 **ISO/IEC JTC 1/SC 29/ WG 11 N19564**

**ISO/IEC JTC 1/SC 29/WG 11**

**Coding of moving pictures and audio**

**Convenorship: Japan (JISC)**

|  |  |
| --- | --- |
| **Document type:** | Approved WG 11 document |
| **Title:** | Final Requirements for ISO/IEC 23092 series  extensions |
| **Status:** | Approved |
| **Date of document:** | 2020-07-03 |
| **Source:** | Convenor, ISO/IEC JTC 1/SC 29/WG 11 |
| **No. of Pages:** | 5 |
| **Email of acting convenor** | ostermann@tnt.uni-hannover.de |
| **Committee URL:** | <http://isotc.iso.org/livelink/livelink/open/jtc1sc29> |

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N19564**

**Geneva, Virtual – July 2020**

**Source: Requirements**

**Final Requirements for ISO/IEC 23092 series extensions**

Table of Contents

[1 Introduction Executive summary 3](#_Toc44677204)

[2 MPEG documents 3](#_Toc44677205)

[3 Terminology 3](#_Toc44677206)

[3.1 ISO/IEC 23092 (MPEG-G) hierarchy 3](#_Toc44677207)

[3.2 Relevant data types 3](#_Toc44677208)

[4 Requirements 3](#_Toc44677209)

[4.1 Low Complexity Coding Mode 3](#_Toc44677210)

[4.2 Indexing in the Compressed Domain 4](#_Toc44677211)

[4.3 Transport 4](#_Toc44677212)

# Introduction Executive summary

This document lists the requirements that responses to the Final Joint Call for Proposals for extensions and improvements of ISO/IEC 23092 series shall meet.

# MPEG documents

All MPEG documents cited in this document are available on the MPEG documents repository (<http://wg11.sc29.org/>) as output documents issued during past MPEG meetings.

# Terminology

Definitions for most concepts appearing in the following text, and concerning the production and analysis of sequencing data, can be found in the Terms and Definitions sections of ISO/IEC 23092 (MPEG-G) Parts 1, 2 and 3.

## ISO/IEC 23092 (MPEG-G) hierarchy

Sequencing reads of different lengths (which depend on the technology used to generate them) can be localized at one or more points on the DNA molecule they originate from. They are the basic tokens of information at the foundation of all high-level biological experiments based on sequencing. It is hence only natural for the ISO/IEC 23092 (MPEG-G) hierarchy to be based on reads, which get organized in terms of, from bottom to top, Access Units, Datasets, and Dataset Groups. The proposed technology should be as compatible as possible with the current hierarchy and components of ISO/IEC 23092 series because the objective is to create a new profile providing higher performance and new functionality.

## Relevant data types

All data types supported by the ISO/IEC 23092 series are relevant for this call.

# Requirements

This section introduces the definition of the requirements that an appropriate genomic information representation should meet when the relevant data types are considered. In addition, support for transport should also be provided – the relevant requirements apply to all kinds of data, and are considered in a separate section.

## Low Complexity Coding Mode

|  |  |  |
| --- | --- | --- |
| **Req ID** | **Requirements** | **Rationale** |
| 4.1.1 | The new proposed technology shall represent data in compressed form and offering significantly faster encoding or decoding performance than the current ISO/IEC 23092 series for at least one data item or type of sequencing data. | To improve encoding and decoding speed for some type of data and use cases. |
| 4.1.2 | The solution shall be compatible and interoperable with the ISO/IEC 23092 series. | To enable integration with the existing parts of ISO/IEC 23092. |
| 4.1.3 | Lossless compression of nucleotides sequences supporting a minimum of 5 symbols (A, C, G, T, N) |  |
| 4.1.4 | Lossless compression of quality scores shall be supported. |  |
| 4.1.5 | Lossy compression of quality scores shall be supported. Lossy compression should respect metrics that MPEG needs to identify with their associated boundaries. |  |
| 4.1.6 | Compressed data shall be structured so that parallel processing is enabled and compression efficiency is preserved | This is intended to enable efficient parallel processing of data without affecting compression efficiency. |
| 4.1.7 | Compressed data should be structured so that efficient querying of data is enabled and compression efficiency is preserved | Clustering applications could benefit from efficient querying. |
| 4.1.8 | Compressed data streaming shall be supported. | This implies that data consumption shall be possible before data transfer completion. |
| 4.1.9 | The association among headers, nucleotides and quality scores must be preserved |  |
| 4.1.10 | The association among headers, nucleotides and quality scores must be preserved |  |

## Indexing in the Compressed Domain

|  |  |  |
| --- | --- | --- |
| **Req ID** | **Requirement** | **Rationale** |
| 4.2.1 | Capability of performing queries in the compressed domain. | To improve the capability of finding specific sequencing information for both aligned and unaligned sequencing data. |
| 4.2.2 | Efficiency in retrieving data | To provide efficient selective ways to extract the data identified by the searches |
| 4.2.3 | Compression performance | To provide indexing capabilities in the compressed domain without degrading the compression performance |
| 4.2.4 | The solution shall be compatible and interoperable with the ISO/IEC 23092 series. | To enable integration with the existing parts of ISO/IEC 23092. |

## Transport

|  |  |  |
| --- | --- | --- |
| **Req ID** | **Requirement for both low complexity coding mode and indexing in the compressed domain** | **Rationale** |
| 4.3.1 | The compression process shall support the assessment of integrity. | Integrity check shall be possible by providing appropriate information. |
| 4.3.2 | The solution shall allow conveying information enabling data protection. | Ability to prevent unauthorized access shall be available. Information needed for protection of data (control for access, modification, publication, etc.) shall be conveyed. |
| 4.3.3 | The solution shall allow conveying information enabling accountability and traceability. | Data access and manipulation shall be traceable together with the identity of parties having access to data.  Information on how to verify integrity and authenticity of the data shall be conveyed. |
| 4.3.4 | The solution shall allow conveying information enabling transparency. | How and for which purpose the information is used shall be known. Usage restriction shall be applicable to the data. |
| 4.3.5 | The solutions shall support compressed data streaming. | This implies that data consumption shall be possible before data transfer completion. |
| 4.3.6 | The solution shall be interoperable with the ISO/IEC 23092 series. | To enable integration with the existing parts of ISO/IEC 23092. |