Cytometryml Vs. Fcs, Robert C. Leif, Newport Instruments, ISAC, Cyto 2017 Poster, Prog. Number: 225);

CytometryML vs. FCS

Robert C. Leif, Newport Instruments, rleif@rleif.com, This poster and its associated schemas are available at www.cytometryML.org

Abstract

Introduction: Although FCS is based on extensive, superb domain knowledge, it is an isolated system, does not take advantage of data structures and requires special programs to read and write the data. The large overlap between imaging and flow cytometry provides strong evidence that both modalities should be covered by the same standard. However diagnostic imaging to a large extent is described by DICOM, and eventually Health Level 7.

Problem: One major change in nomenclature of FCS 4.0 is that the term parameter has been changed to dimension. However, to maintain backward compatibility the letter "P" has been retained as the second character in FCS Keywords.

Solution: A study has been started to determine the amount of effort to translate FCS Keywords into XML elements. The XML schemas were translated into XML pages and are being tested with the values in Blimkie_et_al. Individual CytometryML elements have been linked to a FCS Keyword and/or DICOM by the inclusion of an attribute.

Introduction

Although FCS is based on extensive, superb domain knowledge, it has some problems.

1) It is an isolated system;
2) It does not take advantage of data structures;
3) It requires special programs to read and write the data.
4) It lacks the capability to interoperate or work with other standards.

One significant improvement to Cytometry standardization is being solved by developing a consensus that the large overlap between imaging and flow cytometry provides sufficient evidence to cover both of them by the same standard. The major difference between both branches of cytometry are in the sample fluidic system (Plumbing).

Other Standards: Diagnostic imaging, to a large extent is described by the medical imaging standard, DICOM, and eventually Health Level 7 (http://www.hl7.org/implement/standards/fhir/xml.html) HL7 has a computer assisted method of design, which generates schemas from graphical structures and has an emphasis on transmission of files.

The Digital Image and Communication Standard, DICOM, is presently described in 19 volumes (See below). Unless ISAC has greater knowledge on a subject, we should translate the existing DICOM datatypes into XML. When our expertise is greater than theirs, we should try to persuade DICOM to reuse the ISAC datatypes. In any event, DICOM can be translated into XML.

DICOM Volumes

DICOM Part 1: Introduction and Overview

DICOM Part 2: Conformance

DICOM Part 3: Information Object Definitions

DICOM Part 4: Service Class Specifications

DICOM Part 5: Data Structures and Encoding

DICOM Part 6: Data Dictionary

DICOM Part 7: Message Exchange

DICOM Part 8: Network Communication Support for Message Exchange

DICOM Part 10: Media Storage and File Format for Media Interchange

DICOM Part 11: Media Storage Application Profiles

DICOM Part 12: Media Formats and Physical Media for Media Interchange

DICOM Part 14: Grayscale Standard Display Function

DICOM Part 15: Security and System Management Profiles

DICOM Part 16: Content Mapping Resource

DICOM Part 17: Explanatory Information

DICOM Part 18: Web Services

DICOM Part 19: Application Hosting

DICOM Part 20: Imaging Reports using HL7 Clinical Document Architecture

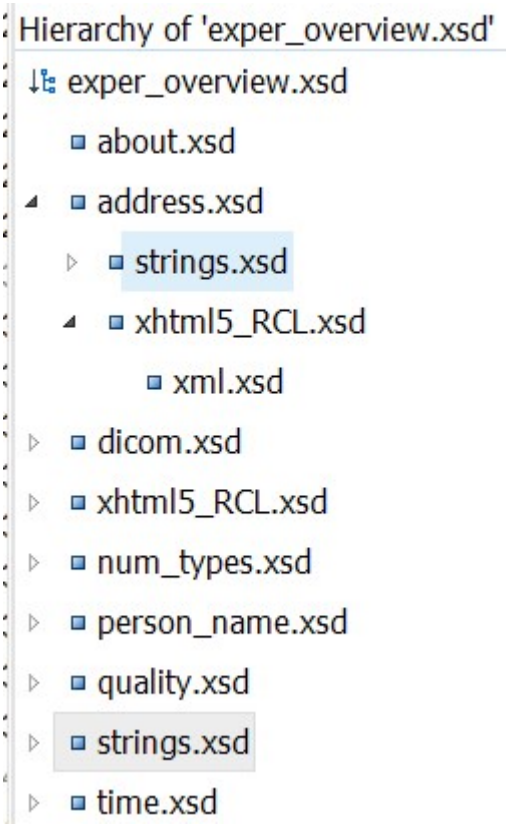
FCS 4.0 Problem: One major change in nomenclature of FCS 4.0 is that the term parameter has been changed to dimension. However, to maintain backward compatibility the letter P has been retained as the second character in the FCS Keywords.

Methods: CytometryML is based on object oriented development. Presently, the XML version of MIFlowCyt is a hierarchy of elements each of which is based upon its own schema. The root of the hierarchy is the MIFlowCyt_Info element, which has been generated from the miflowcyt.xsd XML schema. The MIFlowCyt_Info element includes linkages to the following elements: Experiment_Overview, Specimen_Description, Specimen_Treatment_Description, and a choice between Microscope_Series_and_Instance_Info or Flow_Series_and_Instance_Info elements. Each of these major elements has linkages to similar info and other elements. The Experiment_Overview is described in this poster.

CytometryML

The XML Schema Definition Language, XSD 1.1, from oXygen was used to describe MYFlowCyt. CytometryML unlike HL7, at present, does not include any description of data transmission. The HL7 design methodology sometimes has to avoid complex software constructs in schemas, such as choice elements.

XSD includes: datatypes, elements, and attributes. CytometryML almost exclusively employs datatypes, which can create: other datatypes, elements and attributes. The latter two entities can make data exist by storing, displaying and manipulating it. Many of the schemas serve as libraries of datatypes. A schema can export a datatype from one schema and export that datatype into another schema. Since datatypes are **not part** of the real-world, except for validation, they need minimal testing.



The oXygen **Resource Hierarchy/ Dependency tool** was used to show, in the form of a **tree** (shown on the left) a hierarchy with the **exper_overview schema** as its root. This root imports other schemas, which can each import other schemas. At each level, datatypes are provided to the higher-level schemas

CytometryML differs from FCS in that many of the FCS Keywords employ a single letter to describe an entity. CytometryML has the requirement to maximize readability. CytometryML uses long strings for names and minimizes the use of abbreviations. The abbreviations used include: Num for Number, Dim, which works for many uses of dimension, and exper for experiment. The long strings employ the use of compound words that are based on a sequence of abbreviated terms. In this sequence, the separator is the underscore character "_". The names of the datatypes are the same as the elements except that the suffix "_Type is added.

The ease of interpretation of the CytometryML element names is consistent with the CytometryML requirement for readability. The datatypes for the elements above are presently being integrated in the CytometryML MIFlowCyt schemas. XML files, which are generated from their schemas (design files). These XML schemas are and will be tested with the values in Blimkie_et_al. which is the example of what MIFlowCyt data should be. Blimkie et al. was published in Cytometry Part A; 2010: 77A:546–551.

Solution:

Keywords: A study has been started to determine the amount of effort to translate FCS Keywords into XML elements.

A code fragment from the FCS_4_0.xsd schema is shown below. The Code is augmented by a description, which is enclosed in a documentation element, which essentially an improved comment

```
<annotation>
          <documentation>Since keywords often include their data, Data
types in FCS 4.0 have Keywords starting with '$'. The FCS Keyword is an
attribute, which will facilitate the translation between FCS and
CytometryML. This is similar to what was previously done with the DICOM
elements.
        </documentation>
      </annotation>
      <simpleType name="Keyword_Type">
           <restriction
                base="string">
                <minLength value="2"/>
```

```
            <maxLength value="32"/>
            <pattern
                 value="[$]{1}[A-z]+"/>
        </restriction>
    </simpleType>
```

The FCS Keyword is a subset of strings that are restricted to have a length of from 2 to 32 characters. The first of which is the dollar sign and the rest are upper and/or lower case letters. The keywords are based upon the examples in the FCS documents, which do not include a formal description of in the form of a regular expression or regex. I recommend that the formal description the keyword datatype be a project for the DSTF.

The entire set of schemas and this poster can be found on the CytometryML.org web-page. Some of the keywords have already been coded; see Table 1, below, which shows their CytometryML name and location.

Table 1. Keywords already coded

Keyword	CytometryML Type	Schema
$BEGINSTEXT	Begin_Text_Type	FCS_4_0
$DATE	Acquisition Date_Time_Type	time
$PAR	Dimensions_Count_Type	dimensions
$PnB	Num_Bits_Type	dimensions
$PnE	Mode_Simple_Type	amplification
$PnN	Dimension_Type	channels
$PnR	Dimension_Number _Type	dimension
$PnT	Detector_Numeric_Type	detectors
$PnL	ADC_Max_Value_Type	pulses

A section of the XML page generated from the Dimension schema is included in this poster and will be described and discussed below.

Code Fragment

```
  <annotation>
     <documentation> From Draft of FCS 4.0: 3.3.31 $PAR [REQUIRED]
$PAR/n/ $PAR/12/
        The value of the $PAR keyword specifies the total number of FCS
dimensions stored in each event in the data set. In this example, data for
12 FCS dimensions are stored for each event."</documentation>
<documentation>
From CytometryML, Dimension_Num_Simple_Type, which has a range of 1 to
512, has been extended by the addition of 3 attributes the FCS Keyword
$PAR to produce Dimensions_Count_Type, which is the total number of
dimensions.  </documentation>
```

A small example from the Dimension schema is shown below.

```
<complexType name="Dimensions_Count_Type">
    <simpleContent>
        <extension base="
            dims:Dimension_Num_Simple_Type">
          <attribute
            name="FCS_Keyword"
            type="fcs:FCS_Keyword_Type"
            fixed="$PAR"/>
        </extension>
    </simpleContent>
</complexType>
```

The FCS Keyword is set to"$PAR, which provides a linkage between ACS and CytometryML.

Conclusions

1) The text of FCS could be translated into XML.
2) The text of DICOM could also be translated into XML
3) It should be possible for FCS and DICOM to be translated between each other.
4) FCS and DICOM are not to be replaced by CytometryML, they should be translated into XML
5) It has been much more difficult to display data in XML than html (Data NOT shown). This may in part be the results of differences between browsers.
6) There should be much more communication between standards groups.
7) The design and content of a standard have greater significance than the code.
8) The CytometryML schemas employ range checking. At this preliminary stage of our understanding, these ranges should be considered to be preliminary and not become fixed in the standard.

acknowledgements

The proofreading, suggestions and support of my family have been of great help. The meetings of the DSTF have been highly stimulating, educational, and interesting.