PD ISO/TS 8392:2023



BSI Standards Publication

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

bsi.

National foreword

This Published Document is the UK implementation of ISO/TS 8392:2023.

The UK participation in its preparation was entrusted to Technical Committee IST/35, Health informatics.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

This publication is not to be regarded as a British Standard.

© The British Standards Institution 2023
Published by BSI Standards Limited 2023

ISBN 978 0 539 19952 9

ICS 35.240.80

Compliance with a Published Document cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 May 2023.

Amendments/corrigenda issued since publication

Date	Text affected

PD ISO/TS 8392:2023

TECHNICAL SPECIFICATION

ISO/TS 8392

First edition
2023-05-09

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

Informatique génomique — Règles de description des données génomiques pour les produits et services de détection génétique

Reference number
ISO/TS 8392:2023(E)

© ISO 2023



COPYRIGHT PROTECTED DOCUMENT

© ISO 2023, Published in Switzerland

All rights reserved. Unless otherwise specified, 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
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

© ISO 2023 – All rights reserved

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Data format attribute and description rules **2**
5 Composition and rules of genomic data description **2**
5.1 Identifier 2
5.2 Data format 2
5.3 Data archiving catalogue 2
5.4 Metadata 2
6 Core elements and rules for the description of genomic data **2**
6.1 Identifier 2
6.2 Name 3
7 Requirement of code **3**
7.1 Code structure 3
7.2 Code length 4
7.3 Code type and format 4
7.4 Code list naming 4
8 Compatibility with other rules **5**
Annex A (informative) **6**
Annex B (informative) **Examples of Identifier: Two-level structure DI_V1** **11**
Annex C (informative) **Examples of Metadata Code** **14**
Annex D (informative) **Code length calculation** **16**
Bibliography **17**

© ISO 2023 – 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.

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 ISO/TC 215, *Health informatics*, Subcommittee SC 1, *Genomics informatics*.

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 2023 – 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. From the dimension of data level, there is a lack of data integrity, and medical information has been facing a problem of semantic disunity. These problems have caused great obstacles to downstream applications.

Standardization of data is a prerequisite for data asset management and data storage and applications, which can give better storage for genomic data and enlarge these genomic data used in precision medicine.

This document is based on the actual situation of industry data production, combined with the needs of upstream and downstream industry users. It also takes into account the use made by stakeholders and user friendliness for all common types 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 lay a good foundation for subsequent data storage, data application and data sharing.

© ISO 2023 – All rights reserved

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

1 Scope

This document specifies 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, and to the quality evaluation/assessment of genomic data.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

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

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

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

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

3.5
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

© ISO 2023 – All rights reserved