

Cytometry Metadata In Xml And Xhtml5

Poster Board Number: B267

Robert C. Leif^{1*}, Stephanie H. Leif¹ ¹XML_Med, a Division of Newport Instruments, 3345 Hopi Place, San Diego, CA, rleif@rleif.com

The CytometryML XML schemas and XML pages are available at www.cytometryml.org. You can read this presentation there at your leisure. You can also download the schemas and comment, suggest, improve, and/or use them.

The Difference between CytometryML and FCS

Requirements and uses:

- 1) CytometryML has the requirement that it be readable by cytometrists and related professions, who can be unskilled in computer science.
- 2) It is designed to work with off-the-shelf software. Specifically Digital Imaging and Communications in Medicine (DICOM) image servers and Electronic Publications (EPUBs).
- 3) CytometryML is a tool for describing an experiment in sufficient detail, such that an experiment can be repeated and understood. CytometryML software is developed by bioengineers and histochemists and is in large part technical writing.
- 4) FCS has a different set of requirements, because it is a tool to manage and analyze cytometry data, developed by computer algorithm experts and applied mathematicians. This use often produces useful tools to permit scientists and clinicians to store and analyze data.

MIFlowCyt is: The Minimum Information about a Flow Cytometry Experiment

- Presently, MIFlowCyt is a file or a short appendix to a scientific article, such as one published in Cytometry.
- MIFlowCyt provides a concise description of the technology and findings present in the article. Basically, it is a short description of the analysis, methodology and results of the measurement.
- MIFlowCyt need not contain sufficient information to repeat the experiment, such as would be provided by a well written Materials and Method section of a paper and/or supplementary materials.
- The information contained in a MIFlowCyt document can be textural and/or structured.

MIFlowCyt Completeness

- The present use of MIFlowCyt in a journal, such as Cytometry, requires a referee to attest that the MIFlowCyt is complete. This is a subjective measurement. How does one verify that a MIFlowCyt section of a publication is complete?
- Validation of the XML document would provide significant evidence that the required information was present and conformed to MIFlowCyt.
- Preliminary data indicates that it is possible to create an XML schema (template) that includes the elements that make up MIFlowCyt.
- The elements that describe the minimum could be extended to include the information that would permit the measurement to be reproduced.
- Both conceptually and operationally, it is preferable to have the complete description be an extension of a partial description.

- A disagreement between the MIFlowCyt data and that supplied in the body of the text could require a difficult reconciliation.

Adaptability and Flexibility

Since MIFlowCyt is supposed to provide the minimum information for publication, in its present form it is inadequate for other uses when more information would be appropriate, such as a description suitable for repeating the measurement. One means of accomplishing this is to make the schemas for MIFlowCyt flexible and extendable. For instance, when the DSTF does not believe that an element needs to be required but recording of that element will permit repetition or facilitate understanding of a process or a result, XML has a minOccurs attribute which when set to zero permits the absence of an element in an XML page. The maximum number of an element that can be present can be specified with a maxOccurs attribute. Thus, web pages can be created that can be validated against a schema in the absence of an element, only one value or many values of an element.

A fast safe way to find mistakes and thus be able to minimize their effect is to put them in an easily readable form, such as shown in Table 1, which was created with Microsoft Edge. Table 2, which does not have properly formed vertical columns, was created with Google Chrome. The lack of the data being properly organized into vertical columns also occurred with Firefox (not shown). The images of the 3 Tables were captured with Microsoft Paint. Only Table 1, Microsoft Edge, was acceptable. A major purpose of the different colors was to prove that the elements had been processed by CSS.

The column headings are derived in part from the DSTF discussions, particularly with Kim Blenman; but are presently (May 28, 2016) different from what will be used for FCS4.0.

- The first header is "Name". Chemists tend to refer to this as trivial or common name; another alternative is universal name. Name in the CytometryML schema is defined as that name shown for the compound on the label of the container (bottle).
- Most databases employ keys that are used for lookups. The simplest unique descriptor that can be used to find an entity often is a number. However, the American Chemical Society Chemical Abstracts Service (CAS) numbers, although suitable for a key, have the deficiency that they are unavailable for many dyes and do not provide human readable information concerning the individual dyes. The CAS number format is unavailable for mass and other types of tags or labels. An existing practice is that a short string be followed by a meaningful number. In CytometryML, The key in the column labeled Key consists of a short descriptive letter string attached by a hyphen to the excitation maximum. The short descriptive string should be based on the chemistry and/or the chemistry class of the dye. The excitation maximum is a property, which is of use in selecting the light source. An error in data entry of the descriptive letter string, often should be correctable by inspection of the table of dyes. The use of the hyphen is for readability and ease of parsing. The last column heading is the URI (Internet address). This permits the purchase of the dye and checking of its name.

Table 1 is based on an XML schema and displayed using a Cascading Style Sheet. However, the support by browsers for tables created in XML and CSS appears to be poor. As the style sheet and the XML pages were being developed, often the formation of the columns disappeared. It appears that a simple HTML web page consisting of elements that were plain strings and not XML elements together with a CSS sheet should first be constructed to match the desired format and that CSS sheet reused as a starting point with the XML elements.

Three browsers were tested with the combination of XML page and CSS file. Only Microsoft Edge properly displayed the data in the XML page. Table 2, Google Chrome, and Firefox (not shown), which looks like Chrome, showed the Caption as string on multiple lines packed against the left margin and the data not properly formed into vertical columns.

In any event, it must be emphasized that the contents of the CytometryML tables are tentative. We know that all of us inevitably will make mistakes and omissions. A fast safe way to find them and thus be able to minimize their effect is to put them in an easily readable form, such as shown in Table 1. The lack of the data being properly organized into vertical columns also occurred with Firefox (not shown), except the size of the table was smaller. The images of the 3 Tables were captured with Microsoft Paint. Only Table 1, Microsoft Edge, was acceptable. A major purpose of the different colors was to prove that the elements had been processed by CSS.

Table 1, Microsoft Edge

Dye List 14 May 2016.xml

Key	Common Name	Tag	Excitation Max	Emission Max	URI	CAS NUM
CAS_Num	6-carboxyfluorescein	6-FAM	492	518	http://www.cheap.com/	3301-79-9
Common_Name	Joe	6-Joe	520	550	http://www.cheap.com/	82855-40-1
Common_Name	EuQuantum Dye	EuMac	365	619	http://www.NewportInstruments.com/	Unknown

Dye List 14 May 2016.xml						
Table 2, Google Chrome						
Key	Common Name	Tag	Excitation Max	Emission Max	URI	CAS NUM
CAS_Num	6-carboxyfluorescein	6-FAM	492	518	http://www.cheap.com/	3301-79-9
Common_Name	Joe	6-Joe	520	550	http://www.cheap.com/	82855-40-1
Common_Name	EuQuantum Dye	EuMac	365	619	http://www.NewportInstruments.com/	Unknown

The number of words and lines are compared in Table 3. Clearly the Table is more compact.

Table 3, Size Comparison

	Words	Lines
Table	35	6
XML Table Elements	105	80
Element/Tables	3	13

Present Status of the CytometryML Implementation of MIFlowCyt

1. The software has been tested to produce Table 1, Microsoft Edge, which includes the elements necessary to describe a fluorescent dye provides sufficient flexibility to permit the selection of a Identification Number from Vendor, such as CAS.
2. The standard XML pages validated as expected; however, when the html table formatting elements were added The page would not validate. When these pages were tested with the Namespace-based Validation Dispatching Language, NVDL², file present, they validated.
3. A significant part of MIFlowCyt, has been coded in XML Schema Definition Language, XSD1.1. MIFlowCyt.xml is relatively large with all coded elements present, validates and contains 882 lines of

code. This would be appropriate for a structured document, where the materials and methods was primarily composed of the MIFlowCyt elements. When only the required elements were included, it validates with 432 lines of code, which would be suitable for the present synopsis.

4. MIFlowCyt.xml is based upon a tree of approximately 77 schemas. The next step in this project is to finish the Sample Treatment Description Type.

Conclusions:

1. Cytometry data, because it is XML, can be stored in a database, such as one used by DICOM for image storage. Many high-level elements will be part of a database of which subsets are used for the cytometry documents including reports.
2. The feasibility of using MIFlowCyt to provide the combination of an overview and index to a scientific paper or a report has been strongly suggested.
3. The feasibility of a combination of XHTML5 with XML schema elements has been demonstrated. The use of CSS with these XML web pages is sufficient to be able to monitor the design of a schema from the XML produced from the schema. Unfortunately, the reliability of the formatting of these XML web pages is insufficient to produce publications.

References:

- [1] Robert C. Leif, Stephanie H. Leif, "Cytometry metadata in XML", Proc. SPIE 9711, Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues IX, 971118 (April 6, 2016); doi: 10.1117/12.2218158
- [2] NVDL, ISO/IEC 19757-4 NVDL (Namespace-based Validation Dispatching Language) information is available at", (<http://www.nvdl.org/>)