First edition
2022-11

Biotechnology — Requirements for data formatting and description in the life sciences

Biotechnologie — Exigences relatives au formatage et à la description des données dans les sciences de la vie

Reference number
ISO 20691:2022(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

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 276, *Biotechnology*.

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

Life science research and the application of the obtained results in the biotechnology, diagnostics and pharmaceutical industries depend on complex data obtained from a wide range of assays, biological and functional studies, as well as process descriptions, laboratory and field measurements. This includes the use of the derived data for computational reconstruction, modelling and simulation of biological, biotechnological and physiological processes, as well as their applications in biotechnological workflows. Data enabled life sciences and biotechnology research span across a wide range of biological and biotechnological domains and applications (e.g. human health, genetically engineered organisms, environmental sciences, agriculture, bioremediation, DNA sequencing, chromatography, microscopy). Data driven, data intensive and big data analytical approaches in the life sciences are possible only with the use of computational methods and through consistent description, structuring and integration of data.[1] Data storage, representation, meaning, interpretation, exchange and re-use are all affected by format design. This document satisfies a critical need to set a framework for interoperable and unambiguous data recording, description and transfer by setting fundamental requirements for data recorded, processed, re-used and exchanged in the life sciences enabling the maximum data value and utilization.

These life science data from different sources and recorded at different times must be findable, accessible, interoperable and reusable (F-A-I-R).[2] Data sets are valuable and useful only if they are accessible and stored in well structured, consistent formats. Data versioning, data archiving and tracing data provenance are ensured by timeless and platform independent formats. Complete and updatable metadata (i.e. data describing the data) facilitates locating, use and analysis of data.

This document provides requirements and recommendations for standardized interoperable life science data formats. It provides a conceptual framework for, as well as references to, many different subdomain-specific data formatting and description standards defined by the biotechnological and biological domain communities. A technology-independent framework of minimal requirements and rules for the coherent utilization of the referenced domain-specific formatting and description standards and their concerted interplay is described. This document, therefore, provides rules and guidelines for coherent, subdomain overarching data formatting and description, as a foundation for data integration across domains. Moreover, rules and guidelines for the creation of (sub-)domain specific standards, their interoperability and their implementations are provided.