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

DOI:
10.1109/IEEESTD.2020.9094416

P2791™/D4
Draft Standard for Bioinformatics Analyses Generated by High-Throughput Sequencing (HTS) to Facilitate Communication

Sponsor

Standards Committee
of the
IEEE Engineering in Medicine and Biology Society

Approved <Date Approved>

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. 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: genomics, next generation sequencing, high throughput sequencing, massively parallel sequencing, NGS, HTS, MPS, workflow, pipeline, bioinformatics, analysis, regulatory

OPEN SOURCE NOTICE: This IEEE project incorporates open source software [File download: https://gitlab.com/IEEE-SA/2791/schema/-/archive/v1.3.0-alpha/schema-v1.3.0-alpha.zip; Web link: https://gitlab.com/IEEE-SA/2791/schema/tree/v1.3.0-alpha] under a BSD 3-Clause license either within the draft or as a separate file. Any person contributing material to IEEE open source software during standards development, SA ballot, or Public Review is required to provide the appropriate license to IEEE (IEEE Contributor License Agreement or CLA). Please note that any Contributions to IEEE open source software that is submitted without first providing the appropriate CLA to IEEE will not be eligible for inclusion either in the draft standard or the open source software, and may not be considered by the comment resolution group. The appropriate CLA form (https://app.box.com/s/2nlzxsru3jewdy9ello4rfkokwt9k841) must be submitted to os-fb@ieee.org, if applicable. Note that Federal or Crown employees must submit the CLA that includes Appendix A.•

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

PDF: ISBN 978-0-XXXX-XXXX-X STDXXXXX
Print: ISBN 978-0-XXXX-XXXX-X STDPDXXXXX

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

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. A current 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 ten years. When a document is more than ten 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 all IEEE standards can be accessed on the IEEE-SA Website at the following URL: http://standards.ieee.org/findstds/errata/index.html. Users are encouraged to check this URL for errata periodically.

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 http://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 draft standard was completed, the P2791 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
Jianchao Yao

The following members of the individual/entity balloting committee voted on this standard. Balloters may have voted for approval, disapproval, or abstention.

[To be supplied by IEEE]

Balloter1
Balloter2
Balloter3
Balloter4
Balloter5
Balloter6
Balloter7
Balloter8
Balloter9

When the IEEE-SA Standards Board approved this standard on <Date Approved>, it had the following membership:

[To be supplied by IEEE]

<Name>, *Chair*
<Name>, *Vice Chair*
<Name>, *Past Chair*
Konstantinos Karachalios, *Secretary*

SBMember1
SBMember2
SBMember3
SBMember4
SBMember5
SBMember6
SBMember7
SBMember8
SBMember9

*Member Emeritus

Introduction

This introduction is not part of P2791/D4, Draft Standard for Bioinformatics Analyses Generated by High-Throughput Sequencing (HTS) to Facilitate Communication.

The P2791 specification enables the description of bioinformatic genome analysis workflows in a standardized way. P2791 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 P2791 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. P2791 Objects created according to this standard are intended 1) to be both human and machine readable, 2) to be applied to genomic analysis workflows, and 3) 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.

P2791 is a standard and a P2791 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 P2791 Objects are written in JSON format to encode important information on the execution of computational pipelines, or for the creation of knowledge bases. P2791 can be considered to be process oriented (for software pipelines) and/or product oriented (for knowledge bases). The goal of using a P2791 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 a P2791 Object. A P2791 Object includes:

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

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

[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
[3] Kolarich, Daniel; Rapp, Erdmann; Struwe, Weston B.; Haslam, Stuart M.; Zaia, Joseph; McBride, Ryan; Agravat, Sanjay; Campbell, Matthew P.; Kato, Masaki; Ranzinger, Rene; Kettner, Carsten; York, William S. (1 April 2013). "The Minimum Information Required for a Glycomics Experiment (MIRAGE) Project: Improving the Standards for Reporting Mass-spectrometry-based Glycoanalytic Data". Molecular & Cellular Proteomics. 12 (4): 991–995. doi:10.1074/mcp.O112.026492. ISSN 1535-9476. PMC 3617344. PMID 23378518

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 P2791 Object, the P2791 Object itself is intended to be independent of the execution environment, whether it is a local or a cloud-based infrastructure.

[4] Taylor, 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

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

1. Overview

1.1 General

The P2791 standard captures relevant information from a high throughput sequencing workflow as a P2791 Object in order to enable a user to understand and interpret the workflow efficiently and with high confidence. P2791 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 P2791 Object. Information is recorded using key/value pairs in JavaScript Object Notation (JSON), adhering to the P2791 JSON Schema.

Information in P2791 Objects is organized by domains;

- The Provenance Domain - tracks metadata about the P2791 Object
- The Usability Domain - tracks what was done
- The Extension Domain - provide user-defined fields
- The Description Domain - captures a description of external resources, pipeline steps, and the relationships of I/O objects
- The Execution Domain - describes information needed for deployment, software configuration and running applications in a dependent environment
- The Parametric Domain - captures all parameters that customize a computational flow
- The Input and Output Domain - contains a list of global input and output files
- The Error Domain - describes errors, including the limits of detectability, false positives, false negatives, statistics confidence of outcomes, and description of errors (i.e. empirical or algorithmic).

This document should be read in conjunction with the open source P2791 JSON Schema files (https://w3id.org/2791/) which are referred to from the text, for instance "*provenance_domain.json*" refers to (https://w3id.org/2791/provenance_domain.json). Files are kept separate for organization. References in the P2791 Object schema (`$ref`) to these files should be replaced with the proper domain from the appropriate file. For example, line 142 of "*p2791object.json*" ("`$ref`": "`provenance_domain.json`") is a reference to the structure specified in the *provenance_domain.json* file. The P2791 Object Schema builds on the JSON Schema by adding domains in a way that facilitates the communication of bioinformatics workflows. A description of the domain files follows.

1.2 Scope

This standard establishes detailed and structured communication of bioinformatics protocols in order to facilitate bioinformatics workflow related exchange and communication between regulatory agencies, pharmaceutical companies, bioinformatics platform providers and researchers. Detailed communication helps ensure responsibility, verify bioinformatics protocol, track provenance information and promote interoperability.

1.3 Purpose

The standard allows for the cross platform communications of complex computation from inception to manufacturing of medical products and services. Another goal of this standard is to improve efficiency and speed of communication.

2. Normative references

The following referenced documents are indispensable for the application of this document (i.e., they must be understood and used, so each referenced document is cited in text and its relationship to this document is explained). For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments or corrigenda) applies.

JSON (RFC8259): https://tools.ietf.org/html/rfc8259
JSON Schema: draft-handrews-json-schema-01
JSON Schema Validation: draft-handrews-json-schema-validation-01

3. Definitions, acronyms, and abbreviations

For the purposes of this document, the following terms and definitions apply. The *IEEE Standards Dictionary Online* should be consulted for terms not defined in this clause. [1]

3.1 Acronyms and abbreviations

JSON JavaScript Object Notation

[1] *IEEE Standards Dictionary Online* is available at: http://dictionary.ieee.org.

SCM Source Control Management

4. P2791 Standard

4.1 General

This document describes the P2791 standard for describing bioinformatic workflows. A P2791 "Object" is an instance of the P2791 standard, and is a text file written in JSON data structure that shall consist of all domains required by the P2791 Schema (https://w3id.org/2791/). The P2791 Schema is the formal definition of the standard against which instances of the standard can be validated. JavaScript Object Notation (JSON) is a textual format used by both instances of Objects and the formal P2791 Schema, and the JSON Schema is the language used to express the P2791 Schema.

A valid Object must conform to the P2791 JSON Schema (see section 4.3), and therefore invokes all of the requirements of the JSON Schema (while a valid Object file must conform to the schema, the schema file is not technically required to create the Object file). Later versions of P2791 may be updated for conformance with future JSON Schema versions. The minimum requirement to execute the standard is the fully organized P2791 Object containing all domains in JSON Schema format. Pursuant to JSON schema, the fields required for a valid Object are listed at the top of the *2791object.json* file.

All the files in the repository are linked together (using JSON pointers as described by the JSON Schema), being referenced by the '*2791object.json*' file. The *error_domain.json* is an optional domain to further describe empirical and algorithmic sources and measures of error for a bioinformatics workflow (https://w3id.org/2791/error_domain.json), and the *extension_domain.json* is an optional domain that contains user-defined fields.

At its top-level, Objects have the following three required metadata fields: "`spec_version`", "`object_id`", and "`etag`". These lines are external to all domains. Everything except for the `etag`, `object_id`, and `spec_version` shall be included in the generation of an ETag (see https://tools.ietf.org/html/rfc7232#section-2.3) - which can be "strong" or "weak" (see https://tools.ietf.org/html/rfc7232#section-2.1). It is recommended that the ETag be deleted or updated if the object file is changed (except in cases using weak ETags in which the entirety of the change comprises a simple re-writing of the JSON).

`object_id` is a string that follows the JSON Schema format of namespace/ref shall be a unique identifier. Users are free to number Object files in the manner of their choosing, however, in order to avoid naming conflicts, it is recommended that a domain be registered with a registration authority, such as the one at https://www.biocomputeobject.org/registry.html. For example, *http://www.example.com/exampleproject/1.3.0/schemas/ABC_object0001.json*, where "*ABC*" is a registered domain, and "*_object0001.json*" is an arbitrary identifier, chosen by the owner of that domain.

The remaining top level fields are domains that partition workflow into meaningful subunits. These are the Description domain, Error domain, Execution domain, IO domain, Parametric domain, Provenance domain, and Usability domain.

The Description Domain of an Object contains a description of external resources, pipeline steps, and the relationship of I/O objects (https://w3id.org/2791/description_domain.json).

The Error Domain contains information related to the bounds of detection (such as the minimum sequence depth and minimum sequence coverage), and statistical analyses of the pipeline (such as the false negative and false positive rates). It is recommended that the keys directly under `empirical_error` and

`algorithmic_error` use a full URI. Resolving the URI should give a JSON Schema or textual definition of the field. Other keys are not allowed in `error_domain`.

The Execution Domain of an Object describes details of deployment, software configuration, and running applications in a dependent environment (https://w3id.org/2791/execution_domain.json). This may include scripts, drivers, environment variables, and other software prerequisites.

The IO Domain of an Object is a list of global input and output files that may exist on local machine or on another machine (https://w3id.org/2791/io_domain.json). It does not include references to intermediate files.

The Parametric Domain of a P2791 Object includes any parameters used in a workflow (https://w3id,org/2791/parametric_domain.json). This is intended for use when parameters are changed from default settings.

The Provenance Domain contains metadata related to the Object (https://w3id.org/2791/provenance_domain.json. It is used to track the flow of data from original source to final computation, and includes contributors, reviewers, and versioning. In the event that a P2791 Object retrospectively references an existing Object (such as an example Object), the `derived_from` field within the Provenance Domain shall reference the specific Object by `object_id` field. In the event that the Object is an example Object or is created de novo without reference to existing work, this field is not included. In the event that the Object is an example or template P2791 Object, best practice is to state this in the Usability Domain, along with relevant details (such as completeness of data, whether data is real or artificial, etc.).

The Usability Domain of an Object is a plain language description of what was done in the workflow (https://w3id.org/2791/usability_domain.json). This should align with the actual steps described elsewhere in the Object. The Usability Domain conveys the purpose of the Object.

The Extension Domain allows a user to define additional fields and is optional. The Extension Domain is for the inclusion of any additional structured information. A valid JSON schema for each extension used in this domain is expected to be specified. The schema should be name spaced, and it is recommended that resolving the namespaced URI will provide the extension's JSON Schema. The URL should be provided in the required "`extension_schema`" field. If execution portability is desired, then the included script should be in the Common Workflow Language v1.0 (https://w3id.org/cwl/v1.0/) or later format. In order to avoid potential naming conflicts, it is recommended that users register their domain with

Annex A

(informative)

Bibliography

Bibliographical references are resources that provide additional or helpful material but do not need to be understood or used to implement this standard. Reference to these resources is made for informational use only.

[B1] Community User Guide for Best Practices. https://w3id.org/biocompute/1.3.1

[B2] JSON Schema: A Media Type for Describing JSON Documents. https://tools.ietf.org/html/draft-handrews-json-schema-01