TECHNICAL SPECIFICATION

ISO/TS 23494-1

First edition
2023-04

Biotechnology — Provenance information model for biological material and data —

Part 1: Design concepts and general requirements

Reference number
ISO/TS 23494-1:2023(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 276, *Biotechnology*.

A list of all parts in the ISO 23494 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.

Introduction

Research in life sciences has undergone significant changes during recent years, evolving away from individual projects confined to small research groups to transnational consortia covering a wide range of techniques and expertise. The exchange of research data and biological materials has become essential for the research in life sciences and biotechnology, and consequently interoperability and quality measures of data have become imperative.

At the same time several reports addressing the quality of research papers in life sciences uncovered an alarming number of ill-founded claims. The reasons for the deficiencies are diverse, with insufficient quality and documentation of the biological material used being the major issue.

Hence there is urgent need for standardized and comprehensive documentation of the whole workflow from the collection, generation, processing and analysis of the biological material to data analysis and statistics. This provenance information serves as a quality indicator and provides information on the reliability thus enabling transparency and comparability of research results.

The purpose of these documents is the standardization of provenance information management for the biotechnology domain in a way that allows for meaningful data integration. To this end, provenance information needs to be prepared in a way that enables interoperability between prevailing tools for data generation, processing and analysis. While in information technology well-established approaches to provenance information management in general are available (e.g. OPM[1] or W3C PROV[2]), the implementation for the biotechnology domain and related fields in particular is still a pending issue (as discussed in the results from the Electronic Health Records Systems for Clinical Research (EHR4CR) and TRANSFoRm projects in several papers[3][4]).

Since data in biotechnology mostly originate from analysis of biological material, it is essential that the provenance information covers the entire process chain, from the source of biological material, throughout processing and analysis of the material, generation, and processing of the data to final analysis and interpretation. With the increasing adoption of data-intensive technologies, such as next-generation sequencing (NGS), high-throughput mass spectrometry as used for proteomics or metabolomics, or high-throughput microscopy in digital pathology, and their impact on data collection strategies, consistent and comprehensive documentation of data provenance has become a necessity.

In fact, experimental designs in life sciences have moved from individual experiments with a limited amount of data towards pipelines generating a vast volume of raw digital data using massively parallel acquisition systems demanding complex data processing workflows to extract biologically relevant information. This trend is particularly evident in NGS, where the actual data acquisition device at the wet lab-digital interface (i.e. the sequencer) is completely oblivious to the details of the experiment being performed, with all the specialization pushed to the protocols used by the sample preparation procedure and to the software pipelines processing the data. Software pipelines are continuously changing due to the evolution of analytical algorithms and reference data sets, which is having a significant effect on result concordance.

In addition, particular issues, relevant to scientific domains utilizing biological material and data obtained from humans, must be considered. These include aspects of data privacy, ethics or management of identities. Notably, issues such as withdrawal of an informed consent or communication of incidental findings require the implementation of appropriate mechanisms.

The major objectives for collecting and storing provenance information are summarized as follows:

— retrospective evaluation of experimental results and data analysis with respect to the influence of standard operating procedures (SOPs) and workflow parameters;

— quality monitoring of biological materials and data entered in a workflow or analysis pipeline (e.g. against reference ranges and tolerances);

— automation of quality control procedures (e.g. comparisons between different pipelines);

— profiling of sample and data analysis to identify bottlenecks;

— assessment of fitness for purpose of biological materials and data for the intended use.

To achieve these objectives, a digitally processable description of provenance information is required.

This overarching document will be complemented by appropriate vertical standards for specific fields (e.g. collection of biological material, data generation, processing of biological material and data). The basic requirements contained in this document do not impose any limitations to future, domain-specific standards based on this document.

The standardization of provenance information requires the conceptualization and essential specifications for the generation, management, provisioning and maintenance as described in this document. Not covered in this document are additional fundamental components such as a generic model for provenance information and extensions common to all kinds of provenance information, ensuring security, privacy and non-repudiation. For particular domains in biotechnology, detailed specifications building on a common provenance model are required, covering provenance information describing:

— the life cycle of biological materials, including acquisition, processing, transport and storage.

— the data generation by analytical methods.

— the data processing and analysis in computational workflows.

This document provides definitions for relevant terms used and specifies fundamental requirements for provenance information generation, management and provisioning.

Biotechnology — Provenance information model for biological material and data —

Part 1: Design concepts and general requirements

1 Scope

This document specifies a general concept for a provenance information model for biological material and data and requirements for provenance data interoperability and serialization.

The provenance information model covers any information relevant to the quality and fitness for purpose of the biological material generated throughout the preanalytical phase of the materials life cycle from collection to analysis, data originating from analytical procedures applied to the biological material and results from further mathematical processing of the data.

This document is applicable to organizations, authorities and industries that are:

a) collecting, processing or distributing biological material for research;

b) generating, collecting, analysing or storing data on biological material.

This document does not apply to biological material and data used for other than research or in fields that are regulated by national, regional or international laws, such as medical diagnosis and therapy or food production.

NOTE International, national, or regional regulations or requirements can also apply to specific topics covered in this document.

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.

United Nations Treaty Collection. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity [online]. Available from: https://treaties.un.org/pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVII-8-b&chapter=27&clang=_en

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/