International Standard

ISO/IEC 18670

Information technology — SoftWare Hash IDentifier (SWHID) Specification V1.2

First edition
2025-04

Reference number
ISO/IEC 18670:2025(en)

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 ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Syntax ... **3**
5 Core identifiers ... **3**
5.1 General ... 3
5.2 Contents ... 3
5.3 Directories ... 4
5.4 Revisions ... 5
5.5 Releases ... 7
5.6 Snapshots ... 8
5.7 Compatibility with Git ... 9
6 Qualified identifiers ... **10**
6.1 Qualifiers ... 10
6.2 Fragment qualifiers ... 10
6.2.1 General ... 10
6.2.2 Lines qualifier ... 10
6.2.3 Bytes qualifier ... 10
6.3 Context qualifiers ... 11
6.3.1 General ... 11
6.3.2 Origin qualifier ... 11
6.3.3 Visit qualifier ... 11
6.3.4 Path qualifier ... 11
6.3.5 Anchor qualifier ... 11
6.4 Comparing qualified SWHIDs ... 12
6.5 Recommendations ... 12
Annex A (informative) **Specification versioning** ... **13**
Bibliography ... **14**

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

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 document should be noted (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by JDF [as The SoftWare Hash Identifier (SWHID) Specification Version 1.0] and drafted in accordance with its editorial rules. It was adopted, under the JTC 1 PAS procedure, by Joint Technical Committee ISO/IEC JTC 1, *Information technology*.

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 and www.iec.ch/national-committees.

Introduction

Modern software relies heavily on open source components that are developed collaboratively in a distributed setting, and that are assembled to create complex systems that evolve at a fast pace.

This has strengthened the need to precisely track, ensure availability, and guarantee integrity of the components that go into a given system for a variety of stakeholders. Academia needs to ensure that research results are reproducible, industry needs to improve the traceability of the software supply chain, and developer communities need tools to cope with the increasing complexity.

A key building block for addressing this issue is a system of intrinsic identifiers that allows users to precisely pinpoint the exact version of any software artifact, at all levels of granularity, without relying on any central registry or naming authority.

With this specification, the SWHID working group makes such a system of intrinsic identifiers, originally developed for the Software Heritage universal source code archive,[1] available to all stakeholders.

For the sake of clarity, examples have been drawn directly from the Software Heritage archive; however, it is important to note that systems for the persistent archival of software artifacts, as well as resolution of SWHIDs, are outside the scope of this specification, which does not require the use of Software Heritage.

Information technology — SoftWare Hash IDentifier (SWHID) Specification V1.2

1 Scope

This specification defines a standard data format for referencing software artifacts that match the data model of modern distributed version control systems.

This format includes the typical tree-like structure of a filesystem hierarchy, but also, special nodes to track revisions and releases, as well as the full status of a version control system, with all its development branches.

A key property of SWHIDs is that they can be computed using cryptographically strong functions directly from the digital objects they refer to, by anyone that has access to a copy of those objects. This enables decentralised and independent verification of integrity, without relying on a registry or a central authority.

The computation of the SWHID identifiers is based on Merkle Acyclic Directed Graphs, a natural generalization of Merkle trees.

The resolution of SWHIDs, that is, the process of obtaining a copy of a digital artifact corresponding to a given SWHID, is outside the scope of this specification.

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.

RFC-3174, *US Secure Hash Algorithm 1 (SHA1), The Internet Society Network Working Grouphttps://tools.ietf.org/html/rfc3174*

RFC-3986, *Uniform Resource Identifier (URI): Generic Syntax, The Internet Society Network Working Grouphttps://tools.ietf.org/html/rfc3986*

RFC-3987, *Internationalized Resource Identifiers (IRIs), The Internet Society Network Working Grouphttps://tools.ietf.org/html/rfc3987*

RFC-5234, *Augmented BNF for Syntax Specifications: ABNF, The Internet Society Network Working Grouphttps://tools.ietf.org/html/rfc5234*