BS ISO 4454:2022

BSI Standards Publication

Genomics informatics — Phenopackets: A format for phenotypic data exchange

National foreword

This British Standard is the UK implementation of ISO 4454:2022.

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.

Published by BSI Standards Limited 2022

ISBN 978 0 539 14689 9

ICS 35.240.80

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 July 2022.

Amendments/corrigenda issued since publication

Date	Text affected

BS ISO 4454:2022

INTERNATIONAL STANDARD

ISO 4454

First edition
2022-07

Genomics informatics — Phenopackets: A format for phenotypic data exchange

Reference number
ISO 4454:2022(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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.

Introduction

While great strides have been made in exchange formats for sequence and variation data (e.g. Variant Call Format), the majority of genotype formats do not include a means to share corresponding phenotypic (e.g. observable characteristics, signs/symptoms of disease) information. While some genomic databases have defined their own formats for representing phenotypic information, the lack of uniformity amongst these organizations hinders communication and limits the ability to perform analysis across organizations. For individuals with rare and undiagnosed disease, broad adoption and utilization of uniform, machine-readable, phenotypic descriptions could improve the speed and accuracy of diagnosis by promoting quicker, more comprehensive and cost-effective information acquisition and exchange relevant for research and medical care.

Phenotypic abnormalities of individuals are currently described in diverse places in diverse formats, such as journal/publications databases, laboratory systems, patient registries, health records, and even in social media. The structure of the data in the phenopackets exchange standard will be optimized for integration and efficient data flow across these distributed contexts. Increasing the volume of computable data across a diversity of systems will support large-scale computational disease analysis of combined genotype and phenotype data. Studies of well over 100 000 patients are thought to be required to effectively assess the role of rare variation in common disease or to discover the genomic basis for a substantial portion of diseases. Phenopackets can help integrate geographically distributed cases to build such virtual cohorts and remove the time burden on resources that need to integrate information manually.

Medical coding systems and clinical exchange standards have not to date included rich phenotypic descriptions, as they are largely focused on supporting billing and clinical encounter documentation, rather than the documenting and sharing of the biologically relevant phenotypic information needed for computational use, mechanism discovery, and precision classification. From a clinical perspective, the integration of a standard for phenotypic description and exchange into and out of EHRs would improve disease diagnosis and management, especially for genomic health and precision medicine applications.

Phenopackets enable clinicians, biologists, and disease and drug researchers to build more complete models of disease. It is designed to encourage wide adoption and synergy between the people, organizations and systems that comprise the joint effort to address human disease and biological understanding. The phenopacket proposed in this document is designed to support deep phenotyping, a process wherein individual components of each phenotype are observed and documented. The phenotypes can be constitutional or those related to a sample (such as from a biopsy).

INTERNATIONAL STANDARD **ISO 4454:2022(E)**

Genomics informatics — Phenopackets: A format for phenotypic data exchange

1 Scope

This document specifies a uniform, machine-readable, phenotypic description of an individual, patient or sample in the context of rare disease, common/complex disease or cancer.

It is applicable to academic, clinical and commercial research, as well as clinical diagnostics. While intended for human data collection, it can be used in other areas (e.g. mouse research). It does not define the phenotypic information that needs to be collected for a particular use but represents that information in an appropriately descriptive manner that allows it to be computationally exchanged between systems.

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 (all parts), *Date and time — Representations for information interchange*

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
biosample
unit of biological material from which the substrate for analysis is extracted to support the assessment, diagnosis, treatment, mitigation or prevention of a disease, disorder, abnormal physical state or its symptoms

3.2
boolean
data type having two values: one and zero (which are equivalent to true and false)

[SOURCE: ISO 2146:2010, 4.6.1]

3.3
CURIE
compact URI
generic, abbreviated syntax for expressing *uniform resource identifiers* (3.22)

3.4
deletion
variation in which a part of a chromosome or sequence of DNA is lost relative to a *reference sequence* (3.17)