© ISO #### – All rights reserved

ISO ##### -#:####(X/DTS 8392 2022(E)

Date: 2022-08-15

ISO/TC ###/ 215/SC ##/WG # 1

Secretariat: XXXXKATS

Genomics informatics — Description rules for genomic data for genetic detection products and services

WD/CD/DIS/FDIS stage

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Warning for WDs and CDs

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-9f15d-4bda-86ba-3abb6f427ee7/iso-dts-8392

To help you, this guide on writing standards was produced by the ISO/TMB and is available at

A model manuscript of a draft International Standard (known as "The Rice Model") is available at

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

© ISO 20XX

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

© ISO 2022

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office

CP 401 • Ch. de Blandonnet 8

CH-1214 Vernier, Geneva

Phone: +41 22 749 01 11

Fax: +41 22 749 09 47

Email: copyright@iso.org

Website: www.iso.org~~www.iso.org~~

Published in Switzerland

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

~~© ISO #### – All rights reserved~~ © ISO 2022 – All rights reserved

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ~~*[or Project Committee]*~~ ISO/TC ~~*[or ISO/PC] ###, [name of committee],*~~215, *Health informatics,* Subcommittee SC ~~*##, [name of subcommittee].*~~ 1, *Genomics informatics.*

https://standards.iteh.ai/catalog/standards/sist/5b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

~~This second/third/... edition cancels and replaces the first/second/... edition (ISO #####:####), which has been technically revised.~~

~~The main changes compared to the previous edition are as follows:~~

~~— xxx xxxxxxx xxx xxxx~~

~~A list of all parts in the ISO ##### series can be found on the ISO website.~~

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

~~© ISO #### – All rights reserved~~

© ISO 2022 – All rights reserved

Introduction

The decreasing cost of sequencing and the gradual in-depth study of genomics have led to the generation of more and more genomic data, but the data quality in genomics is not optimal at present. From the dimension of data level, thethere is a lack of data integrity is serious, and the problem of semantic disunity that medical information has been facing. a problem of semantic disunity. These problems have caused great obstacles to the downstream applicationapplications.

Standardization of data is a prerequisite for data asset management and data storage and applications, with the improving ofwhich can give better store and apply thestorage for genomic data and enlarge these genomic data used in precision medicine.

This proposaldocument is based on the actual situation of industry data production, combined with the needs of upstream and downstream industry users, and refined from. It also takes into account the actually use inmade by stakeholders, and user friendlyfriendliness for all common typetypes of genomic data. Solving the problem of data scope and semantic unification can enhance the data association ability, ensure information exchange, improve data flow, improve the data quality from the aspects of data integrity, and data validity and so on, and lay a good foundation for subsequent data storage, data application and data sharing.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

© ISO #### – All rights reserved © ISO 2022 – All rights reserved

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

Genomics informatics — Description rules for genomic data for genetic detection products and services

1 Scope

~~This document applies to all the genomic data used for human genetic detection products and services.~~

This document specifies ~~the~~ requirements on the category definition and quality assessment of genomic data, including the content structure, attribute and description rules of data format, and the compilation rules of data format.

This document applies to all the genomic data used for human genetic detection products and services.

This document applies to genomic data processing and analysis.

~~This document applies~~ and to the quality evaluation/assessment ~~for~~of genomic data.

2 Normative references

~~The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.~~

~~*ISO 8601 : 2004, Data elements and interchange formats – information interchange – representation of dates and times*~~

~~*ISO/IEC 11179-1, Information technology – Metadata registries (MDR) – Part 1: Framework*~~

~~*ISO/IEC 11179-5, Information technology – Metadata registries (MDR) – Part 5: Naming*~~

There are no normative references in this document.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-9f15d-4bda-86ba-xxxxxxxx/iso-dts-8392

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

~~*The list below is always included after each option:*~~

ISO and IEC maintain ~~terminological~~terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
alignment-sequence code
continuous coding of objects in the same series, and reserving of extended space

3.2
code

Field Code Changed

© ISO 2022 – All rights reserved

representation of a piece of information such as a letter, word or phrase in another form, usually briefer

~~[SOURCE: ISO/CD 20691, 3.6]~~

3.3
code structure
representation of the composition and length of a complete code

3.4
equal length code
~~in a code~~coding system~~,~~ in which all coding objects have the same length

3.5
~~**DI**~~
data identifier
DI
identifier that uniquely distinguishes one set of data from all others

3.6
layer code
hierarchical code consisting of membership order of coded objects

iTeh STANDARD PREVIEW
(standards.iteh.ai)

3.7
sequential code
code that represents in the natural order of Arabic numerals, or letters

3.8
variable-length code
~~in a~~ code system~~,~~ in which the length of code is not exactly the same

ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/3b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

3.9
~~**VI**~~
version identifier
~~numeric or symbolic identifier(s) of versions~~

VI
unique number assigned to identify a version of submitted genomic data

4 Data format attribute and description rules

Genomic data can be classified as unstructured data and structured data.

The unstructured data should be described by data format, illustration of data format and archiving catalogue.

The structured data should be described by metadata and data element code.

~~© ISO #### – All rights reserved~~

© ISO 2022 – All rights reserved

5 Composition and rules of genomic data description

5.1 Identifier

The description of genomic data should include data format, data attribute and metadata.

5.2 Data format

Data attribute elements for description of data format are totally classified of 14 attributes in ~~5~~five categories, shown ~~as Annex A~~in Table ~~A1~~ A.1. According to the universal property, there include data element common attribute and data element specific attribute.

5.3 Data archiving catalogue

Data elements for data archiving catalogue are totally classified of 14 attributes in ~~5~~five categories, shown ~~as Annex A~~in Table ~~A2~~ A.2. According to the universal property, there include data element common attribute and data element specific attribute.

5.4 Metadata

Data elements for metadata description are totally classified of 14 attributes in ~~5~~five categories, shown ~~as Annex A~~in Table ~~A3~~ A.3. According to the universal property, there include data element common attribute and data element specific attribute.

6 Core elements and rules for the description of genomic data

6.1 Identifier

Identifier shall use alphanumeric code with two-level structure, including ~~Data Identifier (~~DI~~)~~ and ~~Version Identifier (~~VI~~).~~.

EXAMPLE 1 ~~:~~DI_V1

~~1.~~a) DI consists of alphanumeric characters combining classification code and serial number:

~~—~~ ~~—~~ A category code consists of two uppercase letters;

~~—~~ ~~—~~ A group code is a 2-digit code. The numerical value has no meaning;

~~—~~ ~~—~~ A class code is a 2-digit code. The numerical value has no inherent meanings to humans. If there is no class, the class code is 00. The group code and the class code are separated by a dot;

~~—~~ ~~—~~ A sequential code is a 3-digit code. It represents the data element number under a class, starting from 001. The numerical value has no meaning.

2) VI consists of 4 elements. It follows the structure “V”+“m..m”+“.”+“n..n”, where “m..m” and “n..n” are numbers assigned in increasing order. They should be positive integer. “m..m” represents major version numbers and “n..n” represents minor version numbers.

EXAMPLE 2~~:~~ “V1.2” means the first major version and the second minor version.

~~—~~ ~~—~~ If a valid data exchange can be performed between the versions before and after the data element is updated, the updated major version number remains unchanged, and the minor version number equals to the current minor version number plus one.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/DTS 8392
https://standards.iteh.ai/catalog/standards/sist/2b4a48c0-915d-4bda-86ba-3abb6f427ee7/iso-dts-8392

~~© ISO #### – All rights reserved~~ © ISO 2022 – All rights reserved