INTERNATIONAL STANDARD

ISO/IEC 23681

First edition
2019-05

Information technology — Self-contained Information Retention Format (SIRF) Specification

Reference number
ISO/IEC 23681:2019(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC 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) or the IEC list of patent declarations received (see http://patents.iec.ch).

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 SNIA (as SIRF Specification V1.0) 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.

Introduction

Many organizations now have a requirement to preserve and maintain access to large volumes of digital content indefinitely into the future. Regulatory compliance and legal issues require preservation of email archives, medical records and information about intellectual property. Web services and applications compete to provide storage, organization and sharing of consumers' photos, movies, and other creations. And many other fixed-content repositories are charged with collecting and providing access to scientific data, intelligence, libraries, movies and music. A key challenge to this need is the creation of vendor-neutral storage containers that can be interpreted over time.

Archivists and records managers of physical items such as documents, records, etc., avoid processing each item individually. Instead, they gather together a group of items that are related in some manner — by usage, by association with a specific event, by timing, and so on — and then perform all of the processing on the group as a unit. The group itself may be known as a series, a collection, or in some cases as a record or a record group. Once assembled, an archivist will place the series in a physical container (e.g., a file folder or a filing box of standard dimensions), mark the container with a name and a reference number and place the container in a known location. Information about the series will be included in a label that is physically attached to the container, as well as in a “finding aid” such as an online catalog that conforms to a defined schema and gives the name and location of the series, its size, and an overview of its contents.

This document proposes an approach to digital content preservation that leverages the processes of the archival profession thus helping archivists remain comfortable with the digital domain. One of the major needs to make this strategy possible is a digital equivalent to the physical container — the archival box or file folder — that defines a series, and which can be labelled with standard information in a defined format to allow retrieval when needed. Self-contained Information Retention Format (SIRF) is intended to be that equivalent — a storage container format for a set of (digital) preservation objects that also provides a catalog with metadata related to the entire contents of the container as well as to the individual objects and their interrelationship. This logical container makes it easier and more efficient to provide many of the processes that will be needed to address threats to the digital content. Easier and more efficient preservation processes in turn lead to more scalable and less costly preservation of digital content.

SIRF components, use cases and functional requirements were defined in [1] SIRF use cases and functional requirements, working draft — version 0.5a and further described in [2] "Towards SIRF: Self-contained Information Retention Format." This document goes one step further and details the actual metadata, categories and elements in the container’s catalog. The document also describes how the SIRF logical format is serialized for storage containers in the cloud and for tape based containers. The SIRF serialization for the cloud is being experimented with OpenStack Swift object storage, and the implementation is offered as open source in the OpenSIRF initiative[3].

Creating and maintaining the SIRF catalog requires executing data-intensive computations on the various preservation objects including fixity checks, data transformations. This can be done efficiently via executing computational modules — storlets — close to where the data is stored. The benefits of using storlets include reduced bandwidth (reduce the number of bytes transferred over the WAN), enhanced security (reduce exposure of sensitive data), costs savings (saving infrastructure at the client side) and compliance support (improve provenance tracking). The Storlet Engine[4] (see "Storlet Engine for Executing Biomedical Processes within the Storage System") is an engine to support such storlets computations in secure sandboxes within the storage system, and can be used to create and maintain SIRF containers.

Information technology — Self-contained Information Retention Format (SIRF) Specification

1 Scope

This document specifies the Self-contained Information Retention Format (SIRF) Level 1 and its serialization for LTFS, CDMI and OpenStack Swift.

This document proposes an approach to digital content preservation that leverages the processes of the archival profession thus helping archivists remain comfortable with the digital domain.

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/IEC 17826:2012, *Information technology — Cloud Data Management Interface (CDMI)*

ISO/IEC 20919:2016, *Information technology — Linear Tape File System (LTFS) Format Specification*

Self-contained Information Retention Format (SIRF) use cases and functional requirements, working draft — version 0.5a, SNIA, September 2010, http://www.snia.org/tech_activities/publicreview/SIRF_Use_Cases_V05a_DRAFT.pdf

JSON. ECMA-404, The JSON Data Interchange Standard. http://json.org

OpenSIRF. http://github.com/opensirf

OpenStack Swift. http://swift.openstack.org

PREMIS. PREservation Metadata: Implementation, Strategies, http://www.loc.gov/standards/premis

Rabinovici-Cohen S., Henis E., Marberg J., Nagin K. Storlet Engine for Executing Biomedical Processes within the Storage System", Proceedings of the 7th International Workshop on Process-oriented Information Systems in Healthcare (ProHealth), September 2014, Eindhoven, the Netherlands

Rabinovici-Cohen S., Baker M.G., Cummings R., Fineberg S., Marberg J. Towards SIRF: Self-contained Information Retention Format", Proceedings of the Annual International Systems and Storage Conference (SYSTOR), May 30-June 1, 2011, Haifa, Israel. https://www.research.ibm.com/haifa/projects/storage/datastores/papers/systor56-rabinovici-cohen.pdf

Rabinovici-Cohen S., Cummings R., Fineberg S. Self-contained Information Retention Format for Future Semantic Interoperability", Proceedings of the 4th International Workshop on Semantic Digital Archives (SDA), September 2014, London, UK

W3C Prov Model Primer http://www.w3.org/TR/prov-primer/

