File (Element 4 and XML Code Fragment 6) describes the file, starting with Element 2. The data analysis requires the image file to be processed and the file name, URI and format are provided. The creation date and time (Element 3), the value of an optional DICOM UID (Element 4), originality (Element 6), which also can describe data

derived from a parent file (not shown) are provided. The creator of the data (Element 7) can be either a person or computer. The name of a person is shown in Elements 8-13. This is followed by a description (Element 14).

The detailed description of the actual file format: TIFF, JPEG, JPEG2000, DICOM, Raw, and Other, as well as the extension, is included in (Element 15). Since images obtained with a microscope are large binary objects, it is reasonable to keep the actual image data in a binary format. This minimizes memory usage, disk space, and transmission time. It also maximizes software processing efficiency by eliminating the necessity for conversion of the data. Although the actual data elements are binary, other data in formats, such as TIFF[22,23] also include text. Thus, these files are in a mixed format. Both lossy (Element 17) and lossless (not shown) compression have been included and characterized including the compression ratio (Element 18) and quality (Element 19) of the compressed image. Other elements in the Image_Componants (Element 20) describe: 1) Whether the image is primary (created as a direct result of the examination) or secondary (after the initial examination). 2) The binning subsequently used to produce the file that was not originally performed by the camera. Element 21 describes whether the image is single or a member of a stereo image pair. Elements 22 and 23 describe respectively the number of pixels in the rows and in the columns. The intended format of the pixel data (monochrome, color (RGB etc), or multispectral) is given in Element 25. Multispectral includes any positive number of measurements per pixel. The planar configuration (Element 26) can either be color-by-pixel, where the pixel is a vector or color-by-plane, where the pixels in each plane only include a single value.

Since the information in Image_Specific_File on the image and pixels is sufficient to decode a raw image file, the need for mixed binary and text files has been significantly reduced. The use of ACS containers with only XML and pure (raw) binary files will facilitate the development of new image software by scientists because it removes the complex operation of extracting the raw data from a mixed file.

3.4. Description of the Individual Elements in Code Fragment 5 from the Image_Context Schema

Returning to the Image_Context_Type in the Image_Context_Type.xsd (Code Fragment 5. (Schema). The individual elements will be described below and to facilitate their description they are repeated above the paragraph that describes them.

5. <element name="Application" type="specimen:Application_Type"/>

The applications of interest to the ISAC membership include many that are unrelated to medicine. The present applications include: human, veterinary, environmental, and other samples. In the case of human specimens, the following items are included: accession number, specimen identifier, patient ID, organ and organ part, information on the container including its accession number, barcode and type: jar, vial, cassette, slide, blood_tube, or other.

Since the container is a manufactured item, it and all of the other manufactured items described below can include: an XML identifier (ID), the names of the manufacturer and model, a description, and a DICOM UID.

6. <element name="Slide" type="slide:Slide_Info_Type"/>

This element includes the slide material: glass, fused silica (quartz) etc., whether the slide is transparent, dimensions, the location of the area: that is referenced, and the area that contains the specimen including its center point. The mounting medium and coverslip are also described.

7. <element name="Transmission_Medium"

type="trans:Transmission_Medium_Info_Type

The type of transmission meeting is specified: air, oil, glycerol, water, optical cement, proprietary, unknown, and other.

8. <element name="Optical_Parts" type="context:Optical_Parts_Type"/>

This includes the objective and condenser, as well as the apertures. The numerical aperture, contrast mechanism, if any, and immersion medium of the objective and condenser are specified and the color correction and flatness of field of the objective is also specified.

9. <element name="Channel_Reference"

Each type of measurement made with the imaging device and the configuration of the device for an individual measurement is described by an individual channel, which is associated with a specific number. The configuration

includes and provides the order in the light path through the optical elements, such as filters and dichroic mirrors. A description of the light source is also included.

10. <element name="Detector_Info" type="context:Detector_Info_Type"/>

There are two classes of detectors: cameras, which produce images, and detectors that produce single measurements or unidimensional arrays of single measurements (Numeric_Info). The cameras are specified in terms of the technology used to manufacture their sensors: CCD, EM_CCD, and CMOS and the presence of an intensifier and/or back-thinning. The cameras are also described by their numbers of columns and rows, binning (binning can be done in most cameras and separately in software), exposure duration, temperature during the exposure, and wavelength cut-offs at a specified quantum efficiency.

The single measurement detectors include: PMTs, multi-anode PMTs, diodes, avalanche-diodes, diode-arrays, voltage, current, position, software (for calculated values), and other. Pulse measurements, detector settings, the numbers of bits actually needed to be stored and allocated can also be specified.

A description of an amplifier is provided for both types of detectors. The mode (linear, log, derivative, and other) and the gain can be specified.

11. <element name="Pixels" type="pixel:Pixel_Type"/>

There are two types of pixels: vectors composed of single parameter data, presently numeric data, and spectra, which are arrays of single parameter data. The Pixel Aspect_Ratio and number of samples (channels) per pixel can also be specified. The numeric data-type for each parameter (component) of the vector or the parameter of the array is specified including the resolution of the measurement (number of bits stored) and the number of bits of the bytes required to store the data.

4. DISCUSSION

As was shown in this preliminary description of the reuse of DICOM in XML, it has been feasible to create XML schemas that in large part are based on DICOM designs, data structures, and text. DICOM sequences can be mapped into XML schema complexTypes; however, the mapping of DICOM objects should yield precedence to existing XML data structures.

Since XML schemas do not include any methods (functions or procedures), these could not be included in the CytometryML schemas. However, the XML files could be in the form of either an office type product, which is an XML application and can be validated against XML schemas, such as Microsoft Word or an XML form based on either XForms[24] or XHTM™ Modularization 1.1[25].

The CytometryML schemas besides being prototypes are incomplete. This incompleteness is inevitable when a member of one group (society) with expertise in cytometry authors a project that it is in need of another group's (physicians) knowledge. A means needs to be developed to permit each group to create standards for its own area of expertise and have these standards interoperate with and thus complement other groups' standards. Modular XML schemas appear to be a reasonable construct to permit this interoperability. If groups are going to cooperate including sharing the use of software, then it is absolutely essential that the totality of the software (code, design, descriptions, etc.) be understandable by individuals who are not expert in the fields of the developers. Overuse of abbreviations, use of jargon, and obscurity must be minimized. Understandability must be a major requirement.

Reuse is essential. Firstly, the design cost of reliable software is quite significant. Although extensive good experience with a software construct cannot prove correctness, it does provide significant probability of reliability and thus should provide increased credibility with regulatory authorities. Wherever reasonable, semantics should be preserved; however, syntax has to be changed when DICOM is translated into XML. Translation also provides the opportunity to retire obscure terminology and replace it with easier to comprehend terminology. For instance, a terminology for numerical types based on that created by ECMA was employed instead of that of DICOM or that presently existing in XML. Int32 is a compact notation and is clearly a 32 bit integer; whereas, the common term double could either be a 32 bit integer or a 64 bit floating point number. There is also the problem of how to extend the current terminology for 64 bit integers (Int64). DICOM has a similar problem with its bounded strings. It is not intuitively obvious what the number of bytes is in a Short String, Long String, Short Text, and Long Text are; however, the number of bytes in Bd_16, Bd_64, Bd_1024 or Text_1024, and Bd_10240 or Text_10240 is clear.

One question that needs to be answered is the role of the DICOM Tags and VRs in XML documents? Presently, solely for the purpose of avoiding clutter, these two attributes and their values have been omitted from the XML documents. Is it sufficient to limit them to the schemas or should they be included in the XML pages validated by the schemas? Should every object have a UID and what is the status of this UID? Is it mandatory or optional?

One of the present models for the Advanced Cytometry Standard permits information to be either stored in the Series XML page or the Instance XML page depending on whether it was constant for the entire series. Is this model acceptable?

Often, schemas have been created in an attempt to fully describe an object or a cohesive collection of objects: strings, numeric types, units, etc. or other generalities, such as light, optics, or chemistry. In some cases, it is hoped that these will serve as place holders that will be replaced by the appropriate group (scientific society). CytometryML was originally created to encode in XML a very small part of DICOM that is highly relevant to Cytometry and has been extended to include what DICOM lacked.

5. CONCLUSIONS

The XML page based on the Image_Context schema contains the necessary metadata directly related to an image for that image to be of use to a researcher or clinician.

It has been possible to reuse DICOM in XML schemas. DICOM should eventually evolve into an XML based standard. However, this should be an evolution not a revolution. For the present and near future, it should be possible for DICOM and XML applications to interoperate. A start on this has been undertaken by DICOM Working Group 27. The development of a medical-scientific XML software continuum probably will require 1) coordination between groups (societies), 2) agreed upon style guides, and 3) Software quality assessment procedures.

Modularity has significantly facilitated the development of the CytometryML schemas. Migration of complexTypes from one schema to another have been made with little effort. It has been possible to reuse elements from an original set of helper schemas that act as libraries. Modularity will minimize maintenance cost and maximize and reuse. Modularity permits one person to work on one schema, while someone else works on another schema. This capacity to work in parallel should speed development and shorten the time to fix bugs.

The only item that the author of this paper is absolutely convinced of is that his own work and/or the collective work of the ISAC DSTF and similar organizations will inevitably omit essential material and that the nature of science and technology will result in these omissions becoming obvious, which will require that they be promptly fixed and added to the ACS. For example, one major omission is how should data be accessed over the Internet? Fortunately, Internet access is being worked upon by DICOM Working Group 27, which is updating DICOM Part 18: Web Access to DICOM Persistent Objects (WADO). This has resulted in some XML becoming visible to DICOM users. Other omissions are whole slide and three dimensional images. Other just as important material is probably unknown to the author and/or the other members of the DSTF. The extended DICOM WADO update should facilitate a microscope that has software based on the ISAC ACS (XML) standard exchanging messages including data with DICOM. This capability could be extended to flow cytometers.

ACKNOLEDGEMENTS

This work was sponsored by Newport Instruments. The author wishes to thank the members of the ISAC DSTF and members of DICOM Working Groups 26 and 27 for providing knowledge respectively of the ISAC ACS and of DICOM. However, any mistakes concerning either the ACS or DICOM are the author's own. The views expressed are solely those of the author and need not be than of any group of which he is a member.

REFERENCES

1] Leif R. C., Spidlen J., Brinkman R. R., "Cytometry standards continuum," Proc. SPIE 6859, 68590Q-1-8 (2008).

2] Lee J. A., Spidlen J., Boyce K., Cai J., Crosbie N., Dalphin M., Furlong J., Gasparetto M., Goldberg M., Goralczyk E. M., Hyun B., Jansen K., Kollmann T., Kong M., Leif R. C., McWeeney S., Moloshok T. D., Moore W., Nolan G., Nolan J., Nikolich-Zugich J., Parrish D., Purcell B., Qian Y., Selvaraj B., Smith C., Tchuvatkina O., Wertheimer

A.,Wilkinson P., Wilson C., Wood J., Zigon R., Scheuermann R. H., and Brinkman R. R., “MIFlowCyt: The minimum information about a flow cytometry experiment," Cytometry A, 73A, 926-930, (2008).
3] “Digital Imaging and Communications in Medicine (DICOM) Part 3: Information Object Definitions, PS 3.3-2006, National Electrical Manufacturers Association, NEMA,” Available at: http://medical.nema.org/dicom/2008/ (2008).
4] Pianykh O. S., [Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide], Springer Publishers, Berlin & Heidelberg, (2008).
5] “HL7 Standard Version 3 Ballot January 2009,” Available at: http://www.hl7.org/v3ballot/html/welcome/introduction/index.htm (2008).
6] Leif R. C., Leif S. H., Leif S. B., “CytometryML, a markup language for analytical cytology," Proc. SPIE 4962, 288-297 (2003).
7]Spidlen J., Brinkman R. R., Leif R. C., and other members of the ISAC Data Standards Task Force, “Advanced Cytometry Standard (ACS) Requirements for a data file standard format to describe cytometry and related analytical cytology data, Version 0.070920," Available at: http://superb-east.dl.sourceforge.net/sourceforge/flowcyt/Requirementsv070920.pdf (2007).
8] “Open Microscopy Environment (OME),” Available at: http://www.openmicroscopy.org/site
9] Goldberg I. G., Allan C., Burel J.-M., Creager D., Falconi A., Hochheiser H., Johnston J., Mellen J., Sorger P. K., Swaddle J. R, “The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging,” Genome Biology, 6, R47 (2005).
10] Leif R. C., Spidlen J., Brinkman R. R., “A Container for the Advanced Cytometry Standard (ACS),” Proc. SPIE 7264, (2009).
11] Clunie D. A., [DICOM Structured Reporting], PixelMed Publishing, Available at: http://www.pixelmed.com/srbook.html (2000).
12] Leif R. C., “CytometryML, Binary Data Standards," Proc. SPIE 5699, 325-333 (2005).
13] Parnas D. L., “On the Criteria To Be Used in Decomposing Systems into Modules," Communications of the ACM 15, 1053 - 1058, (1972).
14] Parnas D. L., Clements P. C., Weiss D. M., “Enhancing reusability with information hiding," ITT Proceeding of the Workshop on Reusability in Programming, cse.msu.edu. (1983).
15] Boehm B. W., “A spiral model of software development and enhancement,"Computer IEEE," 21, 61-72, (1988).
16] XML Schema Part 1: Structures Second Edition W3C Recommendation 28 October 2004. Available at: http://www.w3.org/TR/xmlschema-1/ (2004).
17] Seamer L. C., Bagwell C. B., Barden L., Redelman D., Salzman G. C., Wood J. C., Murphy R. F., “Proposed new data file standard for flow cytometry, version FCS 3.0,” Cytometry 28, 118–122 (1997).
18] “ECMA-335, 4th Edition / June 2006 Common Language Infrastructure (CLI) Partitions I to VI, Section 8.2.2 Built-in value and reference types, Table 1: Special Encoding”, ECMA International, Available at: http://www.ecma-international.org/publications/files/ECMA-ST/Ecma-335.pdf (2006).
19] XML Schema Part 2: Datatypes Second Edition, W3C Recommendation 28 October 2004 Available at: http://www.w3.org/TR/2004/RECxmlschema-2-20041028/ (2004).
20] Warmsley P., [Definitive XML Schema], Prentice Hall, Available at: http://www.phptr.com (2002).
21. Leif R. C., Leif S. B., “A DICOM Compatible Format for Analytical Cytology Data,” Proc. of SPIE 3260, 282-289, (1998).
22] TIFF, Revision 6.0 Final — June 3, 1992, Available at: http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf (1992).
23] AWARE SYSTEMS, “TIFF, Tag Image File Format, FAQ,” Available at: http://www.awaresystems.be/imaging/tiff/faq.html
24] “XForms 1.0 (Third Edition), W3C Recommendation 29 October 2007,” Available at: http://www.w3.org/TR/2007/REC-xforms-20071029/ (2007).
25] “XHTML™ Modularization 1.1, W3C Recommendation 8 October 2008,” Available at: http://www.w3.org/TR/2008/REC-xhtml-modularization-20081008 (2008).