IEEE SA
STANDARDS ASSOCIATION

IEEE Standard for Bioinformatics Analyses Generated by High-Throughput Sequencing (HTS) to Facilitate Communication

IEEE Engineering in Medicine and Biology Society

Developed by the
IEEE Standards Committee

IEEE Std 2791™-2020

IEEE

STANDARDS

IEEE Standard for Bioinformatics Analyses Generated by High-Throughput Sequencing (HTS) to Facilitate Communication

IEEE Standards Committee
of the
IEEE Engineering in Medicine and Biology Society

Approved 30 January 2020

IEEE SA Standards Board

Abstract: A major goal of this standard is to improve communication of bioinformatics protocols and data in order to facilitate bioinformatics workflow related exchange and communication between regulatory agencies, pharmaceutical companies, bioinformatics platform providers and researchers. A detailed communication helps ensure responsibility, reproducibility, verify bioinformatics protocol, track provenance information and promote interoperability. In addition, this standard also defines the assurance program for evaluating and certifying products against those requirements.

Keywords: analysis, bioinformatics, genomics, high-throughput sequencing, HTS, IEEE 2791™, massively parallel sequencing, MPS, next generation sequencing, NGS, pipeline, regulatory, workflow

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

PDF: ISBN 978-1-5044-6466-6 STD24065
Print: ISBN 978-1-5044-6467-3 STDPD24065

Important Notices and Disclaimers Concerning IEEE Standards Documents

IEEE documents are made available for use subject to important notices and legal disclaimers. These notices and disclaimers, or a reference to this page, appear in all standards and may be found under the heading "Important Notices and Disclaimers Concerning IEEE Standards Documents." They can also be obtained on request from IEEE or viewed at http://standards.ieee.org/ipr/disclaimers.html.

Notice and Disclaimer of Liability Concerning the Use of IEEE Standards Documents

IEEE Standards documents (standards, recommended practices, and guides), both full-use and trial-use, are developed within IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association ("IEEE SA") Standards Board. IEEE ("the Institute") develops its standards through a consensus development process, approved by the American National Standards Institute ("ANSI"), which brings together volunteers representing varied viewpoints and interests to achieve the final product. IEEE Standards are documents developed through scientific, academic, and industry-based technical working groups. Volunteers in IEEE working groups are not necessarily members of the Institute and participate without compensation from IEEE. While IEEE administers the process and establishes rules to promote fairness in the consensus development process, IEEE does not independently evaluate, test, or verify the accuracy of any of the information or the soundness of any judgments contained in its standards.

IEEE Standards do not guarantee or ensure safety, security, health, or environmental protection, or ensure against interference with or from other devices or networks. Implementers and users of IEEE Standards documents are responsible for determining and complying with all appropriate safety, security, environmental, health, and interference protection practices and all applicable laws and regulations.

IEEE does not warrant or represent the accuracy or content of the material contained in its standards, and expressly disclaims all warranties (express, implied and statutory) not included in this or any other document relating to the standard, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; and quality, accuracy, effectiveness, currency, or completeness of material. In addition, IEEE disclaims any and all conditions relating to: results; and workmanlike effort. IEEE standards documents are supplied "AS IS" and "WITH ALL FAULTS."

Use of an IEEE standard is wholly voluntary. The existence of an IEEE standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard.

In publishing and making its standards available, IEEE is not suggesting or rendering professional or other services for, or on behalf of, any person or entity nor is IEEE undertaking to perform any duty owed by any other person or entity to another. Any person utilizing any IEEE Standards document, should rely upon his or her own independent judgment in the exercise of reasonable care in any given circumstances or, as appropriate, seek the advice of a competent professional in determining the appropriateness of a given IEEE standard.

IN NO EVENT SHALL IEEE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Translations

The IEEE consensus development process involves the review of documents in English only. In the event that an IEEE standard is translated, only the English version published by IEEE should be considered the approved IEEE standard.

Official statements

A statement, written or oral, that is not processed in accordance with the IEEE SA Standards Board Operations Manual shall not be considered or inferred to be the official position of IEEE or any of its committees and shall not be considered to be, or be relied upon as, a formal position of IEEE. At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that his or her views should be considered the personal views of that individual rather than the formal position of IEEE.

Comments on standards

Comments for revision of IEEE Standards documents are welcome from any interested party, regardless of membership affiliation with IEEE. However, IEEE does not provide consulting information or advice pertaining to IEEE Standards documents. Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments. Since IEEE standards represent a consensus of concerned interests, it is important that any responses to comments and questions also receive the concurrence of a balance of interests. For this reason, IEEE and the members of its societies and Standards Coordinating Committees are not able to provide an instant response to comments or questions except in those cases where the matter has previously been addressed. For the same reason, IEEE does not respond to interpretation requests. Any person who would like to participate in revisions to an IEEE standard is welcome to join the relevant IEEE working group.

Comments on standards should be submitted to the following address:

Secretary, IEEE SA Standards Board
445 Hoes Lane
Piscataway, NJ 08854 USA

Laws and regulations

Users of IEEE Standards documents should consult all applicable laws and regulations. Compliance with the provisions of any IEEE Standards document does not imply compliance to any applicable regulatory requirements. Implementers of the standard are responsible for observing or referring to the applicable regulatory requirements. IEEE does not, by the publication of its standards, intend to urge action that is not in compliance with applicable laws, and these documents may not be construed as doing so.

Copyrights

IEEE draft and approved standards are copyrighted by IEEE under US and international copyright laws. They are made available by IEEE and are adopted for a wide variety of both public and private uses. These include both use, by reference, in laws and regulations, and use in private self-regulation, standardization, and the promotion of engineering practices and methods. By making these documents available for use and adoption by public authorities and private users, IEEE does not waive any rights in copyright to the documents.

Photocopies

Subject to payment of the appropriate fee, IEEE will grant users a limited, non-exclusive license to photocopy portions of any individual standard for company or organizational internal use or individual, non-commercial use only. To arrange for payment of licensing fees, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; +1 978 750 8400. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Updating of IEEE Standards documents

Users of IEEE Standards documents should be aware that these documents may be superseded at any time by the issuance of new editions or may be amended from time to time through the issuance of amendments, corrigenda, or errata. An official IEEE document at any point in time consists of the current edition of the document together with any amendments, corrigenda, or errata then in effect.

Every IEEE standard is subjected to review at least every 10 years. When a document is more than 10 years old and has not undergone a revision process, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE standard.

In order to determine whether a given document is the current edition and whether it has been amended through the issuance of amendments, corrigenda, or errata, visit IEEE Xplore at http://ieeexplore.ieee.org/ or contact IEEE at the address listed previously. For more information about the IEEE SA or IEEE's standards development process, visit the IEEE SA Website at http://standards.ieee.org.

Errata

Errata, if any, for IEEE standards can be accessed via https://standards.ieee.org/standard/index.html. Search for standard number and year of approval to access the web page of the published standard. Errata links are located under the Additional Resources Details section. Errata are also available in IEEE Xplore: https://ieeexplore.ieee.org/browse/standards/collection/ieee/. Users are encouraged to periodically check for errata.

Patents

Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. If a patent holder or patent applicant has filed a statement of assurance via an Accepted Letter of Assurance, then the statement is listed on the IEEE SA Website at https://standards.ieee.org/about/sasb/patcom/patents.html. Letters of Assurance may indicate whether the Submitter is willing or unwilling to grant licenses under patent rights without compensation or under reasonable rates, with reasonable terms and conditions that are demonstrably free of any unfair discrimination to applicants desiring to obtain such licenses.

Essential Patent Claims may exist for which a Letter of Assurance has not been received. The IEEE is not responsible for identifying Essential Patent Claims for which a license may be required, for conducting inquiries into the legal validity or scope of Patents Claims, or determining whether any licensing terms or conditions provided in connection with submission of a Letter of Assurance, if any, or in any licensing agreements are reasonable or non-discriminatory. Users of this standard are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. Further information may be obtained from the IEEE Standards Association.

Participants

At the time this IEEE standard was completed, the IEEE 2791 Working Group had the following membership:

Raja Mazumder, *Chair*
Vahan Simonyan, *Vice Chair*

Ogan Abaan
Jonas Almeida
Gil Alterovitz
Payal Banerjee
Amanda Bell
Surajit Bhattacharya
Lee Black
Ben Busby
Kristy Cloyd-Warwick
Ryan Connor
Michael Crusoe
Dennis Dean
Paul Duncan
Josep Gelpi
Carole Goble
Jeremy Goecks
Jonathan Jacobs
Robel Kahsay
Jonathon Keeney
Charles Hadley King
Jonathan LoTempio
Xeandong Meng
David Michaels
Hiroki Morizono
Rahi Navelkar
Asa Oudes
Janisha Patel
John Penn
Megan Pottersbusch
Jonathan Pryke
Stian Soiland-Reyes
Dan Taylor
Jason Travis
Paul Walsh
Jiancha Yao

The following members of the individual Standards Association balloting group voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

Robert Aiello
Pradeep Balachandran
Juris Borzovs
Demetrio Bucaneg Jr
Carole Carey
Michael Crusoe
David Fuschi
Carole Goble
Randall Groves
Werner Hoelzl
Atsushi Ito
Piotr Karocki
Jonathon Keeney
Stuart Kerry
Yongbum Kim
Charles King
Dwayne Knirk
Susan Laird
Daozhuang Lin
Raja Mazumder
Stian Soiland-Reyes
Harry Solomon
Thomas Starai
Walter Struppler
Mark-Rene Uchida
John Vergis
Oren Yuen

When the IEEE SA Standards Board approved this standard on 30 January 2020, it had the following membership:

Gary Hoffman, *Chair*
Vacant Position, *Vice Chair*
Jean-Philippe Faure, *Past Chair*
Konstantinos Karachalios, *Secretary*

Ted Burse
J. Travis Griffith
Grace Gu
Guido R. Hiertz
Joseph L. Koepfinger*
John D. Kulick
David J. Law
Howard Li
Dong Liu
Kevin Lu
Paul Nikolich
Damir Novosel
Jon Walter Rosdahl
Dorothy Stanley
Mehmet Ulema
Lei Wang
Sha Wei
Philip B. Winston
Daidi Zhong
Jingyi Zhou

*Member Emeritus

Introduction

This introduction is not part of IEEE Std 2791-2020, IEEE Standard for Bioinformatics Analyses Generated by High-Throughput Sequencing (HTS) to Facilitate Communication.

IEEE Std 2791-2020 enables the description of bioinformatic genome analysis workflows in a standardized way. IEEE 2791 addresses the tremendous variability and uncertainty in communicating bioinformatics workflows and data related to analysis as a result of high-throughput sequencing (HTS). The need to resolve issues in communication was felt particularly strongly between the United States Food and Drug Administration (FDA) and the entities that submit any work to the FDA for regulatory analysis that includes an HTS component[1,2] (https://doi.org/10.5731/pdajpst.2016.006734 and PMC5510742). A plan for what would become IEEE Std 2791 and initial goals of the project were drafted in a collaboration between the George Washington University and the FDA in 2014. The project has grown since then to include publications, workshops, applied use cases, and a large community of participants and collaborators. IEEE 2791 Objects created according to this standard are intended to do the following:

a) To be both human and machine readable

b) To be applied to genomic analysis workflows

c) To able to capture details related to a workflow in such a way as to facilitate efficient communication and improve reproducibility and interoperability.

Efforts were made to accommodate as many tools, platforms or scripts as possible, and to be adaptable to future developments in this field under a unified set of descriptions to standardize and streamline the representations of such complex bioinformatics processes.

IEEE 2791 is a standard and a IEEE 2791 Object is an instance of that standard. High-throughput sequencing (HTS), also referred to as next-generation sequencing (NGS) or massively parallel sequencing (MPS), has increased the pace at which we generate, compute and share genomic data in biomedical sciences. As a result, scientists, clinicians and regulators are now faced with a new data paradigm that is less portable, more complex and most of all poorly standardized. The IEEE 2791 Objects are written in JSON format to encode important information on the execution of computational pipelines, or for the creation of knowledge bases. IEEE Std 2791 can be considered to be process oriented (for software pipelines) and/or product oriented (for knowledge bases). The goal of using a IEEE 2791 Object is to streamline communication of these otherwise difficult to elucidate details between stakeholders in academia, industry and regulatory agencies.

Standardized HTS data processing descriptions and data formats will promote interoperability and simplify the verification of the bioinformatics protocols applied against data. To do this, a schema has been developed to represent instances of computational analysis as an IEEE 2791 Object. An IEEE 2791 Object includes:

— Information about parameters and versions of the executable programs in a pipeline

— References to input and output test data for verification of the pipeline

— A usability domain

— Keywords

— A list of agents involved along with other important metadata, such as their specific contribution

[1]Alterovitz G et al. Enabling Precision Medicine via standard communication of NGS provenance, analysis, and results. PLoS Biol. 2018 Dec; 16(12):e3000099 DOI: https://doi.org/10.1371/journal.pbio.3000099.

[2]Simonyan V, Goecks J and Mazumder R. Biocompute Objects—A Step towards Evaluation and Validation of Biomedical Scientific Computations. PDA J Pharm Sci Technol. 2017 Mar-Apr;71(2):136-146.

Knowledge of input data is intended to be captured according to existing efforts, such as Minimim Information Required about a Glycomics Experiment (MIRAGE),[3] Minimum Information about a Proteomics Experiment (MIAPE),[4] Standards for Reporting Enzymology Data (STRENDA)[5] and to be in accordance with Minimum Information Standards.[6] In addition to all the information captured in the IEEE 2791 Object, the IEEE 2791 Object itself is intended to be independent of the execution environment, whether it is a local or a cloud-based infrastructure.

[3]Kolaric, Daniel; Rapp, Erdmann; Struwe, Weston B.; Haslam, Stuart M.; Zaia, Joseph; McBride, Ryan; Agravat, Sanjay; Campbell, Matu... Kato, Masaki; Ranzinger, Rene; Kettner, Carsten; York, William S. (1 April 2013). "The Minimum Information Required fo... Glyco...ics Experiment (MIRAGE) Project: Improving the Standards for Reporting Mass-spectrometry-based Glycoanalytic Data". Mole...lar ... Cellular Proteomics. 12 (4): 991–995. doi:10.1074/mcp.O112.026492. ISSN 1535-9476. PMC 3617344. PMID 23378518.
[4]...tor, C. F.; Paton, N. W.; Lilley, K. S.; Binz, P. A.; Julian Jr, R. K.; Jones, A. R.; Zhu, W.; Apweiler, R.; Aebersold, R.; Deutsch, E. W.; Dunn, M. J.; Heck, A. J. R.; Leitner, A.; Macht, M.; Mann, M.; Martens, L.; Neubert, T. A.; Patterson, S. D.; Ping, P.; Seymour, S. L.; Souda, P.; Tsugita, A.; Vandekerckhove, J.; Vondriska, T. M.; Whitelegge, J. P.; Wilkins, M. R.; Xenarios, I.; Yates Jr, J. R.; Hermjakob, H. (2007). "The minimum information about a proteomics experiment (MIAPE)". Nature Biotechnology. 25 (8): 887–893. doi:10.1038/nbt1329. PMID 17687369.
[5]Tipton, K.F., Armstrong, R.N., Bakker, B.M., Bairoch, A., Cornish-Bowden, A., Halling, P.J., Hofmeyr, J.-H., Leyh, T.S., Kettner, C., Raushel, F.M., Rohwer, J., Schomburg, D., Steinbeck, C. (2014) Standards for Reporting Enzyme Data: The STRENDA Consortium: What it aims to do and why it should be helpful. Perspect. Sci. 1(1.6):131-137. DOI: 10.1016/j.pisc.2014.02.012
[6]Taylor, Chris F (2008). "Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project". Nature Biotechnology. 26 (8): 889–896. doi:10.1038/nbt.1411. PMC 2771753. PMID 18688244

Contents

1. Overview 10
1.1 General 10
1.2 Scope 11
1.3 Purpose 11

2. Normative references 11

3. Acronyms 11

4. IEEE Std 2791 1

Annex A (informative) Bibliography 14

IEEE Standard for Bioinformatics Analyses Generated by High-Throughput Sequencing (HTS) to Facilitate Communication

1. Overview

1.1 General

This standard captures relevant information from a high-throughput sequencing workflow as a IEEE 2791 Object in order to enable a user to understand and interpret the workflow efficiently and with high confidence. IEEE Std 2791 is a standard that was initially created with a goal of improving efficiency in regulatory review. Pursuant to this, workflow steps and prerequisites to execute workflow steps are recorded in detail in a IEEE 2791 object. Information is recorded using key/value pairs in JavaScript Object Notation (JSON), adhering to the IEEE 2791 JSON Schema.

Information in IEEE 2791 Objects is organized by the following domains:

- **Provenance Domain:** tracks metadata about the IEEE 2791 Object
- **Usability Domain:** tracks what was done
- **Extension Domain:** provide user-defined fields
- **Description Domain:** captures a description of external resources, pipeline steps, and the relationship of I/O objects
- **Execution Domain:** describes information needed for deployment, software configuration and running applications in a dependent environment
- **Parametric Domain:** captures all parameters that customize a computational flow
- **Input and Output Domain:** contains a list of global input and output files
- **Error Domain:** describes errors, including the limits of detetecability, false positives, false negatives, statistic confidence of outcomes, and description of errors (ie., emperical or algorithmic)

This document should be read in conjunction with the open source IEEE 2791 JSON Schema files (https://w3id.org/ieee/ieee-2791-schema/) which are referred to from the text, for instance "*provenance_domain.json*" refers to (https://w3id.org/ieee/ieee-2791-schema/provenance_domain.json). Files are kept separate for organization. References in the IEEE 2791 Object schema ($ref) to these files should be replaced with the proper domain from the appropriate file. For example, line 142 of "*IEEE2791object.json*"