

TECHNICAL NOTE



Data File Standard for Flow Cytometry, Version FCS 3.2

Josef Spidlen,^{1*} Wayne Moore,^{2‡} David Parks,^{3‡} Michael Goldberg,⁴ Kim Blenman,⁵ James S. Cavenaugh,⁶ the ISAC Data Standards Task Force,[†] Ryan Brinkman^{7,8,9}

¹Informatics, BD Life Sciences – FlowJo, Ashland, Oregon

²Genetics Department, Stanford University School of Medicine, Stanford, California

³Stanford Shared FACS Facility, Stanford University, Stanford, California

⁴BD Biosciences, San Jose, California

⁵Yale School of Medicine, New Haven, Connecticut

⁶Consultant, Rochester, New York

⁷Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia, Canada

⁸Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

⁹Cytapex Bioinformatics Inc, Vancouver, British Columbia, Canada

Received 3 July 2020; Revised 26 August 2020; Accepted 28 August 2020

Grant sponsor: Natural Sciences and Engineering Research Council of Canada, Grant number: R01GM118417

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Josef Spidlen, 385 Williamson Way, Ashland, Oregon 97520.

Email: josef.spidlen@bd.com

†Full list of the ISAC Data Standards Task Force provided in the Supporting Information Appendix D.

‡These authors contributed equally to this work.

Abstract

FCS 3.2 is a revision of the flow cytometry data standard based on a decade of suggested improvements from the community as well as industry needs to capture instrument conditions and measurement features more precisely. The unchanged goal of the standard is to provide a uniform file format that allows files created by one type of acquisition hardware and software to be analyzed by any other type. The standard retains the overall FCS file structure and most features of previous versions, but also contains a few changes that were required to support new types of data and use cases efficiently. These changes are incompatible with existing FCS file readers. Notably, FCS 3.2 supports mixed data types to, for example, allow FCS measurements that are intrinsically integers (e.g., indices or class assignments) or measurements that are commonly captured as integers (e.g., time ticks) to be more represented as integer values, while capturing other measurements as floating-point values in the same FCS data set. In addition, keywords explicitly specifying dyes, detectors, and analytes were added to avoid having to extract those heuristically and unreliable from measurement names. Types of measurements were formalized, several keywords added, others removed, or deprecated, and various aspects of the specification were clarified. A reference implementation of the cyclic redundancy check (CRC) calculation is provided in two programming languages since a correct CRC implementation was problematic for many vendors. © 2020 International Society for Advancement of Cytometry

Key terms

flow cytometry; FCS; FCS 3.2; data standard; file format; bioinformatics

THE Flow Cytometry Data File Standard was published in 1984 (1) to facilitate interoperability of flow cytometry data analysis software with data acquired on different types of flow cytometry instruments. The standard was revised to FCS 2.0 (2) in 1990, FCS 3.0 (3) in 1997, and FCS 3.1 (4) in 2010. Here, we present FCS 3.2, which is a revision addressing such needs and also incorporating a decade of suggested improvements from the cytometry community at large. The normative version of the FCS 3.2 specification can be found in the Supporting Information SI as well as at <http://flowcyt.sf.net/fcs/fcs32.pdf>. It will also be posted at ISAC website upon the publication of this manuscript. This Technical Note provides a summary of changes and guides implementors through some of the new features of the FCS 3.2 specification.

SUMMARY OF CHANGES

Support for Mixed Numeric Data Types

Historically, the FCS file format supported several data types including integers and both single and double precision floating point values, but those data types had to be the same for all data in the FCS data set. As of FCS 3.2, different data types can be used for different types of measurements. This is particularly useful in cases



Published online 16 September 2020 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.24225

© 2020 International Society for Advancement of Cytometry

where certain measurements are well represented as integers (such as time ticks measured in discrete units of the \$Timestep variable, indices or class assignments) while the remaining measurements are well captured as floating-point values. This also allows one or more measurements to be represented with double precision without paying a large penalty in the file size by representing all of the measurements that way. Note that taking advantage of this feature breaks the backwards compatibility of the FCS specification, but including it was considered to be critical to keep up with the increased number of measurements acquired by modern instrumentation as well as the increase in automated computational analysis of those data sets.

Support for mixed numeric data types was introduced by the addition of the optional \$PnDATATYPE keywords. The value of the \$PnDATATYPE keyword describes the type of data for FCS measurement *n* written in the DATA segment of the data set. If specified, this keyword indicates that FCS measurement *n* is stored as a different datatype than the default datatype specified by the \$DATATYPE keyword.

Concepts of Detectors and Measurement Types and Probes

Optional keywords supporting the concepts of probes, detectors, and analytes were added to eliminate the need for unreliable heuristics based on parsing those from the values of \$PnN and \$PnS keywords. The \$PnDET keyword specifies the detector name. \$PnTYPE and \$PnFEATURE keywords were added to formally specify FCS measurement signal types and evaluation features (e.g., area, height, or width), respectively. All of those keywords are optional. The \$PnTAG and \$PnANALYTE keywords were added to specify dyes used for FCS measurements and target molecules or processes associated with FCS measurements, respectively. The ISAC Probe Tag Dictionary (5) located at <https://github.com/ISAC-DSTF/ProbeTagDictionary> may be leveraged to populate values of those keywords in a standardized way.

Cyclic redundancy check Reference Implementation

While the cyclic redundancy check (CRC) is a useful mechanism to ensure the integrity of an FCS file, it is rarely used in practice (6). We attribute this to both the lack of clarity of the CRC description and the difficulties that vendors may have faced in correctly implementing this functionality. Consequently, we clarified the description and also provided free, open source reference implementations in C++ and C#.

Required \$CYT Keyword

Most existing software tools leverage the \$CYT keyword to determine the instrument used to acquire data for the data set in order to customize data scaling and presentation options

as well as other instrument-specific preferences that users may want to set. Consequently, the \$CYT keyword was made required as it was historically very useful and requiring it for FCS 3.2 does not represent a significant burden for data producers when creating the FCS file.

Various Improvements

Previously deprecated keywords such as \$GATE, \$GnE, \$GnF, \$GnN, \$GnP, \$GnR, \$GnS, \$GnT, and \$GnV were removed. Previously deprecated support for histograms was removed along with related keywords, such as \$PKn and \$PKNn. The format of the ANALYSIS segment was relaxed and related keywords (\$CSMODE, \$CSVBITS, and \$CSVnFLAG) removed. Keywords that were no longer required, such as \$GATING, \$RnI, \$RnW, and \$PnP were deprecated. The \$DATE, \$BTIM, and \$ETIM keywords were deprecated in favor of \$BEGINDATETIME and \$ENDDATETIME, which follow ISO standards. The \$PLATEID, \$PLATENAME, and \$WELLID keywords were deprecated while new more generic keywords \$CARRIERID, \$CARRIERTYPE, and \$LOCATIONID were added to identify the carrier holding the sample that was the source of a data set. The \$MODE keyword was deprecated and, if used, the value of this keyword was fixed to "L" as list mode is the only supported mode at this point. Values of \$PnB keywords that are not divisible by 8 were deprecated. The \$PnG keywords were restricted to integer data type only. This means that channel values and scale values are the same for floating point data. An optional offset was added to the format of the value of the \$PnCALIBRATION keywords. The \$BEGINTEXT, \$ENDTEXT, \$BEGINANALYSIS, and \$ENDANALYSIS keywords were made optional as they are rarely needed.

Other Additions

The optional \$UNSTAINEDCENTERS and \$UNSTAINEDINFO keywords were added to capture the vector of central values of an unstained sample and document how they were obtained. The \$UNSTAINEDCENTERS keyword represents the average background and autofluorescence of a selected unstained cell or particle population. Please see the text in the normative specification for further details including how those can be used.

The \$FLOWRATE keyword was added to capture the acquisition flow rate setting. It should contain the name of the setting as used by the data acquisition instrumentation or software. Note that this keyword is not intended to capture the actual flow rate as a volume/time value.

Improved Documentation

The terminology and structure of the specification were improved and several ambiguities were clarified throughout the text, including the description of the \$BEGINTEXT and

TECHNICAL NOTE

\$ENDSTEXT keywords. In a few cases, the description of FCS keywords was extended to be more explicit or specific. Former FCS parameters are now more appropriately referenced as FCS measurements.

CONCLUSIONS

FCS 3.2 represents a much-needed revision of the successful flow cytometry standard file format. This revision includes features and improvements requested by the community and necessary to keep up with the advancements in flow cytometry instrumentations, reagents as well computational data analysis. While FCS continues to be the main venue for raw flow cytometry data and metadata describing the acquisition conditions, additional metadata are still required to unambiguously describe experiments and analyses as specified by the Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) guidelines (7).

ACKNOWLEDGMENTS

In memory of Robert C. Leif, PhD, PMIAC—an outstanding scientist and a valuable long-time contributor to cytometry standardization efforts as well as the field of cytometry in general. The authors would like to thank everyone who provided comments and suggestions for the FCS 3.2 specification. Grant sponsors: Natural Sciences and Engineering Research Council of Canada (NSERC) and R01GM118417.

AUTHOR CONTRIBUTIONS

Wayne Moore: Conceptualization; investigation; methodology; validation; writing-review and editing. **David Parks:**

Conceptualization; investigation; methodology; validation; writing-original draft; writing-review and editing. **Michael Goldberg:** Methodology; software; validation; writing-review and editing. **Kim Blenman:** Methodology; validation; writing-review and editing. **James S. Cavenaugh:** Methodology; validation; writing-review and editing. **Ryan Brinkman:** Conceptualization; investigation; methodology; project administration; resources; supervision; validation; writing-review and editing.

CONFLICTS OF INTEREST

Members of ISAC Data Standards Task Force (ISAC DSTF) include representatives of companies selling flow cytometry instrumentation, software, and services.

LITERATURE CITED

1. Murphy RF, Chused TM. A proposal for a flow cytometric data file standard. *Cytometry* 1984;5:553–555.
2. Dean PN, Bagwell CB, Lindmo T, Murphy RF, Salzman GC. Data file standard for flow cytometry. *Cytometry* 1990;11:323–332.
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 30. *Cytometry* 1997; 28:118–122.
4. Spidlen J, Moore W, Parks D, Goldberg M, Bray C, Bierre P, Gorombe P, Hyun B, Hubbard M, Lange S, et al. Data file standard for flow Cytometry, version FCS 31. *Cytometry Part A* 2010;77A(1):97–100.
5. Blenman KRM, Spidlen J, Parks DR, Moore W, Treister A, Leif R, Bray C, Goldberg M, ISAC Data Standards Task Force, Brinkman R. ISAC probe tag dictionary: Standardized nomenclature for detection and visualization labels used in cytometry and microscopy imaging. *Cytometry Part A* 2020. <http://dx.doi.org/10.1002/cyto.a.24224>
6. Bras AE, van der Velden VHJ. Robust FCS parsing: Exploring 211,359 public files. *Cytometry Part A* 2020. <http://dx.doi.org/10.1002/cyto.a.24187>
7. Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, et al. MIFlowCyt: The minimum information about a flow cytometry experiment. *Cytometry Part A* 2008;73A(10):926–930.