Unification of CytometryML, DICOM and FCS (P23)

Robert C. Leif[*A]

[A]**XML_Med, a Division of Newport Instruments,**
5648 Toyon Road, San Diego, CA 92115; rleif@rleif.com;
Location of this poster and supporting material:
http://www.newportinstruments.com

ABSTRACT

Introduction-Background: Although FCS 3.1 is a significant improvement over FCS 3.0, it is still a unique standard that requires special supporting programming to interface with other software. The Digital Imaging and Communications in Medicine (DICOM) standard has been extended to include pathology and is very slowly being transformed into being XML based. Health Level 7 Version 3 and the ISAC ACS are both being developed in XML. CytometryML is a proposed extension of ACS, DICOM, and FCS that should 1) permit interoperability and 2) provide sufficient information to permit the measurements to be reproduced.

Methods: DICOM and FCS designs, as well as ISAC Data Standards Task Force (DSTF) suggestions, and consensus items are being translated into the XML Schema Definition Language 1.1 (XSDL). The design of the schemas is object oriented and is checked by computer generation of XML pages that are each based on a key element (class) of one of the schemas. These XML pages are subsequently filled-in. CytometryML presently consists of 4 major schemas: Series, Instance, Instrument, and Specimen; it also includes Image and List-mode schemas. Metadata that is specific for an entire collection of images and/or list-mode files produced by a single instrument and derived from a single specimen constitute a series. Metadata that is specific for individual or closely related images and/or list-mode files produced by a single instrument and from a single specimen together with the binary data constitute an instance. A container (ZIP) file is used to store a series file together with its associated metadata files. A similar container file stores an instance file together with its associated metadata files and binary files.

Results: The Instance and Instrument schemas have each undergone one major revision, where two child schemas (Flow and Microscope) of Instrument were created. These schemas, and the other major schemas and XML pages created from each of them have been validated and tested. As of the end of March 2010, 77 XSDL schemas totaling in excess of 900,000 bytes (disk space) have been created.

Conclusions: Reuse of the well tested DICOM model resulted in a great decrease in the design and documentation efforts and increased probability of reliability and interoperability. CytometryML when combined with the compatible work of the ISAC Data standards Task Force (DSTF) and the XML development work of the DICOM working groups should meet the DSTF requirements and be suitable for incorporation as part of the proposed US Health Informatics Technology Strategic Plan as part of a national and/or international medical-biological data standard.

The CytometryML prototype of the Advanced Cytometry Standard (ACS) has the benefits of including image and list-mode data and being based on XML. The use of XML should permit compatibility and interoperability with existing medical and scientific informatics standards.

1. INTRODUCTION

1.1. Plethora of Biomedical Software standards

From Robin Cover's Pages http://xml.coverpages.org/healthcare.html

XML in Clinical Research and Healthcare Industries

Provisional [work in progress] collection of references to standards activities and formal specifications used in clinical research and healthcare industries. Not intended to be complete.

* Standards, Standards Bodies, and Healthcare Initiatives

How many of these can interoperate? (24 standards were listed. FCS was <u>not</u> included.)

Examples:

ASTM Committee E31 on Healthcare Informatics, Clinical Data Interchange Standards Consortium (CDISC, Digital Imaging and Communications in Medicine (**DICOM**), Health Insurance Portability and Accountability Act (HIPAA), Health Level

Seven (**HL7**), Integrating the Healthcare Enterprise (**IHE**), Logical Observation Identifiers Names and Codes (**LOINC**), Health Record Foundation (openEHR), Systematized Nomenclature of Medicine (**SNOMED**),

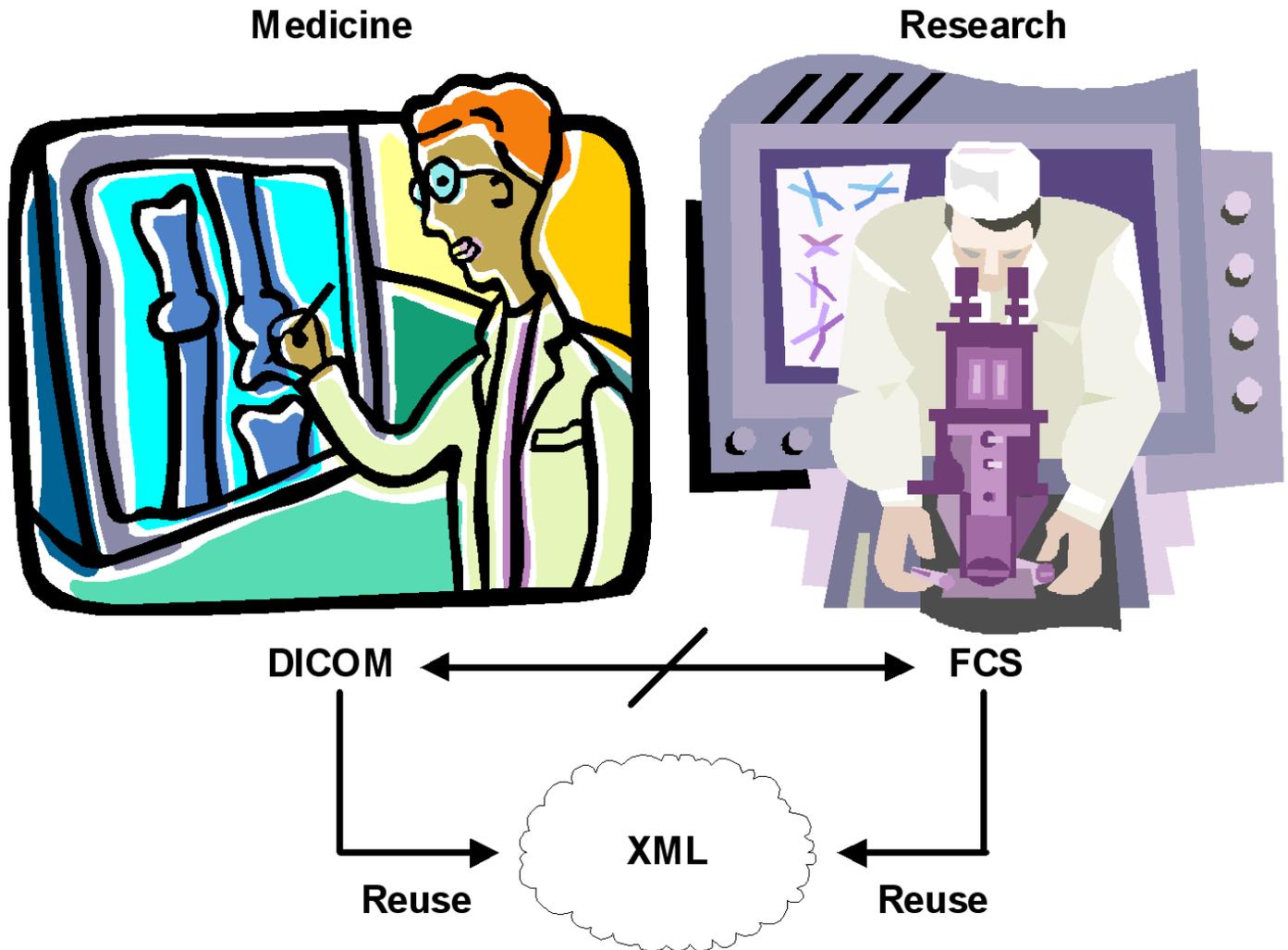## 1.2. Cytometry & Cytology-Pathology

Bridging two standards

Fig. 1. There is significant overlap in the coverage of cytometry and/or cytology in the existing standards of two groups, the pathologists and the cytometrists.

- Transfer of technology from the cytometrists to the pathologists.

- Different data file standards.

- College of American Pathologists is adopting the Digital Imaging and Communications in Medicine (DICOM) standard (http://medical.nema.org/) which includes Supplements 122 & 145:

122: Specimen Module and Revised Pathology SOP (Service-Object-Pair) Classes

- Specimen, Bar Code, Accession number, Worklist, Sampling, and Procedure Step (Staining).

145: Whole Slide Microscopic Image IOD (Information Object Definition) and SOP Classes

- Images, Image Pyramid Storage, Pixel Matrix (x,y,z), Focus, and Optical Path, which is incomplete.

Cytometrists use the International Society for Advancement of Cytometry, ISAC (http://www.isac-net.org/) Data File Standard for Flow Cytometry, Version FCS3.0 and FCS3.1.

- Syntaxes used for both FCS and DICOM are unique and require software interfaces to work with other applications.

- Both groups have started to create software in XML. The ISAC data standards task force has created Gating-ML: XML-Based Gating Descriptions in Flow Cytometry. The DICOM Working Group 27 is creating schemas that are an extension to the capabilities of DICOM Part 18: Web Access to DICOM Persistent Objects (WADO) (ftp://medical.nema.org/medical/dicom/2008/08_18pu.pdf).

- WADO schemas will permit transmission of some data via the Web Services Description Language, WSDL (http://www.w3.org/TR/wsdl20/), between DICOM data stores and XML.

- Leverage this work on WSDL to produce strongly typed (safer) XML.

DICOM Organization
1 Patient (Medical Record)

↓

X Studies
_____

↓

Y Series

↓
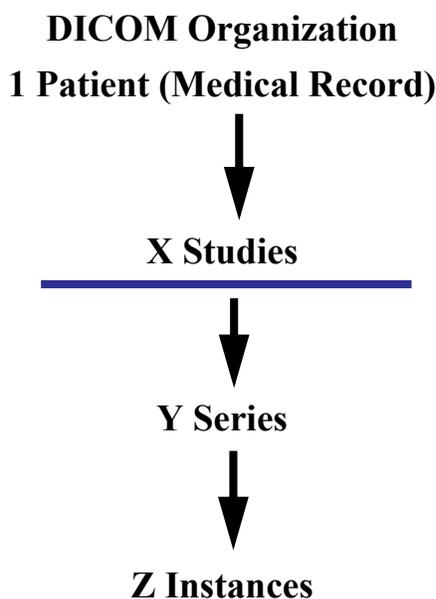
Z Instances

The items below the line are the part of the CytometryML schemas that will be discussed.

2. CYTOMETRYML

- Three schemas instance.xsd, instrument.microscope.xsd or instrument.flow, and series.xsd that include strongly typed data together with many others have been created and tested.
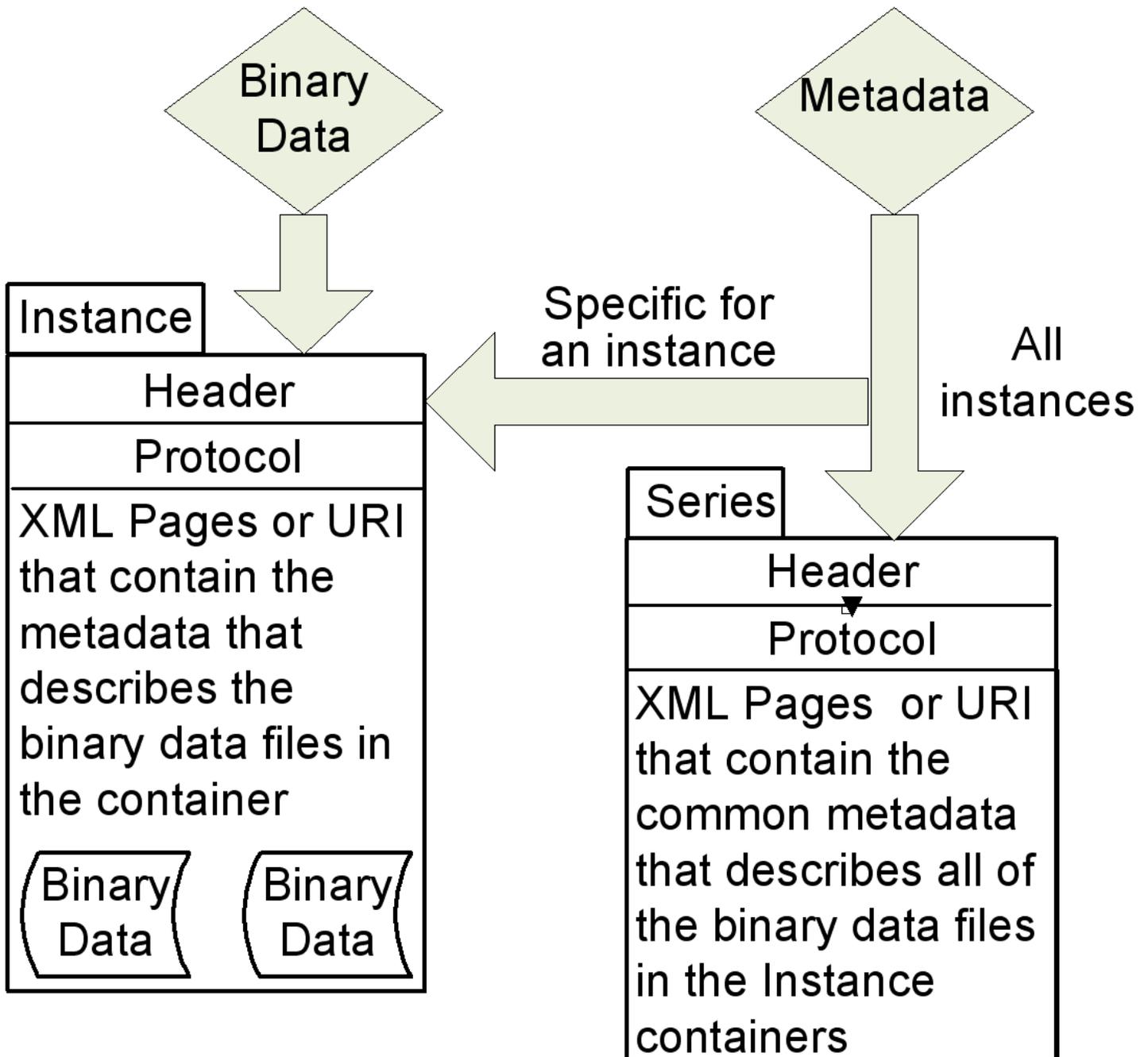
Fig. 2. Diagram showing the placement of binary data and metadata into the instance and series containers. The Series_Data_Type (right) and the Instance_Data_Type (left) and their corresponding elements each contain Header Information and parts of the Protocol.

2.0.1 Series

- A series contains metadata that indicates the locations of one or more measurements, images and/or list-mode files, produced by a single instrument on a specimen or specimens that came from a common ancestor.

- The description of the protocol that is common to the preparation and analysis of **all specimens** prior to the subsequent treatment required for the individual measurements is part of the **series**.

- The series container includes the description of entities that are the **same for all instances** that are part of the series.

- ▬▬ Also only included in the series protocol element are: the total number of instances and a brief summary of the information in each instance container file.
- It is anticipated that the user will often look first at the series file in order to select instances.

2.0.2 Instance

- The description of the protocol that is specific for the creation and characterization of the individual or closely related measurements, such as images and list-mode files, is part of the instance.

- Only one instrument and many channels.

    - ▬▬ Redundant to repeat the information contained in either instrument.flow.xsd or instrument.microscope.xsd with the data that describes each channel.
    - ▬▬ The content of the protocol element of an instance is specific to that instance.

- The instance container file includes the **binary data** (list-mode, images, and index files), which are referenced by the protocol element contained within an instance element.

    - ▬▬ These individual protocol elements also reference the list-mode and image context files that contain or point to the metadata similar to that, which previously had been included in FCS and DICOM, such as FCS, JPEG, or TIFF files.
    - ▬▬ Also reference the individual channels settings for each parameter. Settings such as staining protocol and optical configuration can change for each instance XML document.

2.0.3 ZIP Data Files

- The containers for both the series and instance files are ZIP files. The series and instance files will contain XML including XHTML files. The instance files also contain the binary data.

    - ▬▬ **Firefox** XPI files are a file hierarchy that has been compressed into a zip file."
    - ▬▬ According to **Microsoft** TechNet: "The new XML formats are based on industry-standard XML and ZIP technologies"
    - ▬▬ **Google** Chrome browser extensions are a zipped bundle of files — HTML, CSS, JavaScript, images, and anything else the user or programmer needs to add functionality to the application.

2.0.4 Instrument

- The instrument XML page that includes the elements that describe the fixed parts of the instrument (Flow Cytometer or Microscope)

    - ▬▬ Can be part of the series or pointed to by a URL (manufacturer's web site).
    - ▬▬ The fixed elements have constant values or settings for all acquisitions of the data contained in all instances that are part of the series.
    - ▬▬ Include detailed descriptions of the manufacturer, serial number, and similar data for both the microscope and flow cytometer elements.

3. METHODS

3.1 Disclaimer

If a claim of adequate safety is to be truly valid, experimental data to substantiate the appropriateness of the development and testing techniques needs to be provided.

3.1.1 Unfortunate fact

- The quality of software development techniques has **<u>not</u>** been adequately measured

3.2 Rational

- Much of the information and data-types present in the XML schemas and subsequently XML pages were reused from Digital Imaging and Communications in Medicine (DICOM) standard (http://medical.nema.org/) or Flow Cytometry Standard, FCS.

- New data-types were created and data-types from other CytometryML schemas were reused.

- Both DICOM and FCS were prepared by domain experts.

- Because DICOM is a FDA Class II device, the safety of the software developed as part of a standard should be maximized.

  - Readability, modularity, strong typing, and reuse are four software engineering principles that have been applied to the CytometryML XML schemas.

  - This was possible because of the use of the XML Schema Definition Language (XSDL) structures (http://www.w3.org/TR/xmlschema11-1/) and data-types (XSD) (http://www.w3.org/TR/xmlschema11-2/).

  - Many of the other requirements are met for a data file standard or facilitated by the use of XSDL, and the structure of CytometryML.

  - XSDL schemas validated by oXygen (http://www.oxygenxml.com/) and XMLSpy (http://www.altova.com/).

  - Validated with XSDL 1.1.

  - New XML page produced from each of the main schemas and then filled with the values from the original XML page, and validated.

4. RESULTS

- The Protocol element is one of the two parts of main metadata element of the instance XML file. The other is the Instance_Header.

- A section of the content of each of the CytometryML instance.xsd and instrument.microscope.xsd schemas will be described in terms of the XML pages generated from these schemas.

- Optical path, because it can change, occurs as part of the Protocol element of the instance XML document.

- Each Protocol element contains one Channel_Info element that contains the elements present in Table 1.

Table 1: Channel_Reference Elements (simplified)

Elements	Example of Values
Analyte Reporter	Anti5Brdu
Parameter	FL1-A
Channel Number	3
Measurement	Fluorescence
Long Name	AlexaFluor
Optical Path	Described below
Statistics	CV= 3.0%
Quality Assurance	Bead-based alignment setup

The Example of Values column of Table 1 includes in most cases only one of the values of one of the parts of each element.

- The optical path element of an episcopic illuminated systems are described in detail in Fig. 3.

- The order of the optical path has been defined,

   - Positions of excitation optical elements have negative values;

   - Positions of imaging elements have positive values.

   - The position of the slide or flow cell that holds the specimen is 0.

- The optics go in a positive direction towards the detector and a negative direction towards the light source.

4.1 Optical_Path element example

The CytometryML schemas and XML test pages (including the examples below are located at http://www.newportinstruments.com/cytometryml/cytometryml.htm

Fig. 3 shows the optical path of a fluorescence microscope, which is described in the Optical_Path element of the Channel_Reference element that is located within the instance Protocol element.

The optical parts are numbered with the specimen being 0 and the part of the path that goes to the detector having positive values and the part that emanating from the light source having negative values. The order values of the different optical components are shown in Fig. 3 and lines 3, 13, 14, 15, 35, and 37 of the Optical_Path Code Fragment. Because this is an epiilluminated system, where the excitation and emission beams are separated by a dichroic mirror, some of the components have two values. The objective (1) is also the condenser (-1). The dichroic mirror (line 14) and the objective (line 15) are parts of both the excitation and emission paths.
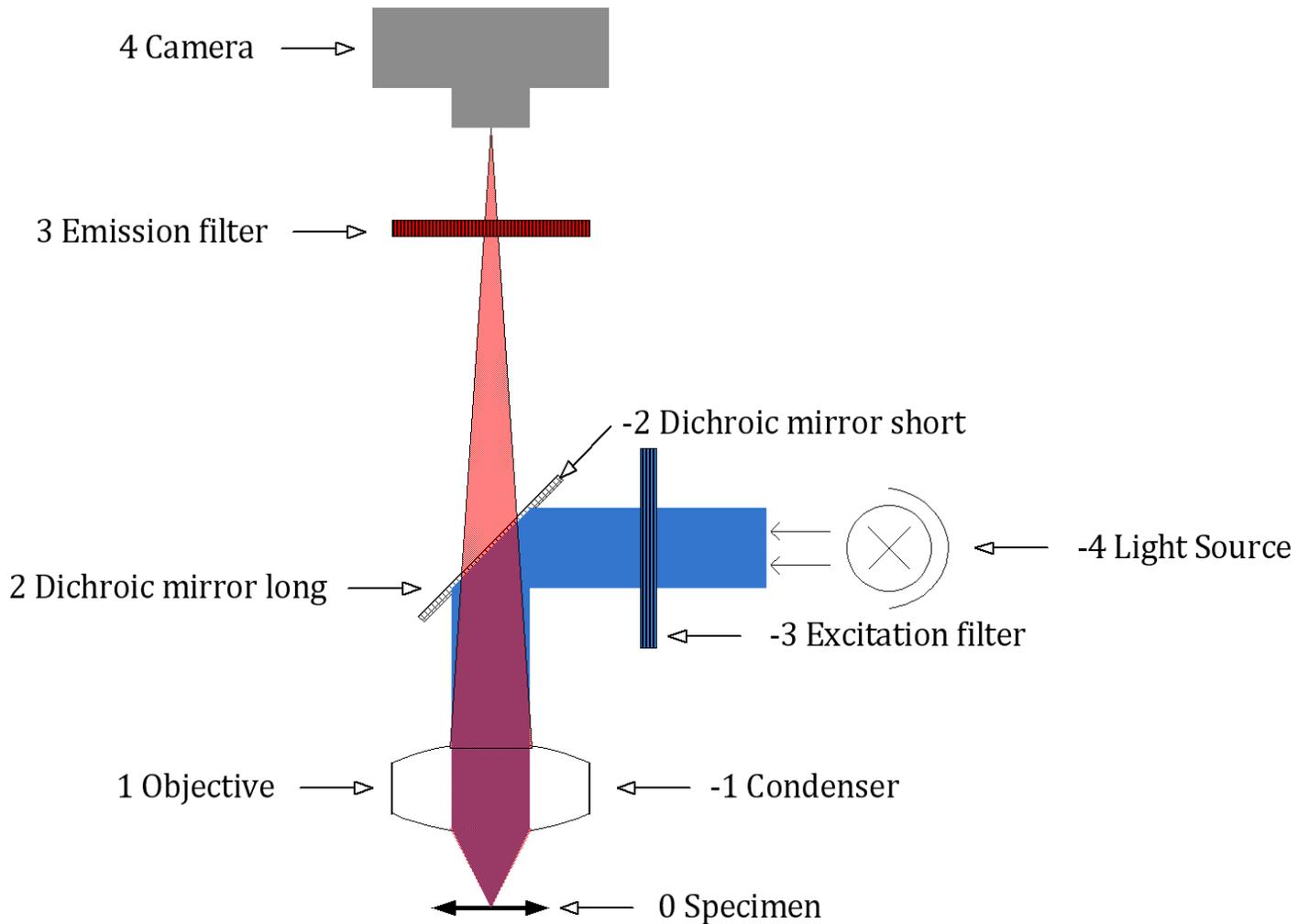
Instance Optical_Path Fragment

Fig. 3 is a cartoon of an episcopic fluorescence optical microscope. Each of the components have been numbered in the order of their presence in the excitation and emission paths.

```
1 <channels:Optical_Path>
2   <channels:Path>Episcopic_Illumination</channels:Path>
3   <channels:Light_Source_Info  Order="-4"
                      UID="1.111.11.112.11.2>
4       <excite:Short_Name>LED365</excite:Short_Name>
5       <excite:Light_Source>LED</excite:Light_Source>
6       <excite:Wavelength_Wavelength="365"
            Units="nm"/>
7       <excite:Power  Units="milliwatt"  Power="200"/>
8       <excite:Polarization>None
9       </excite:Polarization>
10      <excite:Description>Came with microscope
11      </excite:Description>
12  </channels:Light_Source_Info>
13  <channels:Excitation_Filter  Order="-3"
                  Short_Name="Wide365"
                  UID="001.001.0002.001.5"/>
14  <channels:Dichroic_Mirror  Imaging_Order="2"
```

```xml
                        Excitation_Order="-2"  Short_Name="Pass365"
                        UID="001.001.0003.001"
                        Location="Parallel_Area"/>
15<channels:Objective_Info
        Location="Object_Focal_Plane"
                Imaging_Order="1"  Excitation_Order="-1"/>
16    <optics:Magnification>40
17    </optics:Magnification>
18    <optics:NA>0.7</optics:NA>
19    <optics:Contrast>None</optics:Contrast>
20    <optics:Field_Flatness>Plan
21    </optics:Field_Flatness>
22    <optics:Immersion>Air</optics:Immersion>
23    <optics:Chromat>Fluorite</optics:Chromat>
24    <optics:Abbreviated_Info
25        UID_Value="001.001.0003.001.5">
26        <item:Identifier>ID_2</item:Identifier>
27        <item:Manufacturer>Any microscope company
28        </item:Manufacturer>
29        <item:Model_Name>high dry</item:Model_Name>
30        <item:Description>
                            high dry that came with the microscope.
31        </item:Description>
32    </optics:Abbreviated_Info>
33 </channels:Objective_Info>
34 <channels:Detector_Emission_Filter
35    Short_Name="Center530"  Order="3"
36    UID="001.001.0002.001.6"/>
37 <channels:Detector  Order="4">
38    <channels:Camera_Info>
39        <cameras:Abbreviated_Info
40            UID_Value="001.001.0004.001.7">
41            <item:Identifier>ID_10</item:Identifier>
42            <item:Manufacturer>Point Grey
                        </item:Manufacturer>
43            <item:Model_Name>Dragon2
44            </item:Model_Name>
45            <item:Description>
                            Does analog time-gated illumination
46            </item:Description>
47        </cameras:Abbreviated_Info>
48        <cameras:Columns>640</cameras:Columns>
49        <cameras:Rows>480</cameras:Rows>
```

```
50      <cameras:Technology>CCD
51      </cameras:Technology>
52      <cameras:Intensified>
53         <cameras:Not_Intensified>
54            true</cameras:Not_Intensified>
55      </cameras:Intensified>
56      <cameras:Binning>2</cameras:Binning>
57      <cameras:Exposure_Duration
58         Prefix="milli"  Units="Seconds">
59         1.0</cameras:Exposure_Duration>
60      <cameras:Exposure_Off_Duration
61         Prefix="milli"  Units="Seconds">1.0
62      </cameras:Exposure_Off_Duration>
63      <cameras:Summation_Mtd>
64         <cameras:Method>Analog</cameras:Method>
65         <cameras:Num_Exposures_Summed>100
66         </cameras:Num_Exposures_Summed>
67      </cameras:Summation_Mtd>
68      <cameras:Temperature_Centigrade>23
69      </cameras:Temperature_Centigrade>
70   </channels:Camera_Info>
71 </channels:Detector>
72</channels:Optical_Path>
```

Series Microscope Instrument Light Source

```
1 <instr:Light_Source>
2 <excite:Light_Source>LED</excite:Light_Source>
3 <excite:UID>1.111.11.112.21.2</excite:UID>
4 <excite:Emitter>GaAlAs</excite:Emitter>
5 <excite:Wavelength_Wavelength="365"  Unit="nm"/>
            <excite:Max_Power  Units="milliwatt"
                Power="250"/>
6 <excite:Polarization>None</excite:Polarization>
7 <excite:Object_Plane  Units="mm"  Width="0.50"
                Shape="Circular"  Height="0.50"/>
8 <excite:Description>Stock Nichia
        </excite:Description>
9 <excite:General_Info
        UID_Value="1.111.11.112.11.2">
10   <item:Identifier>ID_5</item:Identifier>
11   <item:Manufacturer>Nichia</item:Manufacturer>
12   <item:Model_Name>UV LED</item:Model_Name>
13   <item:Model_Number>12345678
```

```
          </item:Model_Number>
14    <item:Description>UV LED that can be pulsed at 1 kiloHz
15    </item:Description>
16    <item:Item_Serial-number>01234567
          </item:Item_Serial-number>
17    <item:URI_Var>http://www.Nichia.com
          </item:URI_Var>
18 </excite:General_Info>
19</instr:Light_Source>
```

- Content of the instance file substantially differs from that contained in the Instrument (Microscope) file.

- Instrument file is either located within the series container or pointed to by the series XML document.

- Instance file describes the settings and configuration used to perform the measurement

  - Can and often does differ between measurements (instances).

- The Instrument file provides a detailed description of the instrument and its components.

  - These details are unchanged between instances.

  - Instrument file or its URL or both are included in the series container.

Comparison of the Light_Source_Info element [lines 3 to 12] contained in the Instance Optical_Path Fragment with the instr:Light_Source element of the Instrument.Microscope.XML page that is contained in a series container shows: 1) Light_Source_Info contains only 8 items and that only three of these: Light_Source, Wavelength, and polarization (shown underlined) had the same name and values in the Light_Source element. Order numbers in the Light_Source_Info element refer to the actual configuration used for the measurement; whereas, order numbers used in the Light_Source element can refer to elements in a drawing etc.

The content of the Light_Source element of the instrument file is extensive. It includes: the emitting material, the maximum power output, the dimensions and shape of the image of the Light_Source at the object plane, and manufacturer related information including the manufacturer's URL.

It is anticipated that the instrument files for commercial medical devices will be under the control of the manufacturer. In the case of instrument developers and researchers, the instrument file in most cases will by necessity be under the researcher's control.

The instance Optical_Path and other elements in an instance.XML page serves two purposes. The first purpose is to permit the person who reads the XML page to understand how the measurement was made and what was measured. The second purposes to permit the measurement to be repeated. In order accomplish this, a record must be made, preferably at the time of the measurement, that describes the configuration of the microscope (instrument), which together with the detailed information of the instrument components (parts), provided by the instrument.microscope.XML page is sufficient to repeat the experiment.

5. DISCUSSION

This iteration of the code development for CytometryML has demonstrated the feasibility of applying the DICOM design specified organization of series and instances to cytometry data. It has also demonstrated that at least a significant part of DICOM series and instance metadata can be kept in the form of XML pages. This use of XML has the very significant advantages over the present DICOM standard of innate interoperability and being in a format that can be validated.

6. CONCLUSIONS

1. The description of the settings for a Cytometry measurement should be separated into two parts: that which is common for the **series** of measurements and that which is specific for the individual measurement or closely related group of measurements (**instance**).

2. Maximizing reuse including reuse of designs and documentation, besides increasing safety and minimizing development costs, should significantly help to improve international medical informatics infrastructure and facilitate interoperability.

3. It has been possible with XSDL to maximize readability, create a modular structure, and strongly typed, reusable data-types.

4. Much of FCS & DICOM has been translated into XSDL and then into XML.

5. DICOM can and should and is being extended in XML and should, sometime in the future, evolve into an XML based standard, which is compliant with the requirements of US Health and Human Services Health Information Technology.

6. The CytometryML schemas can be used with other applications including Microsoft® Office and to generate XML web pages and in the future HTML 5 pages.

7. This infrastructure improvement should benefit the patients while significantly decreasing health care costs.