

# DEVELOPMENT OF AN INTERSOCIETY LABORATORY FLOW & IMAGING DATA EXCHANGE STANDARD (P139)

Robert C. Leif

Newport Instruments, San Diego, CA. rleif@rleif.com, www.newportinstruments.com

## ABSTRACT

The development of a single standard for cytometry informatics has been greatly confounded by the differences in the requirements, needs, and past histories of the societies and groups (See Appendix) that are involved in their creation. Interoperability between the various standards/specifications can be maximized by

- 1) Combining the description of flow cytometry and digital microscopy and
- 2) Employing common semantics for the description of data-types.

This poster shows in the form of the XML schemas, which are referred to as Cytometry Markup Language, CytometryML, that

- 1) A significant part of a common data standard for digital microscopy and flow cytometry has been created.
- 2) It was possible to reuse the semantics of the data types are based on the Digital Imaging and Communications in Medicine (DICOM), ECMA and other relevant standards.

This standard will allow research and pathology images, list mode data, and accompanying annotations to be exchanged, stored in databases, and facilitate the creation of software by third parties.

## Problem

- "The nice thing about standards is that there are so many of them from which to choose" –anon
- "Standards have become so popular that every society wants to make one for themselves" -anon

## Partial Solution

- Standards paucity the practice of reusing datatypes from other standards works. Most of the data-types and their documentation were reused from existing standards, principally DICOM and ECMA. This reuse minimized the design effort and maximized the reliability of the schemas due to the previous successful use of many of the data-types in implementations of DICOM.
- The creation of new data-types was facilitated by extending and/or enhancing already existing data-types.
- In the future, the CytometryML schema will be enhanced by reuse of data-types and members of enumerations from DICOM Working Group 26, LDIP, Flowcyt, and OME.
- Although the different groups employ different representations (syntax) for their data, the definitions (semantics) of the data-types should, as much as possible, be common to all of the standards.

## METHODS

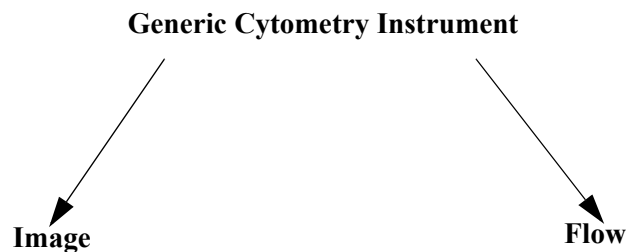
- The CytometryML schemas including the code examples below were developed using the XML Schema language (XSD). They primarily consisted of DICOM data-types with XSD documentation elements that included references to the descriptions of the data-types in the DICOM standard and data-types that could be part of an extension of DICOM data-types.

### Brief Description of XML Schema Syntax

The schema examples below are color coded. XML Schema is a nested language that begins statements which the less than character < to begin a construct and ends constructs that are not nested with />. Schemas include both simpleTypes and ComplexTypes. simpleTypes are data structures that can only be based on a single elementary

type like a string or a number. Attributes are a shorthand description of an object that can be only based on a simpleType. An element is a general purpose description of all other types of objects. ComplexTypes are a data structures that include multiple entities or one or more attributes.

## Solution First Step, A Common Set of XML Schemas (XSD) for Both Flow and Imaging?



First step, create simple (helper) data-types

```
1.<simpleType name="Platform_Type">
2. <restriction base="token">
3.   <enumeration value="Upright"/>
4.   <enumeration value="Inverted"/>
5.   <enumeration value="Plainer"/>
6. </restriction>
7.</simpleType>
```

Statement numbers have been added at the far left and will be referred to in parenthesis. Statement (1), employs the less than character < to begin a construct. The first element is a **simpleType**. Since (1) begins a set of nested statements it ends with the >. The word **name** is an **attribute** that has a value "Platform\_Type". Attributes are always based upon simple\_Types. The value of an attribute is set in quotation marks and follows the equals sign. This is a user defined **simpleType**.

The **Platform\_Type** is derived by restricting (2) a previously defined type of string a **token** (a string that does not include leading or trailing spaces and several nonprinting characters).

As shown on (3), statements that are not nested end with />. Statements (3, 4, and 5) provide the values for this enumerated type. In XML enumerated types are restrictions of the type string. The list of enumeration values ends on (5).

The **restriction**, which is a nested statement, ends on (7) with a />, as does the **simpleType** on (7).

The **Carrier\_Type** and **Viewing\_Type** employ similar syntax

```
1.<simpleType name="Carrier_Type">
2. <restriction base="token">
3.   <enumeration value="Solid"/>
4.   <enumeration value="Fluid"/>
5. </restriction>
6.</simpleType>
```

1. <simpleType name="Viewing\_Type">
2. <restriction base="token">
3. <enumeration value="Mono"/>
4. <enumeration value="Sterio"/>
5. <enumeration value="Other"/>
6. </restriction>
7. </simpleType>

Create a complex datatype (**complexType**) to model a generic instrument

1. <complexType name="Instrument\_Type">
2. <sequence>
3. <element name="Item\_General\_Info" type="item:Item\_General\_Info\_Type"/>
4. <element name="Objective" type="optics:Optic\_Type" maxOccurs="7"/>
5. <element name="Condenser" type="optics:Optic\_Type" minOccurs="0"/>
6. </sequence>
7. <attribute name="Platform" type="instr:Platform\_Type"/>
8. <attribute name="Carrier" type="instr:Carrier\_Type"/>
9. <attribute name="Viewing" type="instr:Viewing\_Type"/>
10. <attribute name="Sorts" type="boolean"/>
11. </complexType>

**Explanation of schema syntax:** The first element (1) is a complexType, because it includes a sequence (2). A sequence contains one or more elements (3,4,5) each with its own type declaration. These elements can also be complexTypes. The type declarations in (3-9) include prefixes, such as **item:** and **optics:**. These prefixes provide the location of the schemas which contained them. Since the primitive type boolean (10) is defined in the schema standard, it does not require a prefix. (4 and 5) specify the cardinality, **maxOccurs** and **minOccurs**, of their elements. (7-10) describe attributes, which are based on simpleTypes and besides standing alone are used to describe and specify types. The terms **name**, **type**, **maxOccurs**, and **minOccurs** are all attributes.

**Transform the generic instrument into a microscope and show differences for a flow cytometer.** The numbers followed by M are statements describing a microscope, those followed by an F are a flow cytometer, and those without a letter are common to both types of instruments.

- 1.M. <complexType name="Microscope\_Type">
- 1.F. <complexType name="Flow\_Cytometer\_Type">
2. <complexContent>
3. <restriction base="instr:Instrument\_Type">
4. <sequence>
5. <element name="Item\_General\_Info"
6. <type="item:Item\_General\_Info\_Type"/>
- 7.M. <element name="Objective" type="optics:Optic\_Type" maxOccurs="7"/>
- 7.F. <element name="Objective" type="optics:Optic\_Type" maxOccurs="1"/>
- 8.M. <element name="Condenser" type="optics:Optic\_Type" minOccurs="0"
- <maxOccurs="1"/>
- 8.F. <element name="Condenser" type="optics:Optic\_Type" minOccurs="0"/>
9. </sequence>
10. <attribute name="Platform" type="instr:Platform\_Type"/>

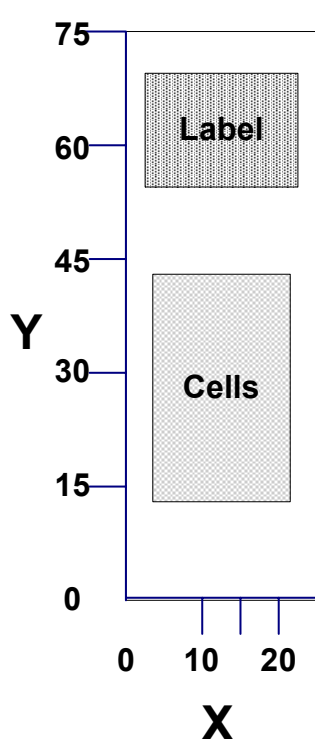
11.M. `<attribute name="Carrier" type="instr:Carrier_Type" fixed="Solid"/>`  
 11.F. `<attribute name="Carrier" type="instr:Carrier_Type" fixed="Fluid"/>`  
 12. `<attribute name="Viewing" type="instr:Viewing_Type"/>`  
 13. `<attribute name="Sorts" type="boolean"/>`  
 14. `</restriction>`  
 15. `</complexType>`  
 16. `</complexType>`

The differences are:

- (1.M and 1.F) where the names of the complexType differ;
- (6.M and 6.F) where a microscope could have up to (**maxOccurs**) 7 objectives and a flow cytometer can have only 1;
- (7.M and 7.F) where a fluorescent microscope can have (**minOccurs**) 0 or 1 condensers and a flow cytometer requires a condenser for each light source (crossed cylindrical lenses);
- (10.M and 10.F) where the microscope has a **fixed** value, **Solid**, for the **Carrier\_Type** and the flow cytometer uses a **Fluid** carrier.

### Image Datatypes

A significant part of the infrastructure of the datatypes used to describe images is based on previously defined CytometryML schemas ([www.newportinstruments.com](http://www.newportinstruments.com)) that described list mode metadata. Most of the rest was based on the DICOM standard (<http://medical.nema.org/dicom/2004/>).



#### Slide

The Offset\_Type, which is based on DICOM, is a vector consisting of the X, Y, and Z offsets. The X and Y offsets are the distance in millimeters from the origin of the slide. As is shown in Figure 1, the origin (0,0) is on a corner of the end of the side that is opposite from the label. The Z offset is the distance in microns from the surface of the slide closest to the objective.

#### Pixel

Since the DICOM pixel type is limited to monochrome or the three primary colors, it is inadequate for cytometry. A simple solution is to incorporate into the Pixel\_Type the **Channel\_Type** (parameter) from the previous List mode schemas. List-mode data produced by either digital microscopy or a flow cytometry being very similar.

- Flow often uses the time after the start of a run as a parameter and digital microscopy uses the X, Y, and often Z coordinates.
- The rest of the parameters for both modalities are overlapping subsets of the class of cytometry parameters, Channel\_Type, which has been defined in the channels schema. The schema code for the complexType Pixel\_Type is shown below.

Figure 1. Locations of the DICOM slide coordinates.

```

1.<complexType name="Pixel_Type">
2.  <sequence>
3.    <element name="Samples_Per_Pixel" type="pixel:Samples_Per_Pixel_Type"/>
4.    <element name="Photometric_Interpretation
      "type=" pixel:Photometric_Interpretation_Type"/>
5.    <element name="Pixel_Aspect_Ratio" type="pixel:Pixel_Aspect_Ratio_Type"/>
6.    <element name="Channels" type="channels:Channel_Type" minOccurs="1"
      maxOccurs="20"/>
7.    <element name="Location" type="slide:Offset_Type"/>
8.  </sequence>
9.</complexType>

```

- The Samples\_Per\_Pixel (3) is the number of parameters measured.
- Samples\_Per\_Pixel\_Type was derived from the Num\_Waveform\_Channels\_Type in the multiplex\_groups schema, which is used to describe list-mode data.
- The Photometric\_Interpretation (4) is an enumeration type that includes: Monochrome, RGB, others specific to DICOM, and Multidimensional, which has been added.
- The Pixel\_Aspect\_Ratio (5) specifies the ratio of the vertical size (Y) and horizontal (X) size.
- The Channels (6) describes each of the channels (parameters); it was imported from the channels schema. Each channel is itself a sequence that describes a complex data structure that includes: a description of the analyte binding species, which includes a description of the analyte; the excitation source including its filter; the detector and its optics; the amplifier; and the data-type of the parameter. At least one and up to 20 parameters can be included.
- The Location (7) specifies a pixel's X and Y slide coordinates and, if it is included, the Z coordinate.

### Image

The image schema provides the metadata for binary image files.

```

1.<complexType name="Image_Type">
2.  <sequence>
3.    <element name="Columns" type="image:Columns_Type"/>
4.    <element name="Rows" type="image:Rows_Type"/>
5.    <element name="Planar_Configuration" type="image:Planar_Configuration_Type"/>
6.    <element name="Pixels" type="pixel:Pixels_Type"/>
7.    <element name="General_Image" type="image:General_Image_Type"/>
8.    <element name="Acquisition_Date_Time" type="time:Acquisition_Date_Time_Type"/>
9.    <element name="Compression" type="image:Compression_Type"/>
10.  <choice>
11.    <element name="Microscope" type="instr:Microscope_Type"/>
12.    <element name="Flow_Cytometer" type="instr:Flow_Cytometer_Type"/>
13.  </choice>
14.  <element name="File_Location" type="anyURI"/>
15.</sequence>
16.<attribute name="Endian" type="dicom:Endian_Type" default="Little_Endian"/>

```

## 17.</complexType>

- The Columns and Rows elements (3 & 4) specify respectively the number of pixels in the horizontal and vertical directions.
- The Planar\_Configuration element (5) specifies whether the image will be by color-by-plane where the image is a stack of monochrome images or color-by-pixel where there is one image with each pixel being a vector of multiple channels (parameters).
- The Pixels element (6) is from the pixel schema. It is identical to the Pixel\_Type described above except that its sequence lacks the Location element, which is only relevant to an individual pixel.
- The General\_Image element (7) includes whether the image is original or has been derived by some process and whether the image is the left or right member of a stereo pair or is monocular.
- The Acquisition\_Date\_Time element (8) is the date and time that the image was acquired or was derived.
- The Compression element (9) can either be lossy or lossless. For both, the image file format type JPEG, JPEG 2000, DICOM, TIFF, etc. is given. For lossy compression, the compression method and ratio are also given.
- Since the instrument, as described above, can be either a Microscope or a Flow\_Cytometer element, the instrument is represented as an XML Schema language Choice element (10).
- Since the File\_Location element location (14) is a Uniform Resource Identifier Reference, URI, it can either be stored on the web, a workstation disc, or on a local server. This permits the XML metadata and the binary image to be stored as separate files and perhaps together as a ZIP or similar file.
- The Endian attribute (16) is from the DICOM schema; In agreement with DICOM, the default is little endian.

## Conclusions

- Flow cytometers and digital microscopes are sufficiently similar that they can be derived from a common ancestor, a generic cytometer.
- The data content of a pixel has been derived from the **Channel\_Type**, which was developed as part of the Channels schema, which describes the parameters acquired by flow cytometers, which is stored in list-mode files.
- Since the DICOM Waveform including its **Channel\_Type have previously been used to model the metadata that describes a list-mode file**, DICOM datatypes can be used for Cytometry.
- Translation into XML schemas of DICOM types and reuse of their descriptions minimized the design effort and will maximize reliability.

## ACKNOWLEDGMENTS

I wish to thank the following individuals for providing information on and explanations of their standardization efforts: Ulysses Balis, Bruce Beckwith, Jules Berman, Ryan Brinkman, Bruce Friedman, and Ilya Goldberg. Newport Instruments internal development funds have supported this project.

## Appendix

### Abbreviations & URLs

Name	Abbreviation	URL
Cytometry Markup Language <sup>1,2,3,4</sup>	CytometryML	<a href="http://www.newportinstruments.com">http://www.newportinstruments.com</a>
Digital Imaging & Communications in Medicine <sup>5,6</sup>	DICOM	<a href="http://medical.nema.org/dicom/2004.html">http://medical.nema.org/dicom/2004.html</a>
Flow Cytometry Standard <sup>7,8</sup>	FCS	<a href="http://www.isac-net.org">http://www.isac-net.org</a>
Flowcyt	Flowcyt	<a href="http://www.flowcyt.org">http://www.flowcyt.org</a>
Health Level 7	HL7® Version 3 Standard	<a href="http://www.hl7.org/v3ballot/html/welcome/environment/index.htm">http://www.hl7.org/v3ballot/html/welcome/environment/index.htm</a>
Laboratory Digital Imaging Project Data Exchange Specification <sup>9</sup>	LDIP	<a href="http://www.ldip.org">http://www.ldip.org</a>
Open Microscopy Environment <sup>10</sup>	OME	<a href="http://www.openmicroscopy.org">http://www.openmicroscopy.org</a>

### Standards

Organization	Name	Status	Format
API	LDIP <sup>9</sup>	Developing	RDF
NEMA, CAP	DICOM WG 26	Starting	DICOM
ISAC	FCS 3.0 <sup>7,8</sup>	Existing	nonstandard
ISAC Members	Flowcyt <sup>11</sup> Gating (Classifying)	In Development	XSD
ISAC Members	Flowcyt <sup>11</sup> Ontology	In Development	OWL (RDF)
HL7	HL7 Ver. 3	Developing	XSD
Newport Instruments	CytometryML	Existing & expanding	XSD
OME	OME	Existing & expanding	XSD & Others

## Organizations

Entity	Abbreviation
International Society for Analytical Cytology	ISAC
College of American Pathologists	CAP
National Electrical Manufacturers Association	NEMA
Association for Pathology Informatics	API
Health Level 7	HL7

## REFERENCES

1. R. C. Leif, S. B. Leif, and S. H. Leif, "CytometryML, An XML Format based on DICOM for Analytical Cytology Data", *Cytometry* 54A pp. 56-65, 2003.
2. R. C. Leif and S. B. Leif, "A DICOM compatible format for analytical cytology data, that can be expressed in XML." In: Farkas DL, Leif RC, editors. Optical diagnostics of living cells IV. Vol. 4260. Bellingham, WA: SPIE Proceedings; 201. pp. 238-248, 2001.
3. R.C. Leif, S.H. Leif, S.B. Leif, "CytometryML, a markup language for analytical cytology", in *Manipulation and Analysis of Biomolecules, Cells and Tissues*, D. V. Nicolau, J. Enderlein, and R. C. Leif, Editors, SPIE Proceedings Vol. 4962, pp. 288-297, 2003.
4. R. C. Leif, "CytometryML, Binary Data Standards", in *Manipulation and Analysis of Biomolecules, Cells, and Tissues II*, D. V. Nicolau, J. Enderlein, R. C. Leif, and D. Farkas, Editors, SPIE Proceeding Vol. 5699 pp. 325-333 (2005).
5. DICOM, Digital Imaging and Communications in Medicine <http://medical.nema.org/dicom/2004.html>
6. DICOM Part 3 from [http://medical.nema.org/dicom/2004/04\\_03PU3.PDF](http://medical.nema.org/dicom/2004/04_03PU3.PDF).
7. L. C. Seamer, C. B. Bagwell, L. Barden, D. Redelman, G. C. Salzman, J. C. Wood, R. F. Murphy, "Proposed new data file standard for flow cytometry", version FCS 3.0. *Cytometry* **28**, pp. 118-122 1997.
8. FCS, Flow Cytometry Standard <http://www.isac-net.org/> Then search for FCS.
9. LDIP, Laboratory Digital Imaging Project (LDIP) Data Exchange Specification, <http://www.ldip.org>
10. OME, Open Microscopy Environment, <http://www.openmicroscopy.org>
11. Flowcyt project (<http://www.flowcyt.org/>)
12. Health Level 7 (<http://www.hl7.org/>)
13. RDF/XML Syntax Specification (Revised), W3C Recommendation 10 February 2004, Latest version: <http://www.w3.org/TR/rdf-syntax-grammar/>
14. OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, <http://www.w3.org/TR/owl-features/>
15. XML Schema Part 1: Structures Second Edition, W3C Recommendation 28 October 2004, Latest version: <http://www.w3.org/TR/xmlschema-1/>
16. XML Schema Part 2: Datatypes Second Edition, W3C Recommendation 28 October 2004, Latest version: <http://www.w3.org/TR/xmlschema-2/>
17. Standard ECMA-335, Common Language Infrastructure (CLI), 3rd edition (June 2005), <http://www.ecma-international.org/publications/standards/Ecma-335.htm>