

STANDARDS CONTINUUM

Robert C. Leif

XML_Med, a Division of Newport Instruments, San Diego, CA

E-mail: rleif@rleif.com; phone 1 619 582-0437

www.newportinstruments.com

This presentation can be found at <http://www.newportinstruments.com/cytometryml/cytometryml.htm>

Abstract

This presentation will provide evidence that: 1) A single standard is sufficient for both flow cytometry and digital microscopy. 2) XML schemas should be used as the primary modality for the description of data-types. However, these data-types can and should be used with RDF triplets 3) DICOM can serve as the source for the design and documentation of data-types for XML schema and other implementations. And 4) ISAC should limit its design effort to subjects where its membership has significant domain knowledge and should collaborate with other groups, such as DICOM and HL7 for clinical informatics designs and, if possible, schemas.

Since no single group or organization has sufficient knowledge or resources to produce a single standard that describes a cytometry experiment or clinical test in its entirety, each group or organization should produce works in their own area of expertise. Ideally, all of the standards should be specified employing a common modality; the same inheritance hierarchy; and common data-type definitions should be employed. The previous use of a data-type by others and its definition should be a major consideration in the selection of data-types. In the context of cytometry, this means that the Digital Imaging and Communications in Medicine (DICOM) and Health Level 7 (HL7) standards provide the clinical information infrastructure; and ISAC provides the cytometry instrumentation, reagent, and analysis details.

Although Network Common Data Form (netCDF) files include both binary & metadata and thus conflict with the ISAC requirements, the advantages of the ubiquity of netCDF and the promised creation of a program that splits a netCDF file into binary and metadata, as well as translates the minimal metadata into XML are sufficient to mitigate this deviation.

A Single Standard for Digital Imaging and Flow Analysis and Other

The Requirements for a data file standard format to describe cytometry and related Analytical Cytometry Standard (ACS) data include:

5 Facilitate support for other analytical cytology data

5.1 Description

Use a common (shared) mechanism/approach/methodology/terminology to store/analyze/transport data and metadata from flow cytometry as well as image and other analytical cytology technologies. (This does not imply that an initial implementation is required to support all types of cytometry data. An initial implementation could be directed to only one modality, such as flow cytometry).

5.2 Rationale

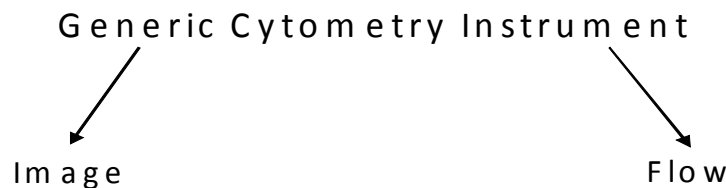
Flow cytometry and digital microscopy represent complimentary technologies used in analytical cytology. Any kind of shared approach has the potential to reduce development costs and facilitate data integration, which is to the benefit of both, developers and end-users.

Advantages of One Standard

1. Since many datatypes can be used for both modalities, the amount of effort it takes to develop, maintain, understand, and extend a combined standard will be considerably less than for two separate standards.
2. The development risk is minimal, since CytometryML (<http://www.newportinstruments.com/cytometryml/cytometryml.htm>), which includes both flow and image cytometry, can serve as a proof of concept.
3. A single standard can be employed for instruments that produce both image and list-mode data.
4. This combination of image and list-mode data will be stored together.
5. Laboratories that use both flow cytometers and digital microscopes will be able to analyze their data using the same tools, which will reduce software development and user education costs.
6. Both flow cytometers and digital microscopes can produce each other's type of data. The Amnis[®] ImageStream (<http://www.amnis.com/>), which is a flow cytometer, produces images and the CompuCyte iColor[™] Fluoro-Chromatic Imaging Cytometer (<http://www.compucyte.com/>), which is a laser scanning microscope, produces Flow Cytometry Standard list-mode. Flow cytometry software is often employed to analyze list mode data obtained with other digital microscopes.

Example

- The Microscope and Flow Cytometer datatypes in CytometryML were both derived from a generic cytometer datatype (Instrument_Type).



Instrument Description

```
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"
maxOccurs="10"/>
6    <element name="Platform" type="instr:Platform_Type"/>
7    <choice>
8      <element name="Stage" type="stage:Stage_Type"/>
9      <element name="Fluidics" type="fluid:Fluidics_Type"/>
    </choice>
10   <element name="Viewing" type="instr:Viewing_Type"/>
11   <element name="Sorts" type="boolean"/>
12   <element name="Comments" type="dicom:Bd_1024_Type" minOccurs="0"/>
</sequence>
```

```

13 <attribute name="Instr_Name" type="dicom:Long_String_Type" use="optional"/>
14   <attribute ref="optics:Objective_List"/>
    </complexType>

```

This example has been taken from (Robert C. Leif, "CytometryML, a data standard, which has been designed to interface with other standards", SPIE Proceedings Vol. 6441, pp. 64410P 1-11, 2007.)

Combined Microscope & Flow Cytometer Descriptions

In the complexType descriptions below, F, M, and I are abbreviations respectively for Flow, Microscope, and Instrument, which is common for both. Elements M1 and F1 each start a separate complexType definition. The elements are common to both definitions.

```

M1. <complexType
      name="Microscope_Type">
F1. <complexType name="Flow_Cytometer_Type">
I2.   <complexContent>
I3.     <restriction
          base="instr:Instrument_Type">
I4.       <sequence>
I5.         <element name="Item_General_Info" type="
"item:Item_General_Info_Type"
          />
M6.         <element name="Objective" type="optics:Optic_Type" maxOc-
curs="7"/>
F6.         <element name="Objective" type="optics:Optic_Type" maxOc-
curs="1"/>
M7.         <element name="Condenser" type="optics:Optic_Type" minOc-
curs="0"
          maxOccurs="1"/>
F7.         <element name="Condenser" type="optics:Optic_Type" minOc-
curs="0"
          maxOccurs="10"/>
I8.         <element name="Platform" type="instr:Platform_Type"/>
M9.         <element name="Stage" type="stage:Stage_Type" minOccurs="1"
          maxOccurs="1"/>
F9.         <element name="Fluidics" type="fluid:Fluidics_Type" minOc-
curs="1"
          maxOccurs="1"/>
I10.        <element name="Viewing" type="instr:Viewing_Type"/>
I11.        <element name="Sorts" type="boolean"/>

```

I12. <element name="Comments" type="dicom:Bd_1024_Type" minOccurs="0"/>
I13. </sequence>
M14. <attribute ref="optics:Objective_List"/>
I15. </restriction>
I16. </complexContent> </complexType>

The <complexContent> element on line (I2) indicates that the elements in the type definition are found in complexTypes. The base type (Instrument_Type) of the restriction element in line (I3) is the parent of this complexType and the new Microscope_Type and Flow_Type are created by narrowing the scope of the parent Instrument_Type type.

Since both instruments can be epi-illuminated for fluorescence, the minimum number of condensers is zero

- The 3 differences between the datatypes are shown in Table 1 below.

Table 1, Differences between the Microscope and Flow Cytometer ComplexTypes

Name	Microscope Type	Flow Type
Objective MaxOccurs	7	1
Condenser MaxOccurs	1	10
Carrier	Solid	Fluid

Since XML schema design language, XSDL, permits limiting values of types to specific ranges, reasonable maxima have been included. In the future, if warranted, these can be easily changed. The inclusion of assertions describing maxima and minima helps prevent data corruption, facilitates the detection of errors, and simplifies the conversion of dictation into text.

DICOM in XML

As described by M. A. Musen & D. L. Rubin in the DICOM Ontology (DO) Project Description

http://bmir.stanford.edu/projects/view.php/dicom_ontology_do_project

“DICOM (Digital Imaging and Communications in Medicine) is the global standard for medical image information; a position it has held for over 20 years. It is pervasive throughout the medical imaging community, and nearly every medical imaging device supports some aspect of the standard. DICOM models the image acquisition process and information objects related to imaging, and it specifies how the image data, the metadata, and related objects are represented in a binary format. For example, DICOM models patients as both clinical and clinical trial subjects, imaging studies that consist of series of images as well as all of the technical parameters of

imaging modalities. Despite its size and complexity, DICOM lacks a Reference Information Model of the imaging domain. A reference information model is a formal description of a domain that enables users to share consistent meaning and establish semantic interoperability beyond a local context.”

“There is a pressing need for an information model of imaging based on DICOM to enable the community to create intelligent imaging-based applications that are interoperable. We are developing the DICOM Ontology (DO), an ontology that will be a single common reference information model for the imaging domain. The DO will be analogous to the Gene Ontology (GO) and serve a similar role in radiology that GO serves in biology. The DO will unify and make explicit all the key entities and relations in DICOM in a human-usable and machine-processable format. The DO will ultimately become a reference ontology—one that comprehensively represents knowledge about the medical imaging domain independent from specific objectives or applications, guided by a theory of the imaging domain and by robust ontology design principles that encourages reuse.”

- This is an excellent idea. However in the interim, it has been possible to create schemas based on DICOM.
- Since these schemas include the DICOM descriptions of the datatypes, they can serve as informal, partial ontologies.

Josef Spidlen and Ryan Brinkman in a recent preprint (FlowRDF: A Demonstration of Describing Flow Cytometry Metadata Using the Resource Description Framework) have demonstrated “how to how to create and reference FCM specific metadata elements and how to reuse standardized RDF metadata elements (defined by resource description communities, e.g., the Dublin Core Metadata Initiative).”

The same metadata can also be described using the combination of XML schemas and DICOM data-types. This approach includes the capacity of using the combination of RDF with DICOM data-types. Analytical Cytometry Standard, ACS, has been substituted for FCS.

```
1<complexType name="ACS_Meta_Type">
2  <sequence>
3    <element name="List_Mode_Location" type="anyURI"/>
4    <element name="Gating_File_Location" type="anyURI"/>
5    <element name="Acquisition_Context_Sequence"
6      type="wave:Acquisition_Context_Sequence_Type"/>
7    <element name="Series_Description" type="dicom:Long_String_Type"/>
8    <element name="Operator_Name" type="pn:Person_Name_Long_Type"/>
9    <element name="Responsible_Organization"
10     type="dicom:Long_String_Type"/>
11   <element name="Responsible_Organization_Address"
12     type="address:Address_Type"/>
13   <element name="Device_Description"
14     type="item:Device_Description_Type" minOccurs="0"/>
15   <element name="Modality" type="instr:Modality_Type"/>
16   <element name="Procedure_Code" type="series:Procedure_Code_Type"/>
17   <choice>
18     <element name="Blood_Specimen" type="spm:Blood_Specimen_Type"/>
```

```

15     <element name="Specimen" type="spm:Specimen_Type"/>
16 </choice>
17 <element name="Abstract" type="dicom:Short_Text_Type"/>
18 <element name="Physician_of_Record"
    type="pn:Person_Name_Long_Type"/>
19 <element name="Acquisition_Date_Time"
    type="time:Acquisition_Date_Time_Type"/>
20 <element name="Snomed" type="snomed:Snomed_Type"/>
21 </sequence>
22</complexType>

```

The order of the elements in the complexType follows the order of the presentation in the Spidlen and Brinkman preprint. The DICOM based terminology is significantly more specific than that provided by the Dublin core.

XMLSpy was used to generate an XML page, which because of the use of previously created data-types from the CytometryML schemas was too large to include in this poster. The large size was also the result of using data-types that contained significantly more information than those in the Spidlen and Brinkman preprint.

XML schema and RDF and their combination will have their uses. The XML page generated from a schema is suitable for the output of a clinical instrument, where the requirements to test and to maximize reliability outweigh the benefit of the flexibility of RDF. Conversely, since researchers need flexibility, they might use RDF in a stand-alone mode or, more probably, to augment material produced by a schema based XML page. Since J. Spidlen (personal communication) has demonstrated that the simpleTypes, such as those present in the CytometryML schemas can be reused in RDF triplets, DICOM in RDF is feasible.

Requirements Conflict

- Req. 8. Each type of information shall only be stored in one file format.
- Req. 28. Reuse existing standards

Network Common Data Form (netCDF)

- netCDF includes both binary & metadata, which is an obvious contradiction to Req. 8.
- netCDF is supported by many scientific organizations (Req. 28):

Unidata (www.unidata.ucar.edu) funded primarily by the National Science Foundation, is 1 of 8 programs in the University Corporation for Atmospheric Research (UCAR) Office of Programs (UOP).

Diverse community of over **160** institutions vested in the common goal of sharing data, and tools to access and visualize that data.

For 20 years Unidata has been providing data, tools

Provides a rich set of services and tools

- Researchers want to reuse netCDF

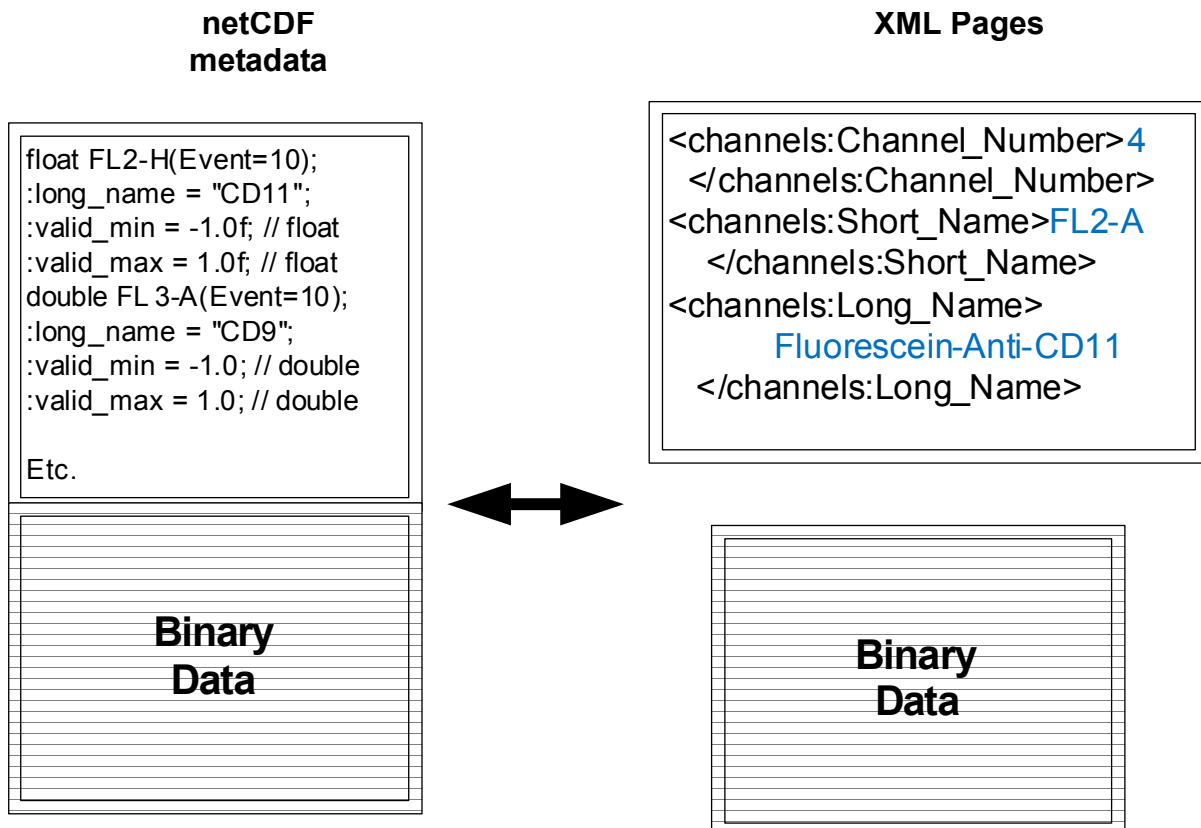
The following is from R. C. Leif, J. Spidlen, R. R. Brinkman, "Cytometry standards continuum", SPIE Proceedings Vol. 6859, 68590Q-1-8 (2008).

netCDF Compound File

<pre>float FL2-H(Event=10); :long_name = "CD11"; :valid_min = -1.0f; // float :valid_max = 1.0f; // float double FL3-A(Event=10); :long_name = "CD9"; :valid_min = -1.0; // double :valid_max = 1.0; // double Etc.</pre>	<table border="1"><thead><tr><th>Rec.#</th><th>FSC-H</th><th>SSC-H</th><th>FL1-A</th><th>FL2-A</th><th>FL3-A</th></tr></thead><tbody><tr><td>1</td><td>990</td><td>111</td><td>505</td><td>139</td><td>638</td></tr><tr><td>2</td><td>889</td><td>30</td><td>843</td><td>161</td><td>20</td></tr><tr><td>3</td><td>721</td><td>400</td><td>156</td><td>962</td><td>992</td></tr><tr><td>4</td><td>250</td><td>981</td><td>158</td><td>92</td><td>670</td></tr><tr><td>5</td><td>970</td><td>166</td><td>877</td><td>208</td><td>359</td></tr><tr><td>6</td><td>737</td><td>553</td><td>504</td><td>390</td><td>752</td></tr><tr><td>7</td><td>522</td><td>463</td><td>866</td><td>535</td><td>316</td></tr><tr><td>8</td><td>438</td><td>521</td><td>894</td><td>330</td><td>515</td></tr><tr><td>9</td><td>46</td><td>138</td><td>326</td><td>727</td><td>598</td></tr><tr><td>10</td><td>780</td><td>430</td><td>648</td><td>388</td><td>261</td></tr></tbody></table>	Rec.#	FSC-H	SSC-H	FL1-A	FL2-A	FL3-A	1	990	111	505	139	638	2	889	30	843	161	20	3	721	400	156	962	992	4	250	981	158	92	670	5	970	166	877	208	359	6	737	553	504	390	752	7	522	463	866	535	316	8	438	521	894	330	515	9	46	138	326	727	598	10	780	430	648	388	261
Rec.#	FSC-H	SSC-H	FL1-A	FL2-A	FL3-A																																																														
1	990	111	505	139	638																																																														
2	889	30	843	161	20																																																														
3	721	400	156	962	992																																																														
4	250	981	158	92	670																																																														
5	970	166	877	208	359																																																														
6	737	553	504	390	752																																																														
7	522	463	866	535	316																																																														
8	438	521	894	330	515																																																														
9	46	138	326	727	598																																																														
10	780	430	648	388	261																																																														

Composed of text (left) and binary data (right).

Solution: Split up and transform



Although Network Common Data Form (netCDF) files include both binary & metadata and thus conflict with our requirements, the advantages of the ubiquity of netCDF and the promised creation of a program that splits a netCDF file into binary and metadata, as well as translates the minimal metadata into XML are sufficient to mitigate this deviation.

Conclusions

- Only one standard is needed to describe cytometry data including both flow and image
- The objects should be defined in the XML schema design language
- The relationships between objects should be defined in RDF
- netCDF files can be part of the ACS.
- Compromises are necessary!
 - netCDF for list-mode; JPEG, TIFF, DICOM & Raw for Imaging.
- ISAC should limit its standard to areas within its expertise and interoperate with HL7, DICOM, and others for the rest.
- ISAC should join Unidata and the HL7-DICOM consortium.
- We do not want HL7 or similar organization to publish a different standard for cytometry.

Acknowledgements

This work would not have been possible without the benefit of the very useful discussions and collaboration that I have had with Josef Spidlen and Ryan Brinkman. However, any mistakes in this poster were made by me. This work and the previous work on CytometryML have been supported by the XML_Med division of Newport Instruments.