FCS 3.1 Implementation Guidance[1]

Chris Bray[1,2], **Josef Spidlen**[2,†], and **Ryan R. Brinkman**[2,*]

[1] Verity Software House, Topsham, ME, USA

[2] Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada

Abstract

The Flow Cytometry Standard (FCS) format was developed back in 1984. Since then, FCS became the standard file format supported by all flow cytometry software and hardware vendors. Over the years, updates were incorporated to adapt to technological advancements in both flow cytometry and computing technologies. However, flexibility in how data may be stored in FCS has led to implementation difficulties for instrument vendors and third party software developers. In this technical note, we are providing implementation guidance and examples related to FCS 3.1, the latest version of the standard. By publishing this text, we intend to prevent potential compatibility issues that could be faced when implementing the FCS spillover and preferred display keywords that have arisen during discussions among some implementers.

Keywords

flow cytometry; FCS; data standard; file format; bioinformatics

Introduction

The Flow Cytometry Data File Standard facilitates the development of software for reading and writing flow cytometry data. The goal of the standard is to provide a uniform file format that allows files created by any type of acquisition hardware to be analyzed by any third-party data analysis tool. The original FCS standard was published in 1984 as FCS 1.0 (1), amended in 1990 as FCS 2.0 (2), in 1997 as FCS 3.0 (3), and finally in 2010 as FCS 3.1 (4). Here we give guidance and data file examples from instrument vendors to prevent potential compatibility issues identified during discussion with implementers.

Spillover Implementation Guidance and Examples

The FCS 3.1 standard does not include a specific example of how to handle writing the spillover matrix when multiple measurement types, such as the height (H) and area (A) of a signal, are involved. The general approach is to set up a sparse spillover matrix that isolates the different measurement types by setting some matrix elements to zero, indicating no spillover between two measurements. By specifying a value of zero for the spillover between different measurement types, the different measurement types are isolated in the matrix. Thus, the spillover for one measurement type can be properly accounted for

independent of any other type using a single matrix. It is the responsibility of the FCS writer to ensure that the $SPILLOVER keyword is encoded in this way and that the matrix is well formed and invertible. By writing the $SPILLOVER keyword in this manner, FCS readers will be able to use the spillover matrix to properly compensate all measurements without any knowledge of the measurement types. FCS readers should make no assumptions about the measurement types. Note that FCS readers are not required to compensate data using the spillover matrix from the FCS file. For example, an alternative matrix can be extracted from a Gating-ML (5) file or a new matrix may be created based on user-supplied compensation control files.

A drawback of the approach that isolates the different measurement types within the spillover matrix becomes apparent when different measurements types are acquired for different fluorescence parameters. Consider the following example. We have three signals that we are interested in (FL1, FL2, and FL3). Please note that parameters are referenced by their identifiers (values of their $PnN keywords) within the $SPILLOVER keyword value. For the purpose of this example it is adequate to identify the parameters by fluorescence detector name instead of using a marker-fluorochrome-like name as typically used when reporting flow cytometry experiments. In our example, we are measuring both height and area of the signal for FL1 and FL2, but only the area for FL3. The example spillover is as follows: 12% FL1-H to FL2-H; 10% FL1-A to FL2-A; 4% FL1-A to FL3-A; 19% FL2-H to FL1-H; 20% FL2-A to FL1-A; 13% FL2-A to FL3-A; 15% FL3-A to FL1-A; 8% FL3-A to FL2-A. The $SPILLOVER keyword for this example is created as follows:

$SPILLOVER/5,FL1-H,FL2-H,FL1-A,FL2-A,FL3-A,

1,0.12,0,0,0,0.19,1,0,0,0,0,0,1,0.1,0.04,0,0,0.2,1,0.13,0,0,0.15,0.08,1/

Table 1 displays the corresponding spillover matrix. The example matrix contains three signals of interest, FL1, FL2, and FL3. Both height and area are being captured for the signal of FL1 and FL2, but only area is being measured for FL3. Consequently, since FL3-H was not collected, the compensation of FL1-H and FL2-H will not be complete – the signal crossover from FL3-H will not be accounted for if the spillover matrix is strictly followed. Intuitively, there is a correlation between FL3-A and FL3-H and therefore, an algorithm could use the knowledge of FL3-A and the existing spillover coefficients in order to perform a better compensation than simply ignoring FL3. However, the area and height of a signal are different measurement types and their combination in this sense is nontrivial and may lead to significant errors. Therefore, it is recommended that appropriate compensation controls with the full range of measurements types are acquired and used for proper compensation.

Use of Suggested Visualization Scale ($PnD)

$PnD is a set of new optional keywords in FCS 3.1 that recommend visualization scale for parameters with two options: linear and logarithmic. However, the FCS 3.1 standard does not include specific guidelines as how to use the suggested visualization scale if a modern "log-like" display method is used.

In the past, analogue-based analyzers commonly used a limited number of channels (e.g., 1,024) to store parameter values using the integer data type. While forward and side scatter were commonly captured and analyzed on a linear scale, most fluorescence parameters utilized logarithmic amplification. See description of $PnE and $PnG in the FCS specification (4) for information regarding the channel to scale conversion. Typically, 1,024 logarithmic channels corresponded to linear scale values (e.g., 0 – 10,000), already compensated for spectral overlaps. The channel values could be simply displayed next to a 4-decade logarithmic scale. Linear parameters could be displayed the same way, only next to

a linear scale. The $PnE keywords specified both the scale and the proper visualization for each parameter.

More recently, the final visualization has been further improved by several "log-like" display methods (6,7) that avoid deceptive effects of logarithmic scaling for low signals and compensated data. Also, in the era of modern digital instruments, all parameters are typically stored linearly as single precision floating point values ($PnE/0,0/ for all parameters) and therefore, the use of $PnE to suggest visualization is no longer possible. Several instrumentation vendors added proprietary keywords to address the need for a mechanism to suggest proper parameter visualization.

The FCS data file standard has always been focused on flow cytometry data with a minimal set of information allowing for basic interpretation, and extensive visualization standardization is beyond the scope of FCS. Therefore, although there are many novel "log-like" display methods, $PnD intentionally supports only two types of scales: linear and logarithmic, in a manner consistent with $PnE. If a "log-like" method is used for visualization while an FCS file is being created then it is recommended that a "similar" logarithmic option is saved in the $PnD keyword.

For example a "logicle" visualization (7) showing a parameter with a resolution of 262,144 (or approximately 5.42 decades) using 4.5 "asymptotic logicle decades" may be approximated as $PnD/Logarithmic,6.42,0.1/. $PnD does not prescribe any transformation of the data, as in scaling to a certain number of decades. Therefore, a simple 4.5 decades logarithmic scale (with offset of 1) would result in a very different visualization pushing large values off the display. A logarithmic visualization closer to the original "logicle" can be obtained by adjusting the top of the scale properly. In addition, the choice of 0.1 as the offset more closely reflects "logicle's" visualization of values around zero. In this case, specifying 6.42 decades reaches the full range of the parameter values ($0.1*10^{6.42} =$ ~263,000).

If multiple visualization scales are being used for the same parameter (e.g., in different plots) then the software tool may choose any of these to suggest visualization of that parameter. FCS is not intended to prescribe details about the representation of flow cytometry data such as types of plots, colors, number of tick marks, position and font of labels, etc. These are left up to analytical software tools and so is the selection of "their favorite" "log-like" scale. The $PnD keywords provide only an initial hint for whether parameters are best viewed on a linear or on a log scale. Optionally, $PnD may also be used to zoom on the "interesting" section of the data. However, it is not mandatory that any software uses the suggested visualization scale. For example, it is to be expected that most tools will replace the suggestion of the logarithmic scale with their own "log-like" display methods, or the end users may still be able to select completely different scales to visualize specific parameters.

Linear visualization scale may be suggested as follows:

$PnD/Linear,f1,f2/ $P3D/Linear,0,1023/

Data should be displayed on a linear scale. Both f1 and f2 parameter values are in "scale" units, not "channel" units. Historically, channel values have been used in FCS files to capture light intensity within a limited set of available bins (typically 256 or 1024). A logarithmic amplifier has commonly been used to maximize the amount of useful information captured in the file. During analysis, these channel values have typically been converted to scale values, which linearly ("more or less") corresponds to the amount of

measured light intensity. With modern instruments (and $PnE/0,0/), scale values are often saved directly in FCS as floating point numbers.

- f1: Lower bound - the scale value corresponding to the left edge of the display

- f2: Upper bound - the scale value corresponding to the right edge of the display

Examples:

- $P1D/Linear,0,1000/

  The $PnD keyword specifies a linear display ranging from 0 to 1,000 (inclusive).

- $P3D/Linear,-5000,262144/

  The $PnD keyword specifies a linear display ranging from -5,000 to 262,144 (both inclusive); note that even before compensation, there may be negative values in the FCS data files.

- $P4D/Linear,0,1000/ $P4B/16/ $P4R/1024/ $P4E/4,1/

  The $PnD keyword specifies that the data should be shown in linear scaling, with only the bottom 10th of the scale values shown since as per the $P4R and $P4E settings, the upper bound for scale values is 10,000 (10^4). This will restrict the display to channel values between 0 and 768 (the bottom 3 decades), with channels being distributed exponentially in the linear display.

Logarithmic visualization scale may be suggested as follows:

$PnD/Logarithmic,f1,f2/ $P2D/Logarithmic,4,1/

Data should be displayed with logarithmic scaling. Both, f1 and f2 parameter values shall be positive values.

- f1: Decades - The number of decades to display.

- f2: Offset - The scale value corresponding to the left edge of the display.

This keyword recommends a logarithmic display to show scale values ranging from f2 to $f2*10^{f1}$. As mentioned before, this is only an initial hint; a particular software tool may choose to use a suitable "log-like" display or any other method instead.

Examples:

- $P2D/Logarithmic,4.5,0.1/

  The $PnD keyword specifies that the data should be shown on a logarithmic scale ranging from 0.1 to $0.1*10^{4.5}$ (approximately 3,162), which is 4.5 decades of display width.

- $P4D/Logarithmic,3,1/ $P4B/16/ $P4R/1024/ $P4E/4,1/

  The $PnD keyword specifies that the data should be shown in logarithmic scaling, with only the bottom 3 decades shown (scale values between 1 and 1000). This will restrict the display to channel values between 0 and 768 (1024*3/4).

Figure 1 illustrates how the $PnD keywords can be used to zoom on a specified section of the data with both linear and logarithmic scale. Flow cytometry data was acquired on an instrument that captures all parameters as floating point numbers on a linear scale with $PnR/262144/. The plots display two parameters: the area of forward scatter (FCS-A, FCS file parameter 1, displayed on the vertical axis) and the area of the CD20 APC fluorescence

parameter (FL1-A, FCS file parameter 3, displayed on the horizontal axis). The FCS file contains $PnD keywords suggesting visualization of these parameters as follows: $P1D/ Linear,40000,200000/ and $P3D/Logarithmic,3,10/. Without considering the $PnD keywords a software tool may choose to display the full range of the data using a linear scale for forward scatter and a logarithmic scale for the fluorescence parameter. Taking $PnD into account a software tool would zoom in to forward scatter values between 40,000 and 200,000 displayed on linear scale and to CD20 APC values between 10 and 10,000 on logarithmic scale (3 decades starting from 10^1)

Conclusion

FCS 3.1 represents a minor revision of a well-established flow cytometry standard file format. Consequently, its adoption by software and hardware vendors should not present any major technological challenges. However, there have been cases historically when either incorrect implementation by a specific vendor or the large amount of options in the FCS format caused difficulties to properly read and unambiguously interpret certain data files. We hope this supplemental information will help prevent potential compatibility issues that could be faced when implementing the FCS spillover and preferred display keywords, and those with additional questions are urged to contact the Data Standards Task Force. To further aid third party software vendors, we have collected FCS files from different instruments and software applications to provide a test set of FCS files for development and testing. These FCS files (including the first FCS 3.1 examples) are publicly available from ISAC's public flow cytometry repository at https://flowrepository.org/id/FR-FCM-ZZZ4 (Experiment id FR-FCM-ZZZ4, "FCS collection for software testing"). While ISAC's Data Standards Task Force does not provide conformance testing, the collection of data should aid developers in understanding how others have interpreted the standard. This collection includes figures to visually demonstrate the representation of the data in software tools provided by vendors of instruments that created the data files. This dataset will allow third party software developers to not only test whether they are able to read various data files but also to compare their visualization against that of the providers to quickly identify potential issues. We hope that this guide together with the mentioned FCS collection will significantly facilitate interoperability among flow cytometry instruments and third party software tools.

Acknowledgments

References

1. Murphy RF, Chused TM. A proposal for a flow cytometric data file standard. Cytometry. 1984; 5(5):553–555. [PubMed: 6489069]

2. Dean PN, Bagwell CB, Lindmo T, Murphy RF, Salzman GC. Data File Standard for Flow Cytometry. Cytometry. 1990; 11(3):323–332. [PubMed: 2340769]

3. Seamer LC, Bagwell CB, Barden L, Redelman D, Salzman GC, Wood JC, Murphy RF. Proposed new data file standard for flow cytometry, version FCS 3.0. Cytometry. 1997; 28(2):118–122. [PubMed: 9181300]

4. Spidlen J, Moore W, Parks D, Goldberg M, Bray C, Bierre P, Gorombey P, Hyun B, Hubbard M, Lange S, Lefebvre R, Leif R, Novo D, Ostruszka L, Treister A, Wood J, Murphy RF, Roederer M, Sudar D, Zigon R, Brinkman RR. Data File Standard for Flow Cytometry, version FCS 3.1. Cytometry. 2010; 77A(1):97–100. [PubMed: 19937951]

5. Spidlen J, Leif RC, Moore W, Roederer M, Brinkman RR. Gating-ML: XML-based Gating Descriptions in Flow Cytometry. Cytometry. 2008; 73A(12):1151–1157. [PubMed: 18773465]

6. Bagwell CB. Hyperlog-a flexible log-like transform for negative, zero, and positive valued data. Cytometry. 2005; 64A(1):34–42. [PubMed: 15700280]

7. Parks DR, Roederer M, Moore WA. A new "logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytometry. 2006; 69A(6):541–551. [PubMed: 16604519]
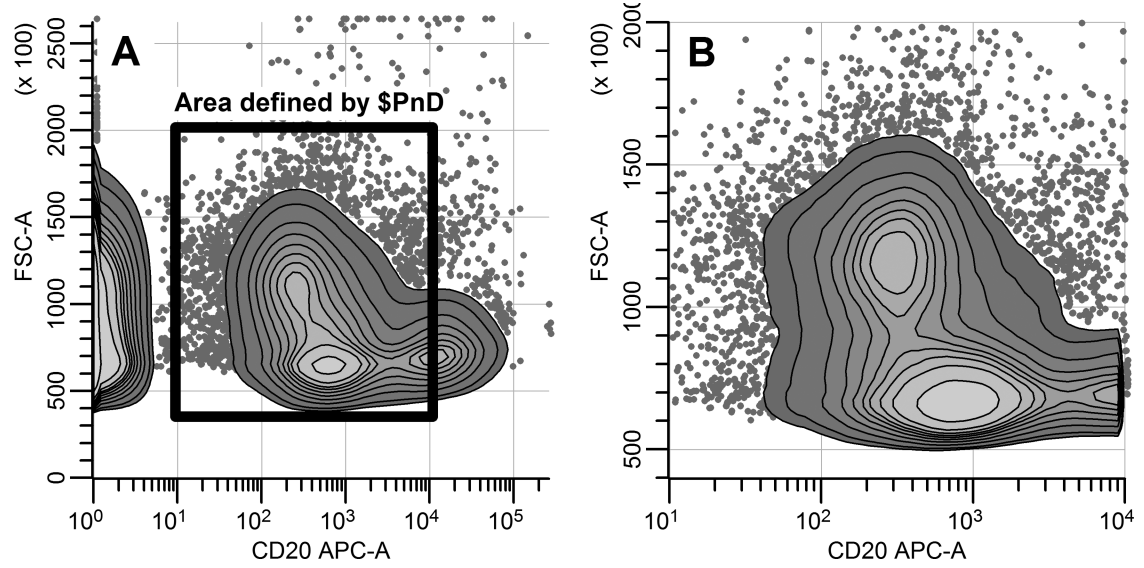
Figure 1.
The $PnD keywords utilized to zoom on a specific part of data. A: Display without considering the $PnD keywords. B: Display taking the $PnD values into account.

Table 1

Example spillover matrix with multiple measurement types of the same signal.

	FL1-H Detector	FL2-H Detector	FL1-A Detector	FL2-A Detector	FL3-A Detector
FL1-H Signal	1.00	0.12	0.00	0.00	0.00
FL2-H Signal	0.19	1.00	0.00	0.00	0.00
FL1-A Signal	0.00	0.00	1.00	0.10	0.04
FL2-A Signal	0.00	0.00	0.20	1.00	0.13
FL3-A Signal	0.00	0.00	0.15	0.08	1.00