FlowRDF: A Demonstration of Describing Flow Cytometry Metadata Using the Resource Description Framework

Abstract:

Metadata and analysis components of flow cytometry data files (FCS) are not recorded in sufficient detail for use for data interpretation by independent parties. Moreover, if metadata annotation takes place at any time subsequent to data capture, the all-inclusive format of FCS necessitates the generation of a new version of the file, which replicates the primary data.

This document reflects the need of providing metadata information in a standardized and extensible way. For this purpose, it does not propose standardization of a new approach; instead, it demonstrates how the Resource Description Framework (RDF) standard could be used for annotation of flow cytometry data and for expressing relations between resources (e.g., between different files).

RDF enables encoding, exchange and reuse of structured metadata. It is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics. Moreover, it provides means for publishing both human-readable and machine-processable vocabularies designed to encourage the reuse and extension of metadata among disparate information communities. Within this document we demonstrate how to reuse standardized RDF metadata elements defined by resource description communities (e.g., the Dublin Core Metadata Initiative), how to create and reference flow cytometry specific metadata elements, and how to use biological ontologies to annotate FCM data.

Keywords: Flow Cytometry, Resource Description Framework (RDF), Ontology

Josef Spidlen (jspidlen@bccrc.ca)
Ryan Brinkman (rbrinkman@bccrc.ca)
and the Bioinformatics Standards for Flow Cytometry Consortium (www.flowcyt.org)

Acknowledgement

This work has been supported by NIH grant EB005034 from the National Institute of Biomedical Imaging And Bioengineering. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Biomedical Imaging And Bioengineering or the National Institutes of Health.

February 15, 2008

EXECUTIVE SUMMARY

The Flow Cytometry Standard (FCS) [B1] specification has kept pace with many years of technological evolution. FCS has been adopted for the common representation of flow cytometry (FCM) data, and this standard is supported by virtually all analytical instrument and third party software suppliers. Scientists can choose among instruments and software with no major compatibility issues for the raw fluorescence values that FCS captures. FCS version 3.0 and earlier includes data, metadata and analysis components within the same file. However, metadata and analysis components, if included at all, are not recorded in a fully standardized fashion and in sufficient detail for use by independent parties to interpret the experiments. Moreover, if metadata annotation takes place at any time subsequent to data capture, the all-inclusive format of FCS necessitates the generation of a new version of the file, which replicates the primary data.

A recent trend in data standard development has been to use Extensible Markup Language (XML) as the preferred mechanism to define data representations. We suggest using XML where appropriate based on the character of stored information, as exemplified by the proposed Gating-ML [B2], Transformation-ML [B3], and Compensation-ML [B4] file formats. However, the syntactic and document-centric XML may not be able to achieve the level of interoperability required for highly dynamic metadata, such as FCM experiment annotation [B5].

RDF enables encoding, exchange and reuse of structured metadata. It is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics. Moreover, it provides means for publishing both human-readable and machine-processable vocabularies designed to encourage the reuse and extension of metadata among disparate information communities. Constructing most useful RDF descriptions requires a standard ontology - standardized description of objects, elements, processes, and relations of a particular domain [B6].

Within this document we demonstrate 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).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

FlowRDF: A Demonstration of Describing Flow Cytometry Metadata Using the Resource Description Framework

1. Overview

1.1 Scope

This document demonstrates how to utilize the Resource Description Framework standard for annotation of FCM data.

1.2 Purpose

Metadata and analysis components of FCM data files (FCS, [B1]) are not recorded in a fully standardized fashion or in sufficient detail for use by independent parties to interpret the experiments. Moreover, if metadata annotation takes place at any time subsequent to data capture, the all-inclusive format of FCS (version 3.0 and earlier) necessitates the generation of a new version of the file, which replicates the primary data. This document reflects the need of providing metadata information in a standardized, extensible, and flexible way. It demonstrates how to use the RDF standard and related technologies for annotation of FCM data.

1.3 Normative references

The following referenced documents represent standards reused within this document and thus are essential for the understanding of this document.

a) W3C Recommendation: RDF Primer, http://www.w3.org/TR/rdf-primer/,[B7].

b) W3C Recommendation: RDF/XML Syntax Specification (Revised), http://www.w3.org/TR/rdf-syntax-grammar/, [B8].

c) W3C Recommendation: RDF Vocabulary Description Language 1.0: RDF Schema, http://www.w3.org/TR/rdf-schema/, [B9].

d) W3C Recommendation: Resource Description Framework (RDF): Concepts and Abstract Syntax, http://www.w3.org/TR/rdf-concepts/, [B10].

e) W3C Recommendation: RDF Semantics, http://www.w3.org/TR/rdf-mt/, [B11].

f) W3C Recommendation, 04 February 2004, Extensible Markup Language (XML) 1.0 (Third Edition), http://www.w3.org/TR/REC-xml/, [B12].

g) Dublin Core Metadata Initiative: DCMI Recommendation, Dublin Core Metadata Element Set, Version 1.1: Reference Description, http://dublincore.org/documents/dces/, [B13].

h) Dublin Core Metadata Initiative: DCMI Recommendation, DCMI Metadata Terms, http://dublincore.org/documents/dcmi-terms/, [B14].

i) Dublin Core Metadata Initiative: DCMI Recommendation, DCMI Type Vocabulary, http://dublincore.org/documents/dcmi-type-vocabulary/, [B15].

j) W3C Recommendation: OWL Web Ontology Language – Overview, http://www.w3.org/TR/owl-features/, [B16].

k) W3C Recommendation: OWL Web Ontology Language – Guide, http://www.w3.org/TR/owl-guide/, [B17].

l) W3C Recommendation: OWL Web Ontology Language – Reference, http://www.w3.org/TR/owl-ref/, [B18].

m) W3C Recommendation: OWL Web Ontology Language – Semantics and Abstract Syntax, http://www.w3.org/TR/owl-semantics/, [B19].

n) W3C (submission): DAML+OIL Reference Description, http://www.w3.org/TR/daml+oil-reference, [B20].

o) W3C (submission): Representing vCard Objects in RDF/XML, http://www.w3.org/TR/vcard-rdf, [B21].

1.4 Acronyms and abbreviations

DDC	Dewey Decimal Classification
FCM	Flow Cytometry
FCS	Flow Cytometry Standard
FuGO	Functional Genomics Investigation Ontology
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
IMT	Internet Media Type
ISAC	International Society for Analytical Cytology
ISO	International Organization for Standardization
LSID	Life Science Identifier
LCC	Library of Congress Classification
LCSH	Library of Congress Subject Headings
MeSH	Medical Subject Headings
OBI	Open Biological Ontology
RFC	Request for Comments
RDF	Resource Description Framework
RDFS	RDF schema
UDC	Universal Decimal Classification
URL	Uniform Resource Locator
URN	Uniform Resource Name
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1.5 Resource Description Framework (RDF)

1.5.1 Introduction

The Resource Description Framework (RDF, [B7], [B8], [B9], [B10], [B11]) developed under the auspices of the World Wide Web Consortium (W3C), is a language that enables the encoding, exchange and reuse of structured metadata. RDF is an application of XML [B12] that imposes needed structural constraints to provide unambiguous methods of expressing semantics. RDF supports the use of conventions that facilitate modular interoperability among separate metadata element sets. These conventions include standard mechanisms for representing semantics that are grounded in a simple, yet powerful, data model (section 1.5.2). RDF additionally provides means for publishing both human-readable and machine-processable vocabularies designed to encourage the reuse and extension of metadata semantics among disparate information communities. Vocabularies are the set of properties, or metadata elements, defined by resource description communities, e.g., the Dublin Core Metadata Initiative ([B13], [B14], [B15]).

The introduction to RDF contained within this document provides only a basic overview to assist understanding of the rest of the document. References [B22] and [B23] are recommended as a further gentle introduction into RDF.

1.5.2 RDF data model

An RDF document is a set of *RDF triples*. Each triple defines a single bit of knowledge, which ideally allows both human and machine understanding. An RDF triple consists of a *subject*, a *predicate*, and an *object*, together interpreted as a simple *RDF statement*. The RDF data model can be illustrated by an example:

- The author of file_1 is Jack Kerouac.
- RDF triple: (“file_1”; “author”; “Jack Kerouac”).
- Graph of the data model (Figure 1):

Figure 1 – Graph of an RDF statement

Figure 1 shows that an RDF data model may be interpreted as a directed labeled graph that represents statements about resources. A *resource* is defined as anything that is uniquely identifiable by a Uniform Resource Identifier (URI). Resources have *properties* (attributes or characteristics; e.g., the author of a resource). The properties associated with resources are identified by predicates (property-types). Predicates express the semantic of relationships and they are also uniquely identifiable by a URI (which makes them special kind of resources too).

In RDF, objects may be atomic in nature (e.g., text strings or numbers) or other resources, which in turn may have their own properties. An atomic object in RDF is called a *literal object*. A collection of properties that refers to the same resource is called a *description*.

1.5.3 RDF blank nodes

A *blank node* represents a resource that is not currently identified. Most commonly, a blank node is used when a resource URI is not meaningful. In RDF graphs, a blank node is represented by an oval (it is a resource), with either no value in the oval or a computer-generated identifier. Most of the RDF tools generate an identifier for blank nodes to differentiate them. Figure 2 illustrates an RDF graph having a blank node.

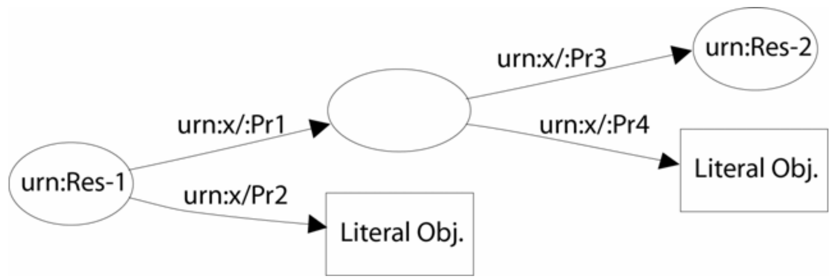

Figure 2 – RDF graph combining resources, literals and a blank node

1.5.4 RDF syntax

RDF/XML [B8] is the official serialization technique for RDF. It provides means of documenting an RDF model in an XML-based format. RDF/XML will be illustrated on examples within this document.

1.5.5 RDF schema

RDF schema (RDFS) is an extension to RDF that provides a framework to describe application-specific classes and properties. RDF schemas are used to declare vocabularies – sets of semantics property-types defined by/for a particular community. RDF schemas do NOT specify the syntax of RDF documents (unlike XML schemas that do prescribe the syntax of XML documents).

2. Flow cytometry specific RDF schema

2.1 RDFS classes for flow cytometry

Table 1 shows an initial proposal (example) of FCM specific RDF Schema classes that would be possible to define in RDFS. They represent semantic types of FCM-specific resources that are reused while describing FCM data and metadata; and, having these formally defined enables for formal definition of FCM-specific properties (2.2).

Table 1 – FCM-specific RDFS classes

Class ID	Label	Comment
fcsdata	FCS Data	Any encapsulation of FCM data, e.g., an FCS data file or an FCS data set.
datafile	FCS Data File	An FCS Data File is a subclass of fcsdata.
dataset	FCS Data Set	An FCS Data Set is a subclass of fcsdata and it is part of a data file.
gating	Gating File	Gating-ML resource (specification of gates).
transformation	Transformation File	Transformation-ML resource (specification of transformations that were used while processing data).
compensation	Compensation File	Compensation-ML resource (specification of compensation used while processing data).

FCS 3.0 and earlier includes experimental metadata within FCS files. Proposed changes separate the binary data from the experimental metadata within separate files, which may be stored using contemporary technologies such as RDF and XML.

Table 2 summarizes how FCS keywords could be replaced by the mentioned RDFS resource classes.

Other metadata may be described using controlled vocabularies (e.g., [B13], [B21]) and referencing relevant terms from biological ontologies (e.g., OBI [B24]) or using XML-based standards (i.e., XML schema defined standards). These may be reused or adapted from previous work (e.g., CytometryML [B25], or reusable XML schemas from HL7 [B26] or DICOM [B27]).

However, prior to deciding these, the minimum required information to interpret a cytometry experiment shall be determined. Standards shall be developed or adopted to encode this information in electronic format. One example of such a minimum description is MIFlowCyt [B28]. Although MIFlowCyt identifies the details necessary to describe a flow cytometry investigation, it does not provide a precisely defined format for data and metadata representation. Such a format is needed to cope with the growing flood of flow cytometry metadata, and to facilitate automated communication between different laboratory information management systems, databases and data analysis tools.

Table 2 – RDFS classes replacing FCS metadata

Class ID	Replaced FCS keywords	
gating	$GATE	Number of gating parameters.
	$GATING	Specifies region combinations used for gating.
	$GnE	Amplification type for gating parameter number n.
	$GnF	Optical filter used for gating parameter number n.
	$GnN	Name of gating parameter number n.
	$GnP	Percent of emitted light collected by gating parameter n.
	$GnR	Range of gating parameter n.
	$GnS	Name used for gating parameter n.
	$GnT	Detector type for gating parameter n.
	$GnV	Detector voltage for gating parameter n.
	$RnI	Gating region for parameter number n.
	$RnW	Window settings for gating region n.
transformation	$PnD	Suggested display scale for parameter n. (proposed in FCS3.1)
compensation	$COMP	Fluorescence compensation matrix.
	$SPILLOVER	Fluorescence compensation (proposed in FCS3.1)

2.2 RDFS properties for flow cytometry

Table 3 presents an example of RDFS properties that could be used to connect FCS data with other RDF resources such as a gating description file. A property may have a domain and a range, i.e., the subject and object resource type. Within this document we use FCS data as the subject (domain) for the gated_by, transformed_by, and compensated_by properties.

Table 3 – Example of FCM-specific RDFS properties

Property ID	Domain	Range	Comment
gated_by	fcsdata	gating	FCS data resource is gated by a gating resource.
transformed_by	fcsdata	transformation	Parameters of an FCS data resource are transformed by a transformation resource.
compensated_by	fcsdata	compensation	FCS data resource is compensated by a compensation resource.

3. Experiment annotation

3.1 Namespace prefixes

Methodology of how to use RDF to annotate FCM experiments will mainly be indicated on examples. The following namespace prefixes are used in this document (Table 4):

Table 4 – Summary of namespace prefixes used within this document

Prefix	URI
daml	http://www.daml.org/2001/03/daml+oil#
dc	http://purl.org/dc/elements/1.1/
dcmitype	http://purl.org/dc/dcmitype/
dcterms	http://purl.org/dc/terms/
flow	http://www.isac-net.org/std/rdf/meta-relations/1.0/flow#
fp	http://downlode.org/rdf/file-properties#
fugo	http://fugo.sourceforge.net/ontology/FuGO.owl#
owl	http://www.w3.org/2002/07/owl#
protege	http://protege.stanford.edu/plugins/owl/protege#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
vCard	http://www.w3.org/2001/vcard-rdf/3.0#
xsd	http://www.w3.org/2001/XMLSchema#
wot	http://xmlns.com/wot/0.1/
wn	http://xmlns.com/wordnet/1.6/

3.2 Assigning an GUID to FCM data

RDF identifies resources by globally unique identifiers (GUID) in the form of URI (i.e., either location dependent URLs or location independent URN). In the following example we assign an LSID [B29] (a special kind of URN used in examples in this document) to a data file. In general, any kind of URI-based GUID can be used. An example in RDF/XML as well as a simple RDF graph follows (Figure 3).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1" />
</rdf:RDF>
```

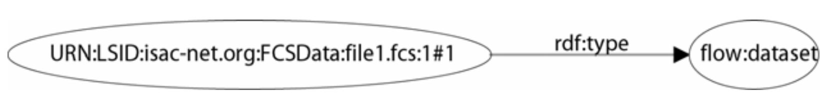

Figure 3 – Assigning an LSID to a dataset

3.3 Using FCM specific classes and properties

3.3.1 Linking a Gating-ML file

A Gating-ML resource may be assigned to an FCS data file by the flow:gated_by property as demonstrated in the following example (Figure 4).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
      <flow:gated_by>
         <flow:gating rdf:about="http://flowcyt.org/rp/gf1.xml#CD3" />
      </flow:gated_by>
  </flow:dataset>
</rdf:RDF>
```

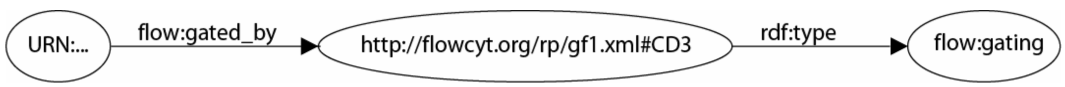

Figure 4 – A Gating-ML file assigned to FCS data

3.3.2 Example of Annotating the Gating-ML file

Dublin Core elements may be used to describe any kind of resource, i.e., also a Gating-ML [B2] file (see 3.4 for more about Dublin Core). The following example (
Figure 5) shows the usage of dc:creator and dcterms:created properties.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
…
    <flow:gated_by>
      <flow:gating rdf:about="http://flowcyt.org/rp/gf1.xml#CD3">
        <dcterms:created>
          <dcterms:W3CDFT>
            <rdf:value>2006-08-12T12:42+08:00</rdf:value>
          </dcterms:W3CDFT>
        </dcterms:created>
        <dc:creator>
          Jack S. Thompson
        </dc:creator>
      </flow:gating>
    </flow:gated_by>
…
</rdf:RDF>
```

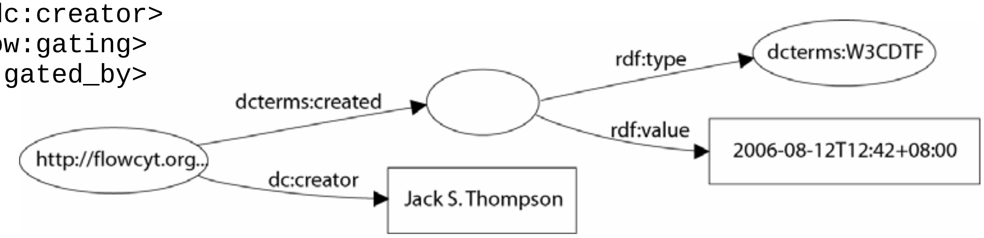

Figure 5 –Annotation of Gating-ML file

3.3.3 Assigning a Transformation-ML file

While analyzing FCM data, various parameter transformations are performed to provide user-friendly visualization and/or to perform statistical analyses and interpret the data. If these data are viewed using a transformed parameter (i.e., on a transformed scale) and any analyses is performed using this visualization, the parameter transformation shall to be connected to the original FCS data file. A Transformation-ML [B3] resource shall be assigned to FCS data by the flow:transformed_by property as demonstrated in the following example (Figure 6).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <flow:transformed_by>
      <flow:transformation rdf:about="http://www.y.z.com/r/f1.xml "/>
    </flow:transformed_by>
  </flow:dataset>
</rdf:RDF>
```

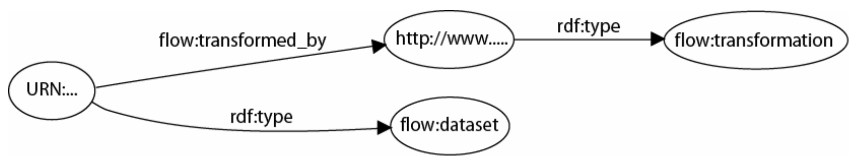

Figure 6 – A Transformation-ML file assigned to FCS data

3.3.4 Assigning a Compensation-ML file

In FCM, the emission spectral overlap of fluorescent labels makes it necessary to correct detected signals before using the values as a basis for any other analyses. If compensation was performed a compensation description file (Compensation-ML file) should be connected to the original FCS data. A Compensation-ML [B4] resource shall be assigned to FCS data by the flow:compensated_by property as demonstrated in the following example (Figure 7).

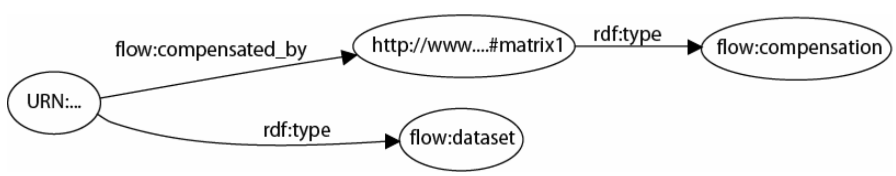

Figure 7 – A Compensation-ML file assigned to FCS data

3.4 Dublin Core properties for FCS data

3.4.1 Overview

The Dublin Core metadata element set ([B13], [B14], [B15]) is a standard for cross-domain information resource description. In the following section we will demonstrate how Dublin Core may be utilized for FCM experiment annotation.

3.4.2 Dublin Core – Title

Typically, title is a name by which a resource is formally known. The data may be named or the name of the experiment may be provided (Figure 8).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:title>Data from the experiment:
              Confusing effects of platelets on flow-cytometric
              analysis experiments using blood-derived cells
    </dc:title>
  </flow:dataset>
</rdf:RDF>
```

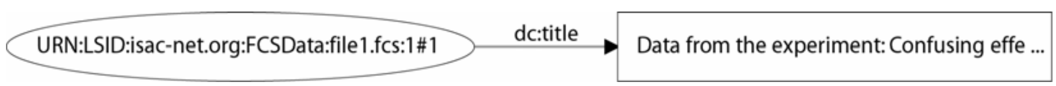

Figure 8 – Assigning a dc:title to a flow dataset

3.4.3 Dublin Core – Creator

A creator is an entity primarily responsible for making the content of the resource. Examples of creator include a person, an organization, or a service. Typically, the name of a creator is used to indicate the entity, e.g., Figure 9.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1" >
      <dc:creator>Karel May BC Cancer Research Centre</dc:creator>
      <dc:creator>Dr. Alexander Hyde</dc:creator>
  </flow:dataset>
</rdf:RDF>
```

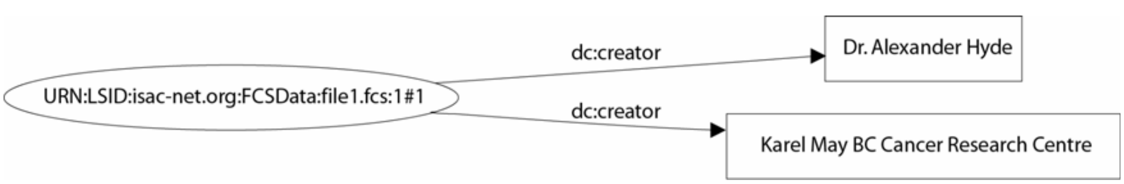

Figure 9 – Assigning a dc:creator to a flow dataset

There are several possibilities how to use vCard [B21] to further structure the identification of an entity. The vCard:N structure may be used to structuralize a name as follows (Figure 10):

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:creator>
      <rdf:Description>
        <vCard:N rdf:parseType="Resource">
          <vCard:Family>Shneiderman</vCard:Family>
          <vCard:Given>Ben</vCard:Given>
          <vCard:Prefix>Dr.</vCard:Prefix>
          <vCard:Suffix>Ph.D.</vCard:Suffix>
        </vCard:N>
      </rdf:Description>
    </dc:creator>
  </flow:dataset>
</rdf:RDF>
```

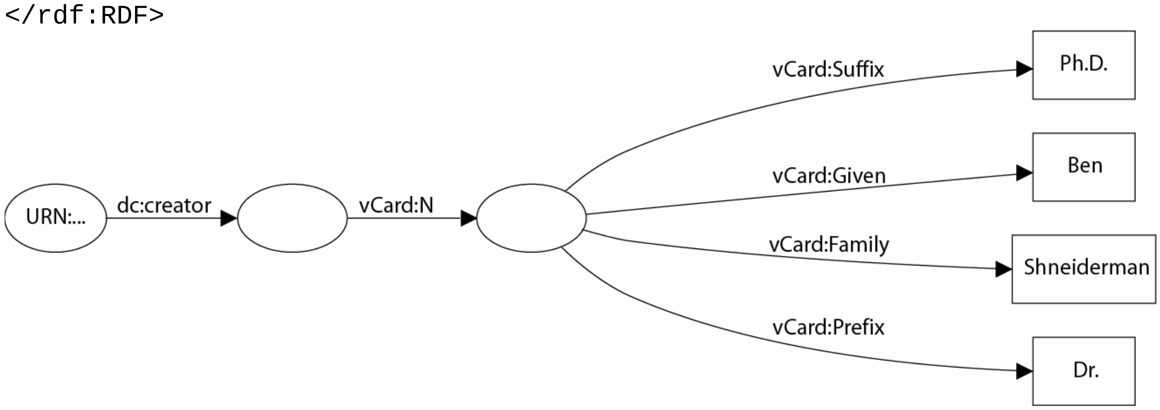

Figure 10 – Using the vCard:N structure to structuralize a name

An address of an entity may be provided either as a vCard:LABEL (Figure 11) or filling in the vCard:ADR structure (Figure 12).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:creator>
      <rdf:Description>
        <vCard:LABEL rdf:parseType="Resource">
          <rdf:type rdf:resource="http://www.w3.org/2001/vcard-
rdf/3.0#postal"/>
          <rdf:value rdf:parseType="Literal">
                  3260 Production Way, Burnaby, BC, V5A 4W4,
                  Canada
          </rdf:value>
       </vCard:LABEL>
```



```
      </rdf:Description>
    </dc:creator>
  </flow:dataset>
</rdf:RDF>
```

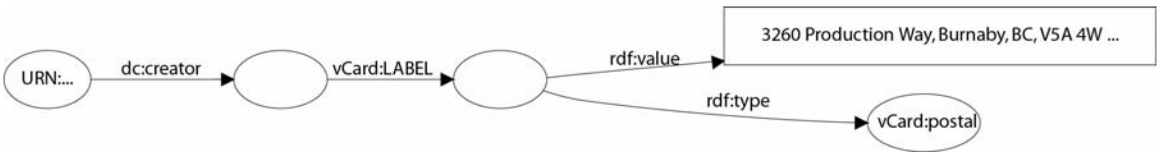

Figure 11 – Using vCard:LABEL/vCard:postal to provide a postal address

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:creator>
      <rdf:Description>
        <vCard:ADR rdf:parseType="Resource">
          <vCard:Street>3260 Production Way</vCard:Street>
          <vCard:Locality>Burnaby</vCard:Locality>
          <vCard:Region>BC</vCard:Region>
          <vCard:Pcode>V5A 4W4</vCard:Pcode>
          <vCard:Country>Canada</vCard:Country>
        </vCard:ADR>
      </rdf:Description>
    </dc:creator>
  </flow:dataset>
</rdf:RDF>
```

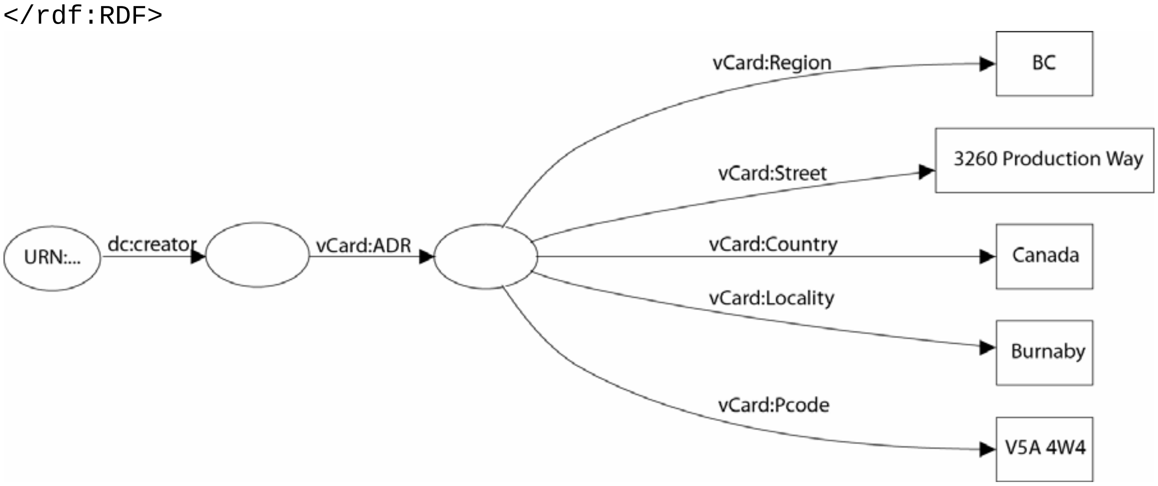

Figure 12 – Structuring an address using the vCard:ADR structure

The Dublin Core creator property may also be used to specify an instrument as the resource that was used to created FCS data. In the following example (Figure 13) a resource labeled *FACSCalibur-5;12th floor* has been used to create the dataset.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
```

```
    <dc:creator>
      <rdf:Description>
        <rdfs:label>FACSCalibur-5; 12th floor</rdfs:label>
      </rdf:Description>
    </dc:creator>
  </flow:dataset>
</rdf:RDF>
```

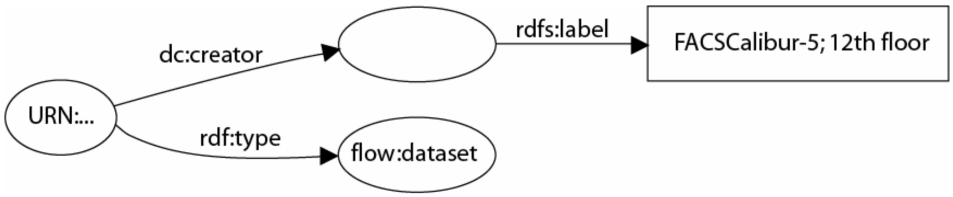

Figure 13 – Referencing the instrument used to generate the data

3.4.4 Dublin Core – Subject

Subject represents the topic of the content of the resource. Typically, subject will be expressed as keywords, key phrases or classification codes that describe the topic of the resource.

There is the possibility of providing comma-separated keywords; however, the rdf:Bag container provides a better structure for this purposes (Figure 14).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
      <dc:subject>
        <rdf:Bag>
          <rdf:li>Blood Platelets</rdf:li>
          <rdf:li>Flow Cytometry</rdf:li>
          <rdf:li>Immunophenotyping</rdf:li>
        </rdf:Bag>
      </dc:subject>
  </flow:dataset>
</rdf:RDF>
```

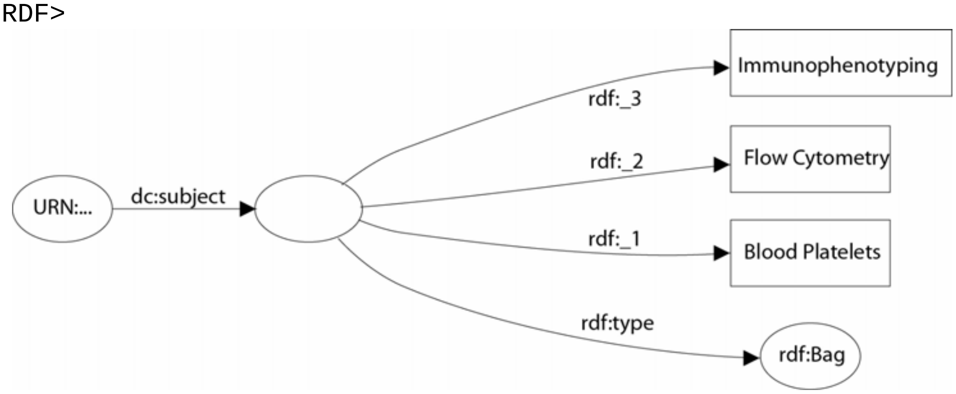

Figure 14 – Subject keywords in a bag container

Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme. Dublin Core defines the following *Element Encoding Schemes* for encoding the subject: DDC, LCC, LCSH, MeSH, and UDC. The following example demonstrates how keywords may be encoded using the MeSH thesaurus (Figure 15).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:subject>
      <rdf:Bag>
        <rdf:li>
          <dcterms:MESH>
            <rdf:value>A15.145.229.188</rdf:value>
            <rdfs:label>Blood Platelets</rdfs:label>
          </dcterms:MESH>
        </rdf:li>
        <rdf:li>
          <dcterms:MESH>
            <rdf:value>E05.200.500.386.350</rdf:value>
            <rdfs:label>Flow Cytometry</rdfs:label>
          </dcterms:MESH>
        </rdf:li>
        <rdf:li>
          <dcterms:MESH>
            <rdf:value>E05.478.600</rdf:value>
            <rdfs:label>Immunophenotyping</rdfs:label>
          </dcterms:MESH>
        </rdf:li>
      </rdf:Bag>
    </dc:subject>
  </flow:dataset>
</rdf:RDF>
```

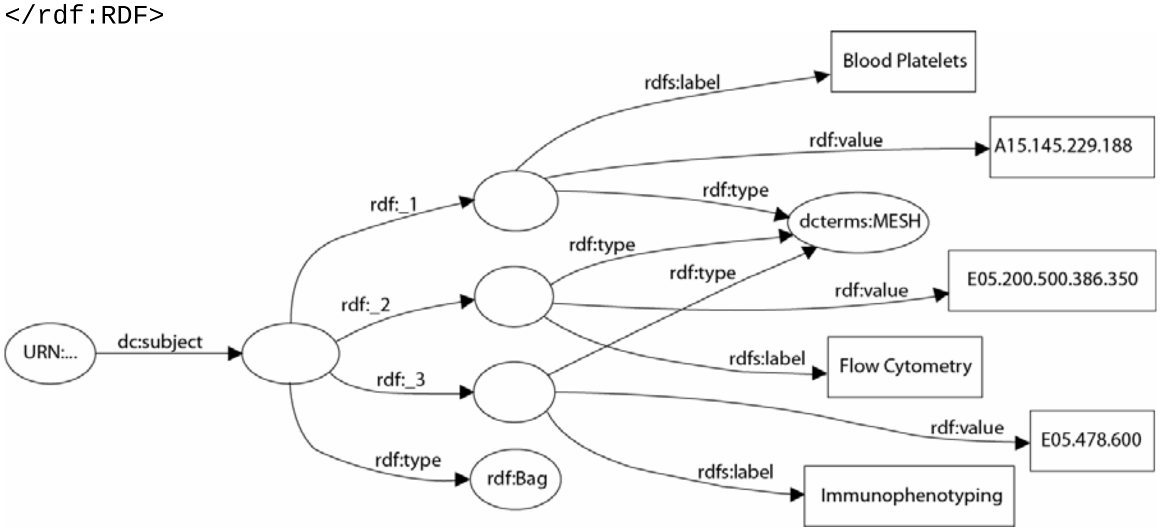

Figure 15 – Selecting keywords from the MeSH thesaurus

3.4.5 Dublin Core – Description

A Dublin Core description element summarizes an account of the content of the resource. Examples of description include, but are not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content. Dublin code defines two *Element Refinements* for a description: *table of contents* and *abstract*. The following example (Figure 16) demonstrate how to combine the dc:description property with one of its qualified refinement.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dcterms:abstract xml:lang="en">
      Flow-cytometric analysis of peripheral blood leukocytes is
      commonplace; however, platelet contamination is typically …
    </dcterms:abstract>
    <dc:description xml:lang="en">Experiment 3</dc:description>
  </flow:dataset>
</rdf:RDF>
```

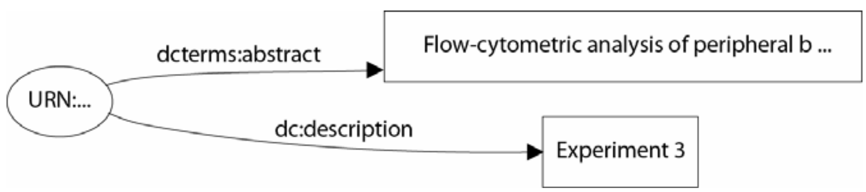

Figure 16 – Dublin Core description and its refinement

3.4.6 Dublin Core – Contributors

A contributor is an entity responsible for making contributions to the content of the resource. Specifically, enumeration of entities (persons) contributing to the particular experiment may be provided.

Typically, the name of a contributor may be used to indicate the entity (Figure 17). Alternatives include using a vCard-defined structure (as described in section 3.4.3) or referencing a particular resource or ontology class.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:contributor xml:lang="en-US">Princess Margaret General Hospital
    </dc:contributor>
    <dc:contributor xml:lang="en-US">Dr. Helen Smith
    </dc:contributor>
  </flow:dataset>
</rdf:RDF>
```

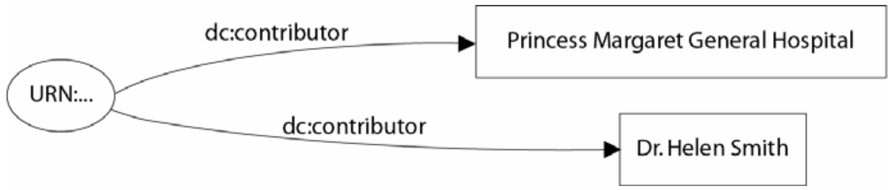

Figure 17 – Assigning contributors to a dataset

3.4.7 Dublin Core – Dates

A date element is used to provide a date of an event in the lifecycle of the resource. Typically, a date will be associated with the creation or availability of the resource. Dublin code defines several *Element Refinements* for a data: *created*, *valid*, *available*, *issued*, and *modified*. Recommended best practice for encoding the date value is defined in ISO 8601 [B30] (e.g., Figure 18). Alternatively, one of the two *Element Encoding Schemes* defined for the date element may be used. The *DCMI Period* encoding scheme represents a specification of the limits of a time interval and the *W3C-DTF* encoding scheme (e.g., Figure 19) stands for W3C encoding rules for dates and times (which are based ISO 8601).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dcterms:created
       rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
       2006-08-04T16:20:30+08:00
    </dcterms:created>
  </flow:dataset>
</rdf:RDF>
```

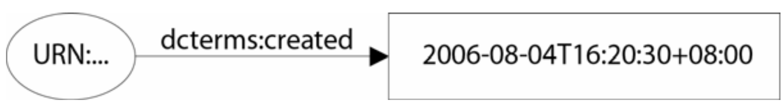

Figure 18 – DC date refined to dcterms:created; expressed as a literal string

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dcterms:created>
      <dcterms:W3CDTF>
        <rdf:value>2006-08-04T16:20:30+08:00</rdf:value>
      </dcterms:W3CDTF>
    </dcterms:created>
  </flow:dataset>
</rdf:RDF>
```


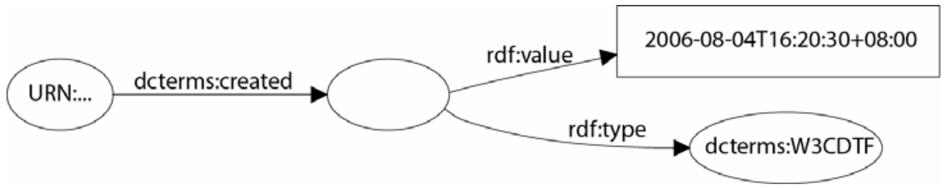

Figure 19 –A resource with an explicit W3CDFT encoding scheme

3.4.8 Dublin Core – Type and Format

The type of a resource expresses the nature or genre of the content of the resource. However, to describe the physical or digital manifestation of the resource, the format element shall be used, which is the case that better correspond to annotation of FCS data. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Recommended best practice is to select a value from a controlled vocabulary, e.g., the list of Internet Media Types (IMT) (also called MIME Media Types, RFC2045 [B31], RFC2046 [B32]).

Similar to expressing dates (section 3.4.7), formats may be expressed as literal string or as a qualified Dublin Core element having an explicit encoding schema provided (Figure 20).

The procedure for registering MIME Types (RFC4288 [B33], RFC4289 [B34]) is not too complicated. We suggest that ISAC officially register the FCS file format as the “application/fcsdata” MIME type through the registration process of IANA. Having the FCS data file format registered as a MIME type makes it a well known format so that for example email or HTTP servers would know how to handle FCS files. The following example assumes such a registration being performed.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
      <dcterms:IMT>
        <rdf:value>application/fcsdata</rdf:value>
      </dcterms:IMT>
  </flow:dataset>
</rdf:RDF>
```

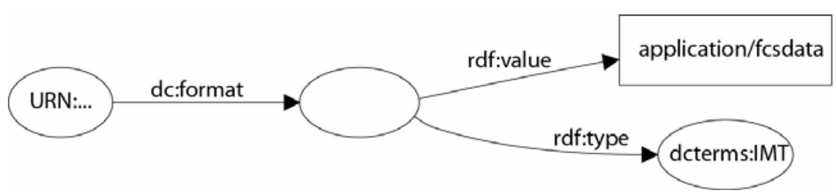

Figure 20 – Format of a resource with an explicit IMT encoding scheme

3.4.9 Dublin Core – Identifier

A resource identifier represents an unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI), the Digital Object Identifier (DOI), the International Standard Book Number (ISBN), etc.

As we have assigned an LSID to the dataset (section 3.2) we do not necessary need to provide another unique identifier. Some data may have more unique resource identifiers. On the following example (Figure 21) we demonstrate how to use the rdf:Alt container (intended to specify alternative resources) to indicate an alternative URL where the FCS data may be obtained.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:identifier>
      <rdf:Alt><rdf:li>
          <dcterms:URI><rdf:value>
             URN:LSID:isac-net.org:FCSData:file1.fcs:1#1
          </rdf:value></dcterms:URI>
        </rdf:li><rdf:li>
          <dcterms:URI><rdf:value>
             http://www.flowcyt.org/files/fcs/file1.fsc#1</rdf:value>
          </dcterms:URI></rdf:li>
      </rdf:Alt>
    </dc:identifier>
  </flow:dataset>
</rdf:RDF>
```

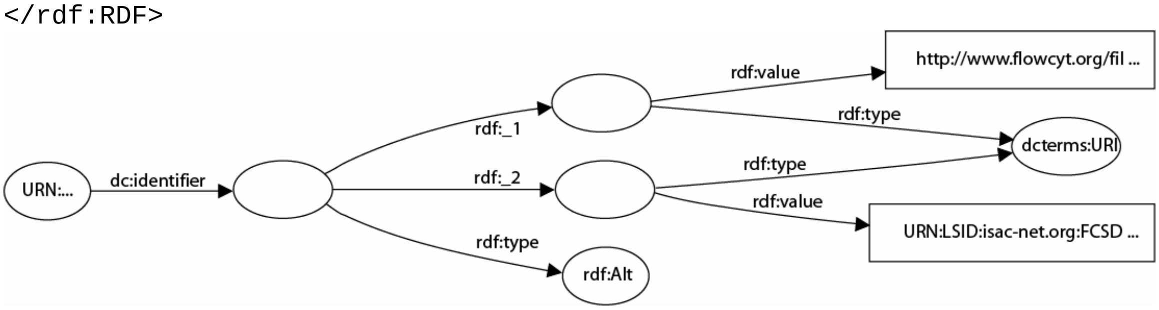

Figure 21 – Using the rdf:Alt container to specify alternative identifiers

3.4.10 Dublin Core – Source

A source represents a reference to a resource from which the present resource is derived. The present resource may be derived from the source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

In FCM experiment annotation, this is the right place where to connect biological source description including genus, species and subspecies, strain, age, gender, cell line, cell culturing / growing description, treatment description, stimulation description, extract preparation description, probes and probe incorporation method description, separation technique description, antibody staining procedures description, wash steps description, etc. The use of external ontologies (e.g., OBI [B24]) is encouraged. The following example (Figure 22) presents a basic classification of the source used to generate FCS data.


```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF …>
  <flow:dataset
    rdf:about="URN:LSID:isac-net.org:FCSData:file1.fcs:1#1">
    <dc:source>
      <rdf:Description>
        <rdf:value>Snomed CT ID: 52501007</rdf:value>
        <rdfs:label>Leukocyte</rdfs:label>
      </rdf:Description>
    </dc:source>
  </flow:dataset>
</rdf:RDF>
```

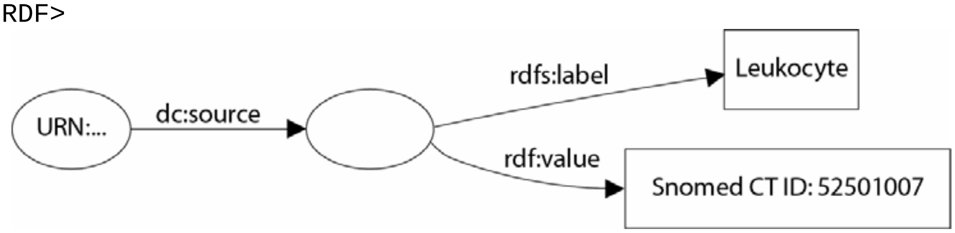

Figure 22 – Classification of the biological source used to obtain FCS data

Annex A

Bibliography

[B1] Seamer L.C., Bagwell C.B., Barden L., Redelman D., Salzman G.C., Wood J.C.S., Murphy R.F.: "Proposed New Data File Standard for Flow Cytometry, Version FCS 3.0", Cytometry 28 (1997), Wiley-Liss, Inc., pp. 118-122, http://murphylab.web.cmu.edu/publications/64-seamer1997.pdf.

[B2] Bioinformatics Standards for Flow Cytometry Consortium: Gating-ML: Draft Standard for Gating Description in Flow Cytometry; http://sourceforge.net/projects/flowcyt/.

[B3] Bioinformatics Standards for Flow Cytometry Consortium: Transformation-ML: Draft Standard for Transformation Description in Flow Cytometry; http://sourceforge.net/projects/flowcyt/.

[B4] Bioinformatics Standards for Flow Cytometry Consortium: Compensation-ML: Draft Standard for Compensation Description in Flow Cytometry; http://sourceforge.net/projects/flowcyt/.

[B5] Wang X., Gorlitsky R., Almeida J.S. (2005). From XML to RDF: how semantic web technologies will change the design of 'omic' standards. Nat Biotechnol 23, 1099-1103.

[B6] Quackenbush J. (2006). Standardizing the standards. Mol Syst Biol 2, 2006 0010.

[B7] W3C Recommendation: RDF Primer, http://www.w3.org/TR/rdf-primer/.

[B8] W3C Recommendation: RDF/XML Syntax Specification (Revised), http://www.w3.org/TR/rdf-syntax-grammar/.

[B9] W3C Recommendation: RDF Vocabulary Description Language 1.0: RDF Schema, http://www.w3.org/TR/rdf-schema/.

[B10] W3C Recommendation: Resource Description Framework (RDF): Concepts and Abstract Syntax, http://www.w3.org/TR/rdf-concepts/.

[B11] W3C Recommendation: RDF Semantics, http://www.w3.org/TR/rdf-mt/.

[B12] W3C Recommendation, Extensible Markup Language (XML) 1.0 (Third Edition), 04 February 2004, http://www.w3.org/TR/REC-xml/.

[B13] Dublin Core Metadata Initiative: DCMI Recommendation, Dublin Core Metadata Element Set, Version 1.1: Reference Description, http://dublincore.org/documents/dces/.

[B14] Dublin Core Metadata Initiative: DCMI Recommendation, DCMI Metadata Terms, http://dublincore.org/documents/dcmi-terms/.

[B15] Dublin Core Metadata Initiative: DCMI Recommendation, DCMI Type Vocabulary, http://dublincore.org/documents/dcmi-type-vocabulary/.

[B16] W3C Recommendation: OWL Web Ontology Language – Overview, http://www.w3.org/TR/owl-features/.

[B17] W3C Recommendation: OWL Web Ontology Language – Guide, http://www.w3.org/TR/owl-guide/.

[B18] W3C Recommendation: OWL Web Ontology Language – Reference, http://www.w3.org/TR/owl-ref/.

[B19] W3C Recommendation: OWL Web Ontology Language – Semantics and Abstract Syntax, http://www.w3.org/TR/owl-semantics/.

[B20] W3C (submission): DAML+OIL Reference Description, http://www.w3.org/TR/daml+oil-reference.

[B21] W3C (submission): Representing vCard Objects in RDF/XML, http://www.w3.org/TR/vcard-rdf.

[B22] Miller E.: An Introduction to the Resource Description Framework, D-Lib Magazine, May 1998, ISSN 1082-9873.

[B23] Shelley Powers: Practical RDF. O'Reilly, 2003, ISBN: 0-596-00263-7. (Online to read at O'Reilly's Safari Bookshelf: http://safari.oreilly.com/0596002637).

[B24] Whetzel P.L., Brinkman R.R., Causton H.C., Fan L., Field D., Fostel J., et al. (2006). Development of FuGO: An Ontology for Functional Genomics Investigations. Omics 10, 199-204.

[B25] R.C. Leif CytometryML, a data standard, which has been designed to interface with other standards (preprint) to be in Manipulation and Analysis of Biomolecules, Cells, and Tissues V, D. Farkas, R. C. Leif, and D. V. Nicolau, Editors, SPIE Proceeding Vol. 6441 (2007).

[B26] Health Level Seven: http://www.hl7.org/.

[B27] DICOM: Digital Communication in Medicine, http://medical.nema.org/

[B28] Bioinformatics Standards for Flow Cytometry Consortium: MIFlowCyt - Minimum Information about a Flow Cytometry Experiment; http://www.isac-net.org/content/view/594/46/, http://flowcyt.sourceforge.net/miflowcyt/.

[B29] Clark T., Martin S., Liefeld T.: Globally distributed object identification for biological knowledgebases. Brief Bioinform. 2004; 5:59-70.

[B30] International Organization for Standardization (ISO): Numeric representation of Dates and Time, http://www.iso.org/iso/en/prods-services/popstds/datesandtime.html.

[B31] RFC2045: Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies, http://www.isi.edu/in-notes/rfc2045.txt.

[B32] RFC2046: Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types, http://www.isi.edu/in-notes/rfc2046.txt.

[B33] RFC4288: Media Type Specifications and Registration Procedures, http://www.isi.edu/in-notes/rfc4288.txt.

[B34] RFC4289: Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures, http://www.rfc-editor.org/rfc/rfc4289.txt.

