Copyright 2015 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited. This paper will appear in the SPIE BiOS 2015 Proceedings as Paper Number 9328-41.

Progress on an Implementation of MIFlowCyt in XML

Robert C. Leif[*a], Stephanie H. Leif[a]

[a]XML_Med, a Division of Newport Instruments, San Diego, CA, USA 92117-3516

ABSTRACT

Introduction: The International Society for Advancement of Cytometry (ISAC) Data Standards Task Force (DSTF) has created a standard for the Minimum Information about a Flow Cytometry Experiment (MIFlowCyt 1.0). The CytometryML schemas, are based in part upon the Flow Cytometry Standard and Digital Imaging and Communication (DICOM) standards. CytometryML has and will be extended and adapted to include MIFlowCyt, as well as to serve as a common standard for flow and image cytometry (digital microscopy).

Methods: The MIFlowCyt data-types were created, as is the rest of CytometryML, in the XML Schema Definition Language (XSD1.1). Individual major elements of the MIFlowCyt schema were translated into XML and filled with reasonable data. A small section of the code was formatted with HTML formatting elements.

Results: The differences in the amount of detail to be recorded for 1) users of standard techniques including data analysts and 2) others, such as method and device creators, laboratory and other managers, engineers, and regulatory specialists required that separate data-types be created to describe the instrument configuration and components. A very substantial part of the MIFlowCyt element that describes the Experimental Overview part of the MIFlowCyt and substantial parts of several other major elements have been developed.

Conclusions: The future use of structured XML tags and web technology should facilitate searching of experimental information, its presentation, and inclusion in structured research, clinical, and regulatory documents, as well as demonstrate in publications adherence to the MIFlowCyt standard. The use of CytometryML together with XML technology should also result in the textual and numeric data being published using web technology without any change in composition. Preliminary testing indicates that CytometryML XML pages can be directly formatted with the combination of HTML and CSS.

Keywords: MIFlowCyt, CytometryML, DICOM, FCS, Cytometry, HTML5, structured document, XSD

1. Introduction

The International Society for Advancement of Cytometry (ISAC) created a standard, MIFlowCyt[1], for Flow Cytometry Publications. This standard provides a concise description of the technology and findings present in an article. The fact that an article meets the MIFlowCyt requirements needs to be signed off by the referee. Many applications of the this standard can be found in the Society's journal, Cytometry A.

MIFlowCyt[1] presently does not require sufficient information to repeat an experiment. It provides considerably less information than would be found in a well written Materials and Method section of an article and/or supplementary materials for publication. The information contained in a MIFlowCyt document presently is textural. Since all information is in the form of user entered strings, there is not a direct way to employ a computer program to determine if the completeness of the information is appropriate for MIFlowCyt. Simple evidence for the diversity of MIFlowCyt is the range in size of the MIFlowCyt statements, which a cursory investigation showed can range from essentially a single page to 8 and a half pages. The longer version of MIFlowCyt is available as a supplement of an article[2] that provided an exemplary use of MIFlowCyt. Presently, MIFlowCyt is a file or a short appendix to a scientific article, such as one published in Cytometry. An alternative structured format for MIFlowCyt is described below.

1.1. Advantages of Structured Data

MIFlowCyt can serve as a structured abstract and/or a table of contents that can in the future include hypertext linkages to the detailed description, which can be contained in the paper or supplementary material. Interested Parties: reviewers, editors, government agencies, such as the FDA, can automatically check for completeness in answering their requirements by checking for the presence of the required elements (objects). The appropriateness of the individual elements is

*rleif@rleif.com; phone 1 619 555-1234; cytometryml.org

enforced by strong typing. The display of the collection of elements that constitute a MIFlowCyt containing and/or based document is controlled by an external cascading style sheet (CSS) file. This capacity to change the display permits reuse including: archival, electronic publishing, print publishing, government reports and grant applications, as well as style consistency, where the editor and/or publisher control the style of the document. This would permit verified, publishable manuscripts to be submitted by the authors, which will reduce publishing costs including the author's time and effort. A simple example of need for this is the list of references at the end of this paper. The publisher employs an on-line parser to check the format of the references; however, like any parser, a slight error may result in the parser being unable to specifically state what and where the error is. If each of the components, such as the type of document, the family names, the Initials of each of the other names, the title, journal, etc. is placed into a separate element, the system only needs to know the order of these elements and what punctuation is required. The references also become portable between applications. The CytometryML schemas also employ an address.xsd schema, which includes separate elements for building number, street name, etc.

Data is useful; however, structured data is organized knowledge. A scientific paper includes elements, such as Materials and Methods, which can be described by XML schemas and probably could be used in two of the present popular structured formats DITA[3] and Docbook[4], which are both XML data models for authoring. HTML5 is ubiquitous and can also be used for authoring including EPUBs. CSS is an interpreter that works with HTML to provide WYZIWYG (What You See Is What You Get).

2. MATERIALS AND METHODS

Because of reuse of the software designs including the element content and documentation from the Digital Imaging and Communications in Medicine (DICOM)[5,6] standard, Flow Cytometry Standard, FCS3.1[7] and the XML Schema Definition (XSD1.1) language[8,9,10], much of the information and data-types present in the XML schemas and subsequently XML pages were prepared by domain experts. New data-types were created and data-types from other CytometryML schemas[11] were reused. Because existing code in the Digital Imaging and Communications in Medicine (DICOM)[5,6] standard could be included in a FDA Class II device, the safety of the software developed as part of the CytometryML standard should be maximized. A strong effort was made to: maximize readability, facilitate finding the sources of elements by adherence to naming conventions, modularizing the code, minimizing coupling between major schemas, maximizing cohesion of individual schemas, and reuse of existing CytometryML[12,13,14] XML schemas. The schemas and XML pages were prepared and validated with oXygen (http://www.oxygenxml.com/) using the Xerxes validation engine for XSD1.1[8,9,10]. Subsequently, Individual major elements of the MIFlowCyt schema were translated by oXygen into XML pages and filled with reasonable data. The simplest and most powerful method for displaying the elements that constitute the XML page is the use XHTML5[15] and CSS[16]. Unfortunately previously, XHTML5 was not defined in terms of an XSD schema that would validate. However, recently a useful XSD schema that can be validated has become available[17].

Below is Code Fragment 1, which shows the elements of MIFlowCyt that are being included in CytometryML.

Code Fragment 1 MIFlowCyt Overview

```
<complexType name="Miflowcyt_Type" mixed="true">
  <sequence>
     <element name="Experiment_Overview"
        type="exper_overview:Experiment_Overview_Type"/>
     <element name="Specimen_Info"
        type="specimen:Specimen_Info_Type"/>
     <element name="Procedure_Operation"
        type="protocol:Procedure_Operation_Type"/>
     <choice>
        <element name="Flow_Series_and_Instance_Info"
           type="flow:Flow_Series_and_Instance_Info_Type"/>
        <element name="Microscope_Series_and_Instance_Info”
```

```
            type="micro:Microscope_Series_and_Instance_Info_Type"
      </choice>
   </sequence>
</complexType>
```

Code Fragment 1 provides a top-level description of much of the contents of MIFlowCyt. Its validation proves that all of its component schemas are valid and can be used to generate XML pages that are based on MIFlowCyt. Because it is cumbersome to develop and software test very large XML pages, the plan is to sequentially develop the schemas that are required for elements 3,4,5,7,and 8 and their component schemas. The XML software described below pertains to the first member of this ensemble, Experiment_Overview (Element 3). The use of essentially the same names for the elements and the data-types from which they were derived is the result of the use of object oriented technology to describe complex data-structures.

3. Problems and opportunities in implementing MIFlowCyt in CytometryML

The purposes of CytometryML extend those of MIFlowCyt. These purposes include: 1) Permit reproduction of an experiment. 2) Serve as the basis of a structured document, such as a Cytometry article or supplemental information. 3) Provide a means to store structured information on methods, data and results that can be used for scholarly and/or clinical publication, reports, and other modes of communication. 4) Reduce the cost of Open Access and standard publication by developing of do-it-yourself documents and report creation tools.

One of the initial requirements for CytometryML is to maximize the reuse of elements based on the data-types present in other standards, particularly; Flow Cytometry Standard (FCS), the Archival Cytometry Standard[12] (ACS), and DICOM, as well as data types from other standards. There can be name and data-type definition conflicts between type definitions produced by different groups including ISAC and other organizations. For instance the ISAC ACS could be considered a subset of EPUB[18], which is also based upon a ZIP file container.

3.1. Ideal Solution

The ideal solution is to use a combination of XHTML5 and XML Schema Description language (XSD1.1) elements. Since the definition of much of XHTML5 is available in the form of an XSD schema[17], this schema should be able to be merged with any other XSD by the use of a defaultOpenContent element. These elements permit the inclusion of elements from one schema in a web page that was generated by another schema. One of these defaultOpenContent elements was added directly underneath the last import element of both the XHTM5 and Experiment_Overview schemas

Code Fragment 2,

```
1 <defaultOpenContent
      mode="interleave">
2     <any namespace="##other"
          processContents="skip"/>
<!--Corrected to replace lax with skip-->
</defaultOpenContent>
```

Since the use of defaultOpenContent elements permits two separate sets of elements to be oblivious to each other. The XHTML5 elements would be invisible to the MIFlowCyt elements and the MIFlowCyt elements are invisible to the XHTML5 phrasingContent elements (h1, h2, h3, h4, h5, p, etc.)

4. Results

One way to establish that a schema produces the appropriate web pages is to compare these pages with actual results, Therefore this study has the goal to produce a formatted XML page to be consistent with the example published in a supplement of Blimkie et al.[2]

.

1.4. Organization

1.4.1. Name

1.4.1.1. Child & Family Research Institute, Department of Pediatrics, University of British Columbia

1.4.2. Address

1.4.2.1. 938 West 28th Avenue, Vancouver, British Columbia, Canada V5Z 4H4

Figure 1, Copy of a section from Blimkie et al.[2] with the numbers and formatting manually completed. In Blimkie et al. there were two organizations and two addresses. For reasons of space and simplicity the references to the second organization were removed. MIFlowCyt is similar to scientific publications in that it kept the two names together and the two addresses together.

1.4. Organization

1.4.1. Name

1.4.1.1. Child and Family research Institute, Department of Pediatrics, University of British Columbia

1.4.2. Address

1.4.2.1. Department of Department of Pediatrics, 938 28th Avenue, Vancouver, BC, Canada V5Z 4H4
http://pediatrics.med.ubc.ca
938 West 28th Avenue, Vancouver, British Columbia, Canada V5Z 4H4

Figure 2 ,View of part of a web page created with the oXygen program. In order to establish that the data had actually been formatted by the CSS file, colors were included in the h2 and h3 styling elements to increase readability and to establish, as mentioned above, that the data had actually been formatted by the CSS file. These colors were not carried over directly into the word processor, Adobe FrameMaker Version 10. However, the paragraph numbers except for the last were conserved when copied over from the Firefox web page screen. The colors and numbering were retained after printing from the web page.

The second address and URL (not shown) were added, as described below. The URL was added because it is a more common way to transfer scientific information than a mailing address. The last line is a string that reproduces the full title for items that accept the address as a single string.

Code Fragment 3, XML Code with CSS styling codes

```
<h2 class="center">1.4. <address_org:Organization_Name_Fixed>Organization
</address_org:Organization_Name_Fixed></h2>
<address_org:Org_Name>
  <h3>1.4.1. <address_org:Name_Fixed>Name</address_org:Name_Fixed></h3>
    <h4> 1.4.1. <address_org:Org_Name>Child and Family research
       Institute,
       Department of Pediatrics, University of British Columbia
    </address_org:Org_Name></h4>
</address_org:Org_Name>
<address_org:Org_Address>
   <h3>1.4.2. <address_org:Address_Name_Default>Address
```

```
        </address_org:Address_Name_Default></h3>
7           <address_org:Other_Address>
8             <h4>1.4.2.1 <address_org:Department_Name>
9                Department of Pediatrics</address_org:Department_Name>,
10            <address_org:Building_Number>938</address_org:Building_Number>
11            <address_org:Street_Name>28th Avenue</address_org:Street_Name>,
12             <!--address_org:Building_Name>Commented out
            </address_org:Building_Name-->
13             <address_org:City_Name>Vancouver</address_org:City_Name>,
14             <address_org:Canadian_Provence_Code>BC
               </address_org:Canadian_Provence_Code>
15             <address_org:Code>V5Z 4H4</address_org:Code>
16             <address_org:Country_Name>Canada</address_org:Country_Name>
17             <address_org:URI>http://pediatrics.med.ubc.ca
               </address_org:URI></h4>
18             <h4> <address_org:Address_Free_Text>938 West 28th Avenue,
                  Vancouver, British Columbia, Canada V5Z 4H4
               </address_org:Address_Free_Text></h4>
          </address_org:Other_Address>
        </address_org:Org_Address>
     </address_org:Organization>
```

Code Fragment 3 is a small section of the Experiment Overview (exper_overview.xsd) schema. Code Fragment 3 only includes the name and address for the institution located in British Columbia. The Experiment_Overview schema and XHTML5.xsd[17] schema after minor modifications, both validate. The web page generated by oXygen from Experiment_Overview schema also validates. However, the inclusion of the HTML markup elements makes the validation fail for Code Fragment 3. It has validated after the addition of a single h2 element. Simple web pages that include the html markup elements do validate. The cause of the validation problem with the Experiment_Overview web page that includes html elements is being actively investigated. However, a marked up filled Experiment_Overview page will produce a formatted but not validated page. As is shown in element 8 of Code Fragment 3, the contents of the components of the address are concatenated by enclosing most of element 8 which includes the address of the British Columbia institution. Code Fragment 3. Element 8 terminates one line before above the end of the code. Since all of the address elements except Address_Free_Text are between the beginning and end of the <h4> element, they are concatenated. This concatenation of multiple elements results in an easily read, compact display of the components of the element, which provides sufficient space to show the two names and each of their addresses. This capacity to concatenate elements permits a standard HTML CSS file to format XML data. It is much easier to check a formatted web page, such as shown in Figure 2 than the text shown in Code Fragment 3.

5. Conclusions

1. The combination of XML elements, XHTML5 and CSS potentially will be very powerful publication tool, particularly when it is coupled with the EPUB Standard[18].
2. A significant part of MIFlowCyt, has been described in XSD1.1.
3. The feasibility of using MIFlowCyt to provide the combination of an overview and index to a scientific paper or a report has been demonstrated.
4. Searching for specific elements should be much more effective than for key words.
5. The feasibility of a combination of XHTML5 with XML schema elements is strongly suggested.
6. Since multiple XHTML5 pages can be combined into one EPUB[18], a potential open, efficient, searchable scientific format can be developed. This would provide a seamless system, where the data is in XML except for the binary files, such as FCS and images. The XML elements could be reused in reports, forms, papers, books, web pages, etc. as well as, the source of instrument displays etc.
7. Scientific papers could be created with web tools and/or EPUB tools. The use of XHTML formatted papers should permit the publisher and/or the editor to control the CSS that was used to format the paper.

6. ACKNOWLEDGMENTS

This work was sponsored by Newport Instruments internal research funds. One of us, R.C.L. wishes to thank Ryan R. Brinkman, Josef Spidlen and the other members of the ISAC DSTF for many enlightening and pleasant discussions. The members of DICOM Working Groups 26 and 27 for the knowledge that they have provided. We thank the staff of SyncRO Soft SRL, the publisher of oXygen, for their helpful software support. the editing of the manuscript by and suggestions from Liza C. Leif are greatly appreciated.

7. References

1] Lee, J., Spidlen J. and Data Standards Task Force, ISAC Recommendation, "Minimum Information about a Flow Cytometry Experiment MIFlowCyt 1.0. A standard for outlining the minimum information required to report the experimental details of flow cytometry experiments," Eds. Lee, J., Spidlen, J., Scheuermann R., and Brinkman R., <http://www.isac-net.org/media/standards/miflowcyt/latest.pdf>

2]. Blimkie, D., Fortuno, E.S. 3rd., Thommai, F., Xu, L., Fernandes ,E., Crabtree, J., Rein-Weston, A, Jansen, K., Brinkman R.R., Kollmann, T.R., "Identification of B cells through negative gating—an example of the MIFlowCyt standard applied, "Cytometry A. 77(6) 546-551. (2010).

3] OASIS, "Darwin Information Typing Architecture (DITA) TC," <https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=dita> (2014).

4] Walsh N., "DocBook 5: The Definitive Guide," Publisher: O'Reilly Media <http://shop.oreilly.com/product/9780596805012.do> (2010).

5] Pianykh O.S., [Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide] Springer Publishers, Berlin & Heidelberg, (2008).

6] "The DICOM Standard 2014B", Medical Imaging & Technology Alliance - a division of NEMA, <http://medical.nema.org/standard.html> (2014).

7] Spidlen J., et al. and the ISAC, Data Standards Task Force (International Society for Advancement of Cytometry), "Data File Standard for Flow Cytometry Version FCS 3.1 Normative Reference", ISAC Publisher <Isac-net.org/PDFS/90/9090600d-19be-460d-83fc-f8a8b004e0f9.pdf> (Copyright (c) 2008-2009 ISAC).

8], W3C, "W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures, W3C Recommendation 5 April 2012"<http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/> (2012).

9] "W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes, W3C Recommendation 5 April 2012" <http://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/> (2012).

10] Walmsley P., [Definitive XML Schema, Second Edition], ISBN: 0132886723 Publisher Prentice Hall (2012).

11] Leif R.C., Leif S.B., and Leif S.H., "CytometryML, an XML format based on DICOM and FCS for analytical cytology data," Cytometry 54A, 56-65 (2003).

12] Leif, R.C., Spidlen, J., Brinkman, R.R., "A Container for the Advanced Cytometry Standard (ACS)," Proc. SPIE 7182, 71821Q (2009).

13] Leif, R.,C.,"Toward the integration of cytomics and medicine," J. Biophoton. 2, 482–493 (2009).

14] Leif, R.C., "CytometryML, Binary Data Standards," Proc. SPIE 5699, 325-333 (2005).

15] W3C WATWIG "HTML5, A vocabulary and associated APIs for HTML and XHTML, W3C Recommendation 28 October 2014" < http://www.w3.org/TR/html/> (2014).

16] "Cascading Style Sheets (CSS) Snapshot 2010, W3C Working Group Note 12 May 2011,"<http://www.w3.org/TR/css-2010/> (2011).

17] Ishacian, O., "A W3C XML Schema for XHTML5., Corresponds to W3C Working Draft 29 March 2012". Publisher Pixware SARL, <http://grepcode.com/file/repo1.maven.org/maven2/com.googlecode.l10n-maven-plugin/l10n-maven-plugin/1.5/xhtml5.xsd> (2012).

18] International Digital Publishing Forum (IDPF), EPUB Publications, "3.0.1, Recommended Specification 26 June 2014"; Eds. Gylling, M., McCoy,W., and Garrish, M., <http://www.idpf.org/epub/301/spec/epub-publications.html> (2014).