

# The Evolution of Flow Cytometry Standard, FCS3.0, into a DICOM Compatible Format

Robert C. Leif and Suzanne B. Leif

Ada Med, a Division of Newport Instruments, 5648 Toyon Road, San Diego, CA 92115; E-mail rleif@rleif.com

## ABSTRACT

The International Society for Analytical Cytology, ISAC, has developed a Flow Cytometry Standard (FCS) to permit data interchange. ISAC will soon replace Flow Cytometry Standard 2.0 (FCS2.0) with FCS3.0. Unfortunately, the proposed FCS3.0 is still fraught with problems, which are of sufficient magnitude as to warrant its early replacement. The most reasonable replacement is as a supplement to the Digital Imaging and Communications in Medicine, DICOM 3.0, standard. The recent digital microscopy extension of DICOM can be extended and modified to include flow cytometry data. DICOM includes: image graphics objects, specifications for describing: studies, reports, the acquisition of the data and the individuals involved, physician, patient, etc. Storing the present FCS data in a database, which has already been accomplished with the QC Tracker software, will facilitate the transition of FCS to DICOM.

KEYWORDS: FCS, Ada, Flow Cytometry, DICOM, Software, Database, Standards

### 1. INTRODUCTION

The International Society for Analytical Cytology, ISAC, has developed a Flow Cytometry Standard (FCS) to permit data interchange. ISAC will soon replace Flow Cytometry Standard 2.0<sup>5</sup> (FCS2.0) with FCS3.0<sup>23</sup>. Unfortunately, the proposed FCS3.0 is still fraught with problems, which are of sufficient magnitude as to warrant its early replacement. We have previously suggested<sup>14</sup> a two step process: 1. Removal of the obvious problems from FCS3.0; and 2. Commence building a Digital Imaging and Communications in Medicine, DICOM<sup>3</sup> standard for Flow Cytometry employing the domain knowledge contained in FCS3.0. The FCS image standard will be replaced by the DICOM Microscopic Image Standard<sup>1</sup>.

It must be emphasized that the knowledge required to determine what information should be contained in a standard is entirely different from how to structure a standard. The selection of the data items contained in ISAC FCS3.0 showed excellent domain knowledge. However, the present design of FCS3.0 shows no evidence of expertise in the creation of software standards. Since the FCS is a detailed specification for software which will be included in a medical device, the products that employ it fall under the jurisdiction of the US FDA and other national and multinational regulatory agencies, such as the European Economic Community.

### 2. FDA Guidance for Software

The latest "ODE Guidance for the Content of Premarket Submission for Medical Devices Containing Software Draft Document" from the FDA CDRH Home Page<sup>6</sup> provides guidance on the regulatory review of premarket medical device software submissions. It will replace the "Reviewer Guidance for Computer Controlled Medical Devices Undergoing 510(k) Review" which was issued in 1991. "This guidance discusses the key elements reviewers look for in a premarket medical device software submission; thereby providing a common baseline from which both manufacturers and scientific reviewers can operate."

The items relevant to an Analytical Cytology standard of Section C.1 Review Checklist are given below. The numbers in the ODE Guidance checklist have been prefaced with FDA. This checklist is intended to be used by reviewers and manufacturers to determine if the software documentation is consistent with the device.

FDA 1. Is the hazard analysis complete, and is it consistent with the device and intended use and level of concern determination?

FDA 2. Are appropriate safety requirements incorporated in the device which address the hazards identified? Have they been appropriately evaluated?

FDA 3. Are the software lifecycle processes and methodologies discussed appropriate for the safety issues and level of concern of the device?

Does the submission discuss how the hazard analysis was performed?

- a. If this is a “lower” level of concern, is it adequate for addressing the hazards that were identified? Are the lifecycle processes, risk control measures, and quality assurance activities reasonable for the device?
- b. If this is a “higher” level of concern, is it adequate for addressing the hazards that were identified? Are the lifecycle processes, quality assurance measures, and risk management activities appropriate in addressing the safety issues for the device?

FDA 4. Are verification and validation activities performed prior to formal release?

FDA 5. Is there adequate documentation generated to assure traceability?

FDA 7. Are the software and system requirements consistent with device claims, labeling, and intended use?

FDA 10. Are security measures consistent for the level of security required for data or device access for preventing data loss?

FDA 17. Do results of tests and analyses demonstrate conformance to all requirements, including safety?

FDA 19. When data are transmitted between devices or device elements, are means provided to ensure there are no transmission errors? How is this verified?

FDA Items 1 and 2 require a hazard analysis; items 5, 7, and 17 concern requirements.

### 3. PRELIMINARY HAZARD ANALYSIS

Whatever standard is finally adopted for Flow Cytometry data interchange will require a hazard analysis and appropriate safeguards built into the standard to prevent hazards from occurring. Since no hazard analysis has previously been performed for any version of the FCS, a draft analysis is given below. FCS3.0 will be used to transfer flow cytometry data from one program to another on the same computer and from one computer file system to another on the same or, most often, on different computers. The data can be moved between computers either by physically transporting movable media: diskettes, tapes, CD-ROM, etc. or by electronic communication, principally over the Internet. The hazards below are listed from higher to lower severity.

Hazard 1. Undetected transmission errors.

Note: Since these errors could effect the treatment of the patient, this is a high priority hazard.

Hazard 2. Erroneous field identification and/or interpretation.

Hazard 3. Erroneous numeric interpretation.

Note: This could be due of interpreting an incorrect byte order (endian).

Hazard 4. The original data logging system can create a file that can not be read and both the primary data and sample are lost.

Note: Reasonable testing of the software would minimize this possibility. This error could either delay or prevent the use of the flow cytometry data.

Hazard 5. The data can not be read at the receiving end.

Note: This can often be fixed by a retransmission. This error could delay the use of the flow cytometry data.

### 4. REQUIREMENTS

In the following comments, when quoting the FCS3.0 text, quotation marks, “FCS3.0” will be used. The only documented requirements in FCS3.0<sup>23</sup> are: “The principal goal of the Standard is to provide a uniform file format allowing files created by one type of acquisition hardware and software to be analyzed by another type. The proposed FCS3.0 standard maintains backwards compatibility with previous versions by retaining the basic FCS file structure.”

The stated goal of providing “a uniform file format allowing files created by one type of acquisition hardware and software to be analyzed by another type” has not been achieved, since the FCS3.0 specification still permits diversity rather than enforcing uniformity.

Since the above requirements are insufficient for creating a flow cytometry standard, a more complete set of requirements is proposed below. The membership of ISAC and other parties interested in Flow Cytometry are encouraged to comment on these requirements. Test plans require an explicit requirements document. Good specifications facilitate validation suite development.

Requirements that we would add are:

Req.1 FCS should be defined as a manufacturer independent data transfer format.

Note: The data generated by a flow cytometer belongs to the institution where it was generated. In many cases, the FCS format is to be applied to information that is transferred from one computer to another or from one program to another. The FCS files are a means to store an image of data organized to be transferred. The software vendor or developer is not required to keep flow cytometry data in FCS format. FCS format is very inefficient for many applications. All of the data describing the flow cytometry run except, possibly, the actual measurements on the cells is far more accessible if stored in a database<sup>15</sup>. If the database stores 8, 16, and/or 32 bit integers, it also is practical to store the list mode data in a database.

Req.2 Undetectable errors in writing, transmitting, reading, and receiving FCS data should be minimized.

Note: Very often as the allowable error rate is reduced, the amount of testing required to determine the actual error rate increases. This relationship need not be linear. Therefore, some reasonable error rate needs to be specified. At the very least, the FCS software developers should publish their measured or best estimate of the error rate.

Req.3 Conformance to the FCS should be testable.

Req.4 The developer (vendor) of software who produces FCS software should, at the very least, be required to submit a statement that the product conforms to FCS.

Req.5 The FCS data format should be optimized for reliable transmission.

Req.6 International standards should be used for the design of and referenced by the FCS.

Note: A large portion of the users of flow cytometers are not residents of the United States. Therefore, the International Organization for Standardization, ISO, has the highest precedence.

Req.7 FCS should be programming language independent.

Req.8 The capacity to interact and interface with other clinical data management systems should be maximized.

Req.9 FCS should be compatible with or based on existing, relevant standards.

Req.10 Human reading of the contents of a FCS data transmission or file should not result in errors or breaches of patient privacy.

Req.11 The data acquisition configuration information included in the standard should be sufficient to permit an experimenter to set up his/her instrument to repeat the acquisition of the data.

Note: At least initially, this may require inclusion of textual comments.

Req.12 The FCS ANALYSIS section should permit an experimenter to understand how the data was analyzed.

Note: The gates and their operations should be available. The exact descriptions of the algorithms employed for producing the results in the Analysis section may be proprietary. At least initially, this may require inclusion of textual comments.

Req.13 The data included in the standard should be sufficient to permit an experimenter to prepare the sample(s).

Note: The DICOM 3 Microscopic Imaging standard includes this capability.

Several goals are specified below. Although these goals are not true software requirements and can not be absolutely predicted or measured, they are still desirable features.

Goal 1 . The costs of implementing and maintaining the standard should be minimized.

Goal 2 . Cytometry experts should exercise as much control as possible concerning the specification of what information is contained in a Flow Cytometry data transfer format.

Note: The specification of what information is contained is entirely separate from the means to convey this information, such as the format.

### 5. Comments on FCS3.0

At this stage of a discussion, agreement should be sought on the hazards and requirements listed above. If these are agreeable, then the criticisms below of FCS3.0 can only be refuted by demonstrating that they are inconsistent with the requirements and hazards listed above.

The recursive structure of FCS is a significant problem in terms of functionality and testing. Many of the readers of this document and/or the users of FCS have had the frustration of trying to download a large file on the Internet. Often, there is a fatal communication disruption. For instance, the probability of success is much higher for downloading 12 one megabyte files than one twelve megabyte file (Req.5).

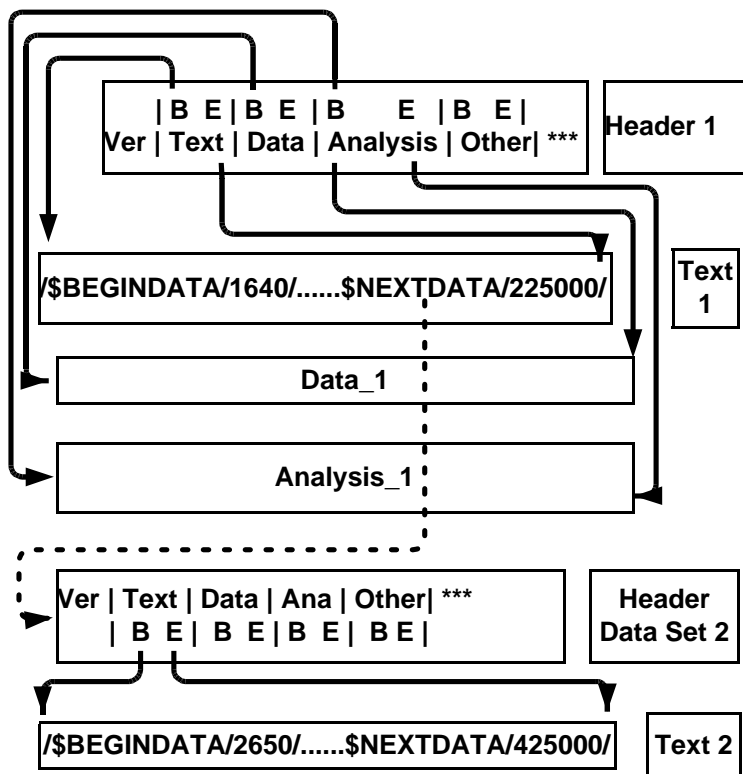


Figure 1. FCS 3.0 Structure. The letters 'B' and 'E' abbreviate respectively beginning and end byte offsets. In order to facilitate readability, only the first letters of the segments have been capitalized. This is a recursive structure. HEADER 1 points to TEXT 1, which in turn (dashed arrow) points to the HEADER in Data Set 2, which points to TEXT 2. In order to accommodate data sets larger than 99,999,999 bytes, a supplemental TEXT section can be included. The location of the supplemental TEXT section is specified in the primary TEXT section.

Figure 1 shows the layout of an FCS3.0 file. Table 1 shows the location of the beginning and ending byte offsets of an FCS3.0 data set. Since there is no limit on the number of TEXT sections that will (except for the last) point to a HEADER which both starts a subsequent data set and will in turn point to a TEXT section, complete testing is impossible (Req.3). A simple interim solution is to limit the number of list mode data files to one and the number of HEADER and TEXT sections to five for each FCS3.0 data file. Large files containing multiple data sets can be broken up into smaller packets for transmission. Each of these packets can include a few key fields such as date, time, and specimen number. These fields should provide enough information to reassemble the file. If very large list mode files are to be sent via the Internet or other electronic means, there should be a standard for dividing them into blocks prior to transmission and reassembling them after transmission.

When international standards are available from ISO, these should be employed (Req.6) rather than the American standards such as those from ANSI. Therefore, the character sets must be those specified by ISO<sup>12</sup>. The eight bit character set, ISO 8859-1 (Latin 1), contains the special characters required for many European languages<sup>13</sup>. This will, at least, permit European Analytical Cytologists to store their own and their patients' names correctly. FCS3.0 presently does not permit many Europeans to enter their names without using 16 bit characters. The present 7 bit character set referred to as ASCII is ISO 646. Characters 0 to 127 of Latin 1 are identical with US-ASCII. The 16 bit character set, ISO 10646, is suitable for many Oriental languages.

Table 1. Example of an FCS 3.0 HEADER Lay Out

| Version | TEXT  |       | DATA  |         | ANALYSIS |         | OTHER   |         |
|---------|-------|-------|-------|---------|----------|---------|---------|---------|
|         | Begin | End   | Begin | End     | Begin    | End     | Begin   | End     |
| FCS3.0  | 256   | 1,451 | 1,640 | 202,687 | 203,000  | 207,538 | 210,438 | 220,512 |

The present standard uses ==, is full of C constructs, and refers to C code. This is totally contrary to programming language independence (Req.7). The C programming language and its descendants are not suitable for medical<sup>16</sup> or other safety critical devices. For instance, N. G. Leveson<sup>17</sup> states, "Not only must a language be simple, but it must encourage the production of simple and understandable programs. Although careful experimental results are limited, some programming language features have been found to be particularly prone to error --among them pointers, control transfers of various kinds, defaults and implicit type conversions, and global variables. Overloading variable names so that they are not unique and do not have a single purpose is also dangerous. On the other hand, the use of languages with static type checking and the use of guarded commands (ensuring that all possible conditions are accounted for in conditional statements and that each branch is fully specified as to the conditions under which it is taken) seem to help eliminate potential programming errors, Some of the most frequently used languages (such as C) are also those that, according to what is known about language design, are the most error prone." Riehle<sup>21</sup> makes a very strong case for Ada where safety is required.

Citing Numerical Recipes in C is insufficient. C code and nomenclature are obscure, hazard prone methods for describing algorithms. Algorithms are better described in plain English, mathematical notation, pseudo-code, and/or flow charts.

A simple example in FCS of a C construct, which is confusing and can lead to errors, is the first byte of the header is numbered zero. Languages which are descendants of Pascal, including Ada, stress readability. Ada can start arrays with the first element being 1 or any other integer.

Another serious defect is the use of sentinels. FCS Section 3.2.1 states, "The TEXT segments (primary and supplemental) contain a series of ASCII encoded keyword-value pairs that describe various aspects of the data set. For example, \$TOT/5000/ is a keyword-value pair indicating that the total number of events in the file is 5,000. \$TOT is the keyword and 5000 is the value. The '\$' character flags this keyword as a standard FCS keyword. In this example, the '/' is the delimiter character." Although, the use of the character '/' as a sentinel for both the beginning and end of a value can cause errors (see below), the hazard associated with these potential errors is overshadowed by allowing the user to select the sentinel character. The sentinel and other special characters should be specified by the standard.

Programming languages, by and large, employ two ways to describe the size of an object. C and its derivatives employ a sentinel character 'nul' to end strings. This provides great flexibility. It also decreases efficiency because the program has to check for the sentinel after reading each character. If the data is corrupted, the string or other similar construct can have extra characters added. Ada and the rest of the Pascal family of languages include the length of the string in the data. If there are either more or less characters than are specified, in Ada the exception constraint error is raised. This serves as a very effective means to warn both the user and the system that something wrong has occurred, Req.2. DICOM 3 specifies the size of its objects. Most if not all of the readers of this document are knowledgeable about frame shift mutations where after 3 mutations, a functioning protein is often produced. This could happen with transmission errors.

If there is to be long term use of FCS, the possibility of error would be decreased by making the starting character be different from the ending character. Care should be taken to ensure that it is a character in the lower ISO 8859-1 character set so that it will not interfere with foreign languages. This is the range of 0 to 127 decimal or 16#00# (nul) to 16#7F# (del). Since the semicolon Latin1(59) serves as a terminator for C, C++, Pascal, Ada, and other languages, it is an obvious choice for being the terminator. The delimiter should not be user defined. It should be specified. The user variable start character should also be specified.

Since the FCS3.0 is defining a standard for a medical device and the authors of FCS3.0 have defined conformance, they should also propose a rigorous method to validate manufactures' code (Req.3). The suggestion below

of a new standard based on DICOM removes the responsibility concerning validation from ISAC or any other single scientific or medical organization (Goal 1).

The types employed for representing the data could be reduced by eliminating ASCII digits. The inclusion of ASCII digits as a data type causes two meanings, bit and byte, for the B in PnB. This ambiguity increases the complexity of the design and code.

The example given, "\$GnF/string/ \$G2F/520LP/" does, at least, define a long pass filter. However, if filters are defined as the objects that are put in filter holders then, at least, the following terms could be used in the first element in a record or equivalent with the enumerated types: Long\_Pass, Short\_Pass, Transmission\_Band, Blocking\_Band, Polarizer, Half\_Wave\_Plate, Quarter\_Wave\_Plate, Dichroic\_Mirror, and Grating. There is nothing in FCS to describe a grating or prism. The maximum or other specific wavelength, band pass and efficiency could also be included as fields. The part number and manufacturer should be seriously considered. The meaning of "\$GnP Percent of emitted light collected by gating parameter n." is not clear. Is it the product of the transmissions?

Byteord - byte order for data acquisition should be specified and not be an option. Presently, there are 24 possible (4 factorial) byte order formats. Support of all of these including testing (Req.3) is not warranted (Goal 1). DICOM Part 5, Data Structures and Encoding Section 7.3 states, "Big Endian vs. Little Endian Byte Ordering. Another component of the encoding of a Data Set that shall be agreed upon by communicating Application Entities is the Byte Ordering. Little Endian byte ordering is defined as follows: In a binary number consisting of multiple bytes (e.g. a 32-bit unsigned integer value, the Group Number, the Element Number, etc.), the least significant byte shall be encoded first; with the remaining bytes encoded in increasing order of significance." DICOM PART 2: Conformance, in its Presentation Context Table lists Little Endian Transfer Syntax. PART 9: Point to Point Communication Support for Message Exchange specifies little endian in 4 places, for instance, "Only one transfer syntax shall be supported for point to point message exchange: little endian." Both the C and Ada programming languages have facilities for producing data with little endian byte ordering. It should be emphasized, that the standardization of byte order for transmission is totally irrelevant to the internal disk storage or analysis of the data.

The use of clustering technology and improved graphics has resulted in gating becoming a moving target. Newer forms of gating include: the locus of points produced by sketching software, ellipsoids of revolution in hyperspace and data trees. The inclusion of a DICOM Short Text field (1,024 bytes maximum) to describe the actual gating may be the easiest of the possibilities. Otherwise, each type of gating requires its own parameters. The mode of gating thus becomes the equivalent of the discriminant of a variant record or in object oriented technology is a child of the gating class.

In the optional FCS keywords some thought should be given to string lengths. Where there will be no loss of understanding, a maximum number of characters should be specified. The specification needs a data dictionary. Presently, there can be a synonym explosion. This data dictionary could be housed on the ISAC web site. The software developers would be required to register all new key words and to start them with a standard symbol; '#', '!', or '@' might be possibilities.

Short FORTRAN style key words are an anachronism (Req.10). What is needed is the standard software engineering practice of separating syllables by underscores. Many modern programming languages are case insensitive for all entities except user defined strings. The convention of starting succeeding syllables with a capital letter has the difficulty that pretty-printers in case insensitive languages can change the case of individual letters.

It must again be emphasized, that the arguments presented above only pertain to the format of FCS3.0. The data items themselves are not the subject being discussed. In fact, most of this domain expertise present in FCS3.0 should be employed in the creation of DICOM objects. There is no need to maintain a unique standard for Flow Cytometry.

## 6. POSSIBLE REPLACEMENTS FOR FCS

### 6.1. Interfile

Interfile is a specification for a file format for the exchange of nuclear medicine image and associated data<sup>4,9</sup>, which is similar to FCS in layout. The Interfile format includes<sup>9</sup>, "two types of data: administrative (header) data and binary (usually image) data. It is still possible to place both administrative data and binary data in a single contiguous

ous file, but this is no longer recommended. It is now suggested that administrative and binary data be kept in separate files, where a pointer to the file containing the binary data is indicated in the administrative data file. The administrative data are only composed of key-value pairs and exist in the form of ASCII text. The Interfile header is text based." "All administrative data are to be supplied in the form of key-value pairs in ASCII with delimiters." "The maximum permitted number of characters for a key or a value or a comment is 255 characters." End of line is employed as the terminal delimitator and ';' to start a comment. For instance:

```
patient name := JOAN DOE <CR><LF>
```

Interfile also includes a for loop structure and conditionals. Image data is described as arrays.

## **6.2.HDF**

The HDF<sup>18</sup> format has been suggested<sup>22</sup> as a possible replacement for FCS. It is an interesting standard for storing and manipulating "Scientific Data Sets"<sup>7</sup>. Of particular interest to this discussion, the HDF web site included a reservation similar to the one above concerning Interfile which also expresses reservations about putting the administrative and binary data in the same file. HDF does not seem to have any connection with Medical Informatics (Req.8) and thus is not a reasonable choice for a successor to FCS. However HDF appears to be a useful format for multidimensional scientific data and has significant similarities to DICOM. The knowledge embodied in the design of HDF, certainly is relevant to medical imaging and flow analysis and should be employed as a guide to implementing new features in DICOM.

## **6.3. DICOM 3.0**

Supporting Organizations: Initially, The American College of Radiology, and the National Electrical Manufacturers Association formed a joint committee to develop DICOM. Liaison was established with: Committee European de Normalization/Technical Committee 251 or in English, Technical Committee 251 (Medical Informatics) of the European Standards Committee (CEN TC251) and Japanese Industry Association for Radiation Apparatus (JIRA) in Japan. The standard has been reviewed by IEEE, HL7 and ANSI in the USA. The American College of Pathologists is now participant.

6.3.1. Description: DICOM 3.0 is a multivolume Standard<sup>3</sup>, which "includes: Information Object Definitions, Service Class Specifications, Data Structure & Semantics, Data Dictionary, Message Exchange, Network Communication Support for Message Exchange, and Point-to-Point Communication Support for Message Exchange<sup>3, Part 1</sup>. Besides the image graphics objects, DICOM includes<sup>3, Part 3</sup> specifications for describing: studies, reports, the acquisition of the data and the individuals involved, physician, patient, etc. The syntax and semantics of a set of device protocols for exchanging information are specified. Both the OSI and TCP/IP protocols are supported<sup>3, Part 1</sup>.

As stated in Part 1<sup>3, Part 1</sup> describing the DICOM Data Dictionary, "Part 6 of the DICOM Standard<sup>3, Part 6</sup> is the centralized registry which defines the collection of all DICOM Data Elements available to represent information. For each Data Element Part 6:

- assigns it a unique tag, which consists of a group and element number;
- gives it a name;
- specifies its value characteristics (character string, integer, etc.);
- defines its semantics (i.e. how it is to be interpreted)."

The following description of the way DICOM describes the representation of data types is taken from Part 6. The Data Element Tag is "A unique identifier for an element of information composed of an ordered pair of numbers (a Group Number followed by an Element Number), which is used to identify Attributes and corresponding Data Elements." "A Tag is represented as (gggg,eeee), where gggg equates to the Group Number and eeee equates to the Element Number within that Group." Both the Group and Element Numbers of DICOM Tags are represented in hexadecimal notation. They are unsigned 16 bit integers.

The Length of the actual value of the object is specified by a long integer (32 bits). In the case of ASCII Text

an even number of bytes is specified.

Table 2. DICOM Data Types

| DICOM Name            | Description  | Length in Bytes                         |
|-----------------------|--|---|
| Decimal String        | Chars '0'-'9', '+', '-', 'E', 'e', '.'                           | Variable, 16 Bytes Max                  |
| Date Time             | CCYY MMDDHHmmSS, Microseconds & an offset in hours and minutes   | Variable, 26 Bytes Max. Offset Optional |
| Floating Point Single | IEEE Format  | 4 Bytes                                 |
| Floating Point Double | IEEE Format  | 8 Bytes                                 |
| Integer String        | '0'-'9', '+', '-'  | Variable, 12 Bytes Max.                 |
| Long String           | Since '\ ' serves as the delimiter, it is excluded.              | Variable, 64 Chars Max.                 |
| Long Text             | All Chars including graphic character set and control characters | Variable, 10,240 Chars Max.             |
| Person Name           | Excluding control characters                                     | Variable, 64 Chars Max.                 |
| Short String          | Since '\ ' serves as the delimiter, it is excluded.              | Variable, 16 Chars Max.                 |
| Signed Long           | Signed binary integer 32 bits                                    | 4 Bytes                                 |
| Signed Short          | Signed binary integer 16 bits                                    | 2 Bytes                                 |
| Short Text            | All Chars including graphic character set and control characters | Variable, 1,024 Chars Max.              |
| Unique Identifier     | Char. subset '0'-'9', '.'  | Variable, 64 Bytes Max.                 |
| Unsigned Long         | Unsigned binary integer  | 4 Bytes                                 |
| Unsigned Short        | Unsigned binary integer  | 2 Bytes                                 |

Table 2 describes the more familiar DICOM data types. Each of the integer and floating point types has a fixed length. When a string type such as Integer String represents a numeric type, then a subset of the Latin 1 character set is employed. Strings like person name can employ the ISO specified supplementary sets (Req.6) for Latin alphabet Numbers 1 to 5, Cyrillic, Arabic Greek, and Hebrew. Work is underway to employ the two byte representations specified for other languages by ISO. DICOM also includes: enumerated, variable image/curve size, and a sequence of Items types. The latter corresponds to a record that can have as an element another record.

The following is an example of a Value Representation a Data Element from DICOM PART 6: Data Dictionary, Section 6 Registry of DICOM data elements the numeric tag is: (0010,1010) which is Patient's Name. The first hexadecimal integer, 0010, identifies the Group as pertaining to the patient. The second hexadecimal integer, 1010, identifies the Element as Patient's Name. The Group and Element Hexadecimal 16 bit unsigned integers are concatenated into one 32 bit unsigned integer. Person Name is a string with an even number of bytes and a maximum value of 64. The third element is the Value Multiplicity (VM) which specifies the number of Values contained in the Value Field. In this case, a 1 at the end indicates that there is one item. DICOM PART 5: Data Structures and Encoding, Table 6.2-1: DICOM Value Representations describes the data structures in great detail.

Since the Data Element Tag (data type) is expressed as a number, a simple table look up or database can be employed to parse the data. This provides significant protection. If the Data Element Tag in a file or transmission does not match an existing one, a warning can be given. Similarly, the layout of the data must agree with the format specified for the Data Element Tag.

One way to implement a DICOM system is to employ the objects and infrastructure created by The Object Man-



agement Group, OMG, or similar organization. The OMG has developed the Common Object Request Broker, CORBA, architecture for distributed processing<sup>19</sup> and has formed a subsection, CORBAMED<sup>2</sup>, to investigate the use of CORBA for medical informatics including the use of DICOM. The proper long term approach is to abstract DICOM into a generalized image management specification for CORBA.

6.3.2. The transition of FCS to DICOM will be facilitated by storing the present FCS data in a database. This has been already accomplished with the QC Tracker software<sup>15</sup>, which parses Flow Cytometry Standard 2.0 files and stores all of data including list mode in the tables of a relational database, AdaSAGE. AdaSAGE is presently employed to retrieve, analyze, and plot data from multiple experiments to assess the stability of flow cytometers. AdaSAGE combined with Ada includes the necessary facilities to reformat, store, retrieve, and manipulate the present FCS data into the future DICOM data structures. This manipulation includes being organized for transmission in a DICOM 3 compatible format. Table 3 FCS3.0 to DICOM 3.0 Mapping is an example of some FCS3.0 data types and their present DICOM equivalents. The data items in the FCS Header do not need to be archived, since they are only relevant to FCS format.

There are two approaches to mapping FCS to DICOM. The first is to employ the DICOM type which syntactically most closely approximates the FCS type. Since all of the FCS types are represented by either single (ASCII) or double width (Unicode) characters, this requires limiting DICOM to character based data types. The second is to select the DICOM data type to match the semantic meaning and ignore the representation of FCS. Table 3 follows the second approach. Matching the semantic meaning has three advantages: minimization of storage size, maximization of execution speed, and protection against errors (Req.2). Utilization of specific types like Integer\_32 permits both the database and compiler to check that the data and storage type match. If not, an error occurs and in Ada an exception is raised.

DICOM 3 technology provides a better match to the contents of the REQUIREMENTS Section than does FCS3.0. It must be emphasized that although DICOM 3 is a very large undertaking, the implementation of the parts of SUPPLEMENT 15 required to store and transmit microscopic images would not be much different than creating software to read and write FCS3.0. In fact, the structure of DICOM should facilitate reading of the DICOM file or data stream compared to the same operations in FCS3.0.

Table 3 FCS3.0 to DICOM 3.0 Mapping

| FCS3.0 Name | FCS3.0 Description  | FCS3.0 Type            | DICOM Type                | Comments                 |
|-------------|---|------------------------|---------------------------|--------------------------|
| \$ABRT      | Events lost due to data acquisition electronic coincidence. | Integer_String         | Unsigned Long             |                          |
| \$BTIM      | Clock time at beginning of data acquisition.                | Time:<br>hh:mm:ss[:tt] | Time:<br>hh:mm:ss.sss     | FCS 1/60; DICOM 0.000    |
| \$CELLS     | Description of objects measured.                            | string                 | Specimen Type Description | Text up to 1,024 Chars   |
| \$COM       | Comment.  | string                 | Long String               | 64 Chars                 |
| \$CYT       | Type of flow cytometer.                                     | string                 | Long String               | Manufacturer; Model Name |
| \$CYTSN     | Flow cytometer serial number.                               | string                 | Long String               |                          |
| \$DATE      | Date of data set acquisition.                               | dd-mmm-yyyy            | Date<br>Decimal_String    | yyyymmdd                 |
| \$EXP       | Name of investigator initiating the experiment.             | string                 | Person Name               | 64 Chars                 |

Table 3 FCS3.0 to DICOM 3.0 Mapping

|                |   |                |                    |  |
|----------------|---|----------------|--------------------|--|
| \$FIL          | Name of the data file containing the data set.  | string         | Referenced File ID | maximum of 8 components, each from 1 to 8 characters |
| \$INST         | Institution at which data acquired.             | string         | Institution Name   | 64 Chars   |
| \$PAR          | Number of parameters in an event                | Integer_String | Unsigned Short     |  |
| \$PnB (Binary) | Number of bits reserved for parameter number n. | Integer_String | Unsigned Short(2)  | Two overloaded methods for stored data?              |
| \$PnR          | Range for parameter number n                    | Integer_String | Unsigned Short     |  |
| \$SMNO         | Specimen (tube or well) label                   | string         | Integer String     | DICOM states an Integer String                       |

DICOM 3.0 is being extended to include digital microscopy. The use of DICOM for Pathology imaging has extensive backing. "The College of American Pathologists (CAP) organized the CAP Image Exchange Committee in 1994 to develop extensions of the DICOM Standard for Pathology imaging modalities in conjunction with ACR and NEMA."<sup>3, Supplement 15</sup> "This Supplement to the Standard is developed in liaison with other Standards Organizations including CEN TC251 (WG3 and WG4) in Europe and MEDIS-DC and JIRA in Japan, with review also by other organizations who are members of the ANSI Healthcare Informatics Standards Board (HISB)<sup>3, Supplement 15</sup>."

Table 4 shows 39 of the new microscopic image data items that should be considered to be included to describe a flow cytometry data set. Many of the original items were pairs with the end of their name being either a "Code Sequence" or a "Description". Their Value Representations were respectively a Sequence of Items and Short Text. Only one member of each these pairs is shown. The Value Representation is abbreviated to Sequence & S\_Text. . The Value Multiplicity is separated by a comma and given in order.

Table 4 Abridged & Simplified PS3.6 Section 6 Registry of DICOM Data Elements

| Name  | Value Representation Description | Value Multiplicity |
|---|----------------------------------|--------------------|
| Anatomic Location of Examining Instrument Description (Code Sequence) | Sequence & S_Text                | 1                  |
| Specimen Collection Date Time   | Date                             | 1                  |
| Specimen Collection Procedure Description (Code Sequence)             | Sequence & S_Text                | 1                  |
| Specimen Sequence   | Sequence of Items                | 1                  |
| Specimen Processing Procedure Description (Code Sequence)             | Sequence & S_Text                | 1                  |
| Source Specimen Number  | Short Text                       | 1                  |
| Specimen Type Description (Code Sequence)                             | Sequence & S_Text                | 1                  |
| Specimen Fixation Description (Code Sequence)                         | Sequence & S_Text                | 1-n, 1             |
| Specimen Stain Description (Code Sequence)                            | Sequence & S_Text                | 1-n, 1             |
| Specimen Counter-Stain Description (Code Sequence)                    | Sequence & S_Text                | 1-n, 1             |
| Specimen Extraction Description (Code Sequence)                       | Sequence & S_Text                | 1-n, 1             |
| Specimen Hybridization Description (Code Sequence)                    | Sequence & S_Text                | 1-n, 1             |
| Hybridization Amplification Description (Code Sequence)               | Sequence & S_Text                | 1-n, 1             |
| Collecting Physician  | Person Name                      | 1                  |
| Description of Specimen by Collecting Physician                       | Long Text                        | 1                  |

Table 4 Abridged & Simplified PS3.6 Section 6 Registry of DICOM Data Elements

|  |                   |        |
|--|-------------------|--------|
| Specimen Handling Precautions                        | Short Text        | 1-n    |
| Specimen Handling Special Requirements               | Short Text        | 1-n    |
| Submitting Service                                   | Long String       | 1      |
| Specimen Accession Number                            | Long String       | 1      |
| Creation Date of Accession                           | Date              | 1      |
| Creation Time of Accession                           | Time              | 1      |
| Light Source Description (Code Sequence)             | Sequence & S_Text | 1-n, 1 |
| Polarization   | Code String       | 1      |
| Emission Filter Description (Code Sequence)          | Sequence & S_Text | 1-n, 1 |
| Illumination Methodology Description (Code Sequence) | Sequence & S_Text | 1-n, 1 |
| Polarization Angle                                   | Decimal String    | 1      |

6.4 Continuing Discussion. Much of the commentary and descriptions of various file formats referred to above was taken from the ISAC Web Page<sup>10</sup>, which “publishes extended comments and alternative views.” If you are interested in future discussions on this subject, please look at the forum on Data File Standard for Flow Cytometry<sup>11</sup>. Other subjects relevant to Cytometry will also be found at the ISAC web site and, of course, at future SPIE meetings including this one on Advanced Techniques in Analytical Cytology.

## 7. CONCLUSIONS

1. FCS3.0 should be an interim standard, which should be replaced by a DICOM Flow Cytometry Standard.
2. If at all possible, all new FCS key words should specify Elements of DICOM Groups.
3. The Flow Cytometry DICOM Specification should be a sibling of the microscope image standard.
4. A relational database can store, manipulate, and translate the data from one format to another.

## 8. ACKNOWLEDGEMENTS

The authors wish to thank the management and Robert Rios of the software engineering department of Phoenix Flow for encouragement and many useful discussions. We wish to thank David M. Coder, Ph.D., Editor, ISAC WWW Home Page for publishing a preliminary version of this document.

## 9. REFERENCES

1. U.J. Balis, W.D. Bidgood, Jr., S. B. Dove, L. Korman, D. Snavely, *Digital Imaging and Communications in Medicine (DICOM) SUPPLEMENT 15, Draft for Public Review, Visible Light Image, Anatomic Frame of Reference, Accession, and Specimen for Endoscopy, Microscopy, and Photography.*, NEMA Diagnostic Imaging and Therapy Systems Division, 1300 North 17th Street, Suite 1847, Rosslyn VA 22209, USA, E-mail: dav\_snavely@nema.org, <http://dumccss.mc.duke.edu/standards/HL7/committees/image-management/DICOM/>. Refer to: VERSION:1.33, FILE: v133lba.doc size: 171K, DATE: November 20, 1996. (There is also available v133lba.rtf 20-Nov-96 06:26 262K).
2. CORBAmed home page: <http://www.omg.org/corbamed/home.htm>
3. *DICOM (Digital Imaging and Communication in Medicine) Standard* consists of multivolumes:
  3. Part 1: Introduction and Overview
  3. Part 2: Conformance

- 3. Part 3: Information Object Definitions
- 3. Part 4: Service Class Specifications
- 3. Part 5: Data Structures and Encoding
- 3. Part 6: Data Dictionary
- 3. Part 7: Message Exchange
- 3. Part 8: Network Communication Support for Message Exchange
- 3. Part 9: Point-to-Point Communication Support for Message Exchange
- 3. Part 10: Media Storage and File Format for Data Interchange
- 3. Part 11: Media Storage Application Profiles
- 3. Part 12: Media Formats and Physical Media for Data Interchange
- 3. Part 13: Print Management Point-to-point Communication Support
- 3. Supplement 15, Draft for Public Review Visible Light Image, Anatomic Frame of Reference, Accession, and Specimen for Endoscopy, Microscopy, and Photography

The DICOM Standard is available from: The National Electronic Manufacturers Association, 2101 L Street, NW, Suite 300, Washington, DC 20037 or ASTM Committee E-31 on Computerized Systems, Chairman Donald A. Nelson, Cedar Rapids Medical Education Program, 1026 A Ave. NE Cedar Rapids IA 52402-5098, (319) 369-7393. All current NEMA Standards including DICOM 3.0 are now available on CD-ROM from Telemanagement Information Handling Services, 800.525.7052, [http:// www.ih.com](http://www.ih.com). ISO and other standards are also available from this source.

DICOM home page: [http://www.xray.hmc.psu.edu/dicom/dicom\\_home.html](http://www.xray.hmc.psu.edu/dicom/dicom_home.html)

From Frequently Asked Questions: <http://chasse-spleen.xray.hmc.psu.edu/home.html>

“Are the DICOM Standard Documents available in electronic format? Yes! FTP access to the ACR-NEMA Standard (DICOM 3.0) Documents is available. All of the final draft versions of the standard can be obtained by anonymous ftp from [ftp.xray.hmc.psu.edu](ftp://ftp.xray.hmc.psu.edu). They are located in the directory /dicom\_docs. There are subdirectories that contain the documents in postscript (/dicom\_docs/dicom\_3.0/postscript), FrameMaker (/dicom\_docs/dicom\_3.0/frame), and Microsoft Word (/dicom\_docs/dicom\_3.0/word\_hqx).”

“The electronic copies of all DICOM documents are final drafts. Official printed standards documents are only available from: NEMA, Office of Publications, 2102 L Street, N.W., Washington, D.C., 20037, USA”

- 4. T. D. Craddock, D. L. Bailey, B. F. Hutton, F. Deconinck, E. Busemann Sokole, H. Bergmann, U. Noelpp. “A standard protocol for the exchange of nuclear medicine image files,” Nucl. Med. Commun. Vol.10, pp. 703-713, 1989.
- 5. P. N. Dean, C. B. Bagwell, T. Lindmo, R. F. Murphy, and G. C. Salzman, (Data File Standards Committee), “Data File Standard for Flow Cytometry,” Cytometry Vol. 11, pp. 323-332, 1990.
- 6. FDA, ODE Guidance for the Content of Premarket Submission for Medical Devices Containing Software Draft Document, FDA CDRH HOME PAGE, [http://www.fda.gov/cdrh/index.html\(dtswguid.html\)](http://www.fda.gov/cdrh/index.html(dtswguid.html))
- 7. HDF Scientific Data Sets (SD API), [http://hdf.ncsa.uiuc.edu/sds\\_api.html](http://hdf.ncsa.uiuc.edu/sds_api.html)
- 8. H. Hoehn, O. Ratib, and C. Girard, “PAPYRUS 3.1, The DICOM compatible image file format, Working Draft 0.1”, Digital Imaging Unit, Center Medical Informatics, University Hospital of Geneva, (Reg. 93015), February 1996.
- 9. Interfile, Anonymous ftp from <nucmed.med.nyu.edu>:/pub/open/interfile.tar.Z. UNIX tar and uncompress, or their PC equivalents are required to read the files. For the PC users go to the Interfile directory and use your FTP download ASCII mode for the README, \*.ASC, and \*.HDR; files, and binary mode for the.V33 and.IMG files. Personal communication David Reddy, President Radio Logic, Inc. Tel: 860-669-9080 x1; e-mail: [reddy@nucmed.med.nyu.edu](mailto:reddy@nucmed.med.nyu.edu)
- 10. ISAC WWW Home Page, <http://nucleus.immunol.washington.edu/ISAC.html>, D. M. Coder, Editor.
- 11. ISAC Web page “Data Standards” [http://nucleus.immunol.washington.edu/ISAC/data\\_standards.html](http://nucleus.immunol.washington.edu/ISAC/data_standards.html), D. M. Coder, Editor, ISAC WWW Home Page, 1997.
- 12. ISO Online <http://www.iso.ch/>

13. ISO 8859 Alphabet Soup. <http://www.cs.tu-berlin.de/~czyborra/charsets/>
14. R. C. Leif and S. B. Leif, "ISAC Flow Cytometry Standard Version 3, FCS3.0", <http://nucleus.immunol.washington.edu/ISAC/fcs3/leif.html>, 30 November, 1996.
15. R. C. Leif, R. Rios, M. C. Becker, C. K. Becker, J. T. Self, and S. B. Leif, "The Creation of a Laboratory Instrument Quality Monitoring System with AdaSAGE". *Advanced Techniques in Analytical Cytology, Optical Diagnosis of Living Cells and Biofluids*, Ed. T. Askura, D. L. Farkas, R. C. Leif, A. V. Priezzhev, and B. J. Tromberg. A. Katzir *Progress in Biomedical Optics Series Editor SPIE Proceedings Series*, Vol. 2678, pp. 232-239, 1996.
16. R. C. Leif, I. Rosello, D. Simler, G. P. Garcia, and S. B. Leif; "Ada Software for Cytometry". *Analytical and Quantitative Cytology and Histology* Vol. 13 pp. 440-450, 1991.
17. N. G. Leveson, *Safeware, System Safety and Computers*, Addison-Wesley, ISBN 0-201-11972-2 pp. 412-413, 1995.
18. NCSA HDF Levels of Interaction, <http://hdf.ncsa.uiuc.edu/logo.html>
19. Object Management Group, home page: <http://www.omg.org/>
20. Papyrus home page: <http://expasy.hcuge.ch/www/UIN/papyrus.html>
21. R. Riehle, "Can Software Be Safe? --An Ada Viewpoint," *Embedded Systems Programming*, Vol. 9 (13) pp. 28-40, Dec. 1996.
22. L. Seamer, "Alternative Data File Formats - A Response," ISAC NetForum - Message Replies, <http://nucleus.immunol.washington.edu/cgi-bin/netforum/fcs3/a/8--3.1.2.1.0>, Dec., 1996.
23. L. Seamer, B. Bagwell L. Barden, M. Christofferson, L. E. Magruder, G. Malachowski, R. F. Murphy, D. Redelman, G. C. Salzman, and J. C.S. Wood. "Data File Standards Committee of the International Society for Analytical Cytology (ISAC), "Data File Standard for Flow Cytometry, Version FCS3.0," ISAC WWW Home Page