TECHNICAL ISO/TR

First edition
2021-05

Biotechnology — Data publication — Preliminary considerations and concepts

Biotechnologie — Publication de données — Considérations et concepts préliminaires

Reference number
ISO/TR 3985:2021(E)

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 whom 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 https://www.iso.org/directives-and-policies.html).

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

The explosion of life sciences data (big data) has created a need to digitally locate data from diverse biological assays, obtained in a wide range of laboratories, and from a wide range of experimental protocols. To be able to extract value from big data, it is necessary that the data are "findable", and that the biology measured in the assay is described in a way that it can be located and interpreted. Data producer's use of a consistent method to describe the biology that their data represents can greatly improve the use of big data. This single, unified description of biological data facilitates locating and extracting value from an abundance of biological data and return increased value to funding organizations.

Many biotech communities have already developed standard data representations specific to their domain[1]. For example, MIAME[2] in the microarray community, OME/OMERO[3] in the imaging and microscopy communities, SBML[4] in the systems biology and reaction kinetics community, and MIABIS in the biobanking domain[5]. What is lacking is a consistent method of describing the represented biological information so that the same search, analysis and mining tools can locate data across the entire range of life science domains. Consensus and guidance are required and provided in this document for the biotech domain-independent annotation of biological data.

The importance of data sharing as an integral part of biological research is recognized in the research community. As a result, a diverse set of stakeholders has developed the FAIR (Findable, Accessible, Interoperable and Reusable) data sharing principles[7]. The intent of FAIR is to act as a guideline for sharing and enhancing the reusability of data holdings. Many life science funding organizations also place increased emphasis on the importance of data sharing. Some require that data sharing plans are included in grant applications and research contracts, i.e. "data must be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data[8]." Data sharing is equally critical for various national and international research and biobank networks. Data sharing is known to encourage diversity of analysis and opinion, the testing of alternative hypotheses and enabling of explorations not envisioned by the original investigators, resulting in increased value to the funding organization.

This document lays out concepts, challenges, issues and benefits that are relevant to developing International Standards for data sharing in life science research and provides an overview for specifying standards and best practices that enable data sharing.