**ISO/IEC JTC 1/SC 29/WG 11  
Coding of moving pictures and audio  
Convenorship: UNI (Italy)**

**Document type: Approved WG 11 document**

**Title: Evaluation Procedure for the Call for Proposals for ISO/IEC 23092-6**

**Status: Approved**

**Date of document: 2019-07-12**

**Source: WG 11**

**Expected action:**

**No. of pages: 10**

**Email of convenor: leonardo@chiariglione.org**

**Committee URL: http://mpeg.chiariglione.org**

**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 N18649**

**Gothenburg, SE – July 2019**

**Source: Requirements**

**Evaluation Procedure for the Call for Proposals for ISO/IEC 23092-6**

Table of Contents

[1 Introduction 3](#_Toc13834829)

[2 Encoded bitstreams 3](#_Toc13834830)

[3 Evaluation Procedure 4](#_Toc13834831)

[3.1 General functionality 4](#_Toc13834832)

[3.2 Mapping statistics 4](#_Toc13834833)

[3.3 Quantitative browser tracks 5](#_Toc13834834)

[3.4 Variants 5](#_Toc13834835)

[3.5 Genome functional annotations 6](#_Toc13834836)

[3.6 Expression values 7](#_Toc13834837)

[3.7 Hi-C-like experiments 8](#_Toc13834838)

[3.8 Transport 9](#_Toc13834839)

[4 Selection criteria 9](#_Toc13834840)

[5 Annex A – Template for the self-assessment of submissions 10](#_Toc13834841)

[6 References 10](#_Toc13834842)

# Introduction

The Evaluation Procedure described in this document will be used to assess proposals in response to the Joint Call for Proposals for ISO/IEC 23092-6 Coding of Genomic Annotations [1]. Some of the terms used in this document are defined in the “Terminology” section of the Requirements document for the Joint Call.

This assessment will identify both the proposal which will become part of the Test Model 0 (TM0) and other proposals from which a set of technologies will be possibly added to or integrated with the TM0. The addition and integration will be achieved through Core Experiments that will be specified by dedicated MPEG documents. A timetable and the procedure for the verification process are specified in the Call for Proposals.

It should be noted that the procedure described above implies that proposals will be evaluated whether they fulfill requirements entirely or partially, but also in the case that they provide technology covering only a subset of the requirements.

This Evaluation Procedure will be used by:

• Proponents of technology/solutions for coding genomic annotations, both in answering to the Call for Proposals, and in self-assessing their proposal before submitting it;

• MPEG, to validate the self-assessment of the proponents and proceed to the evaluation of the proposed technologies.

The requirements defined in the Call for Proposals are divided into major needs (specified by a “shall” in the requirement text) and objectives (specified by a “should” in the requirement text). Major requirements shall be met by TM0 while objectives are less stringent even if highly desirable.

# Encoded bitstreams

Proponents are expected to supply encoded bitstreams obtained from a subset of the data listed in [2].

|  |  |  |
| --- | --- | --- |
| **ID** | **Input Item** | **Encoded bitstream name** |
| 01 | First 4,000,000 lines of item 07 in the sequencing data collection | MPEG\_ST\_SDC07\_mXXXXX.bit |
| 02 | sample-2-10\_sorted.bam from item 09 in the sequencing data collection | MPEG\_ST\_SDC09\_mXXXXX.bit |
| 03 | Item 13 in the sequencing data collection | MPEG\_ST\_SDC13\_mXXXXX.bit |
| 04 | Item 15 in the sequencing data collection | MPEG\_ST\_SDC15\_mXXXXX.bit |
| 05 | chr1 and chr2 from item 30 in the sequencing data collection | MPEG\_ST\_SDC30\_mXXXXX.bit |
| 06 | Item 1.2.3 in the annotations data collection | MPEG\_TR\_ ADC123\_mXXXXX.bit |
| 07 | Item 1.3.1 in the annotations data collection | MPEG\_VR\_ADC131\_mXXXXX.bit |
| 08 | Item 1.3.2 in the annotations data collection | MPEG\_VR\_ADC132\_mXXXXX.bit |
| 09 | Item 1.3.3 in the annotations data collection | MPEG\_VR\_ADC133\_mXXXXX.bit |
| 10 | Item 1.3.5 in the annotations data collection | MPEG\_VR\_ADC135\_mXXXXX.bit |
| 11 | Item 1.3.6 in the annotations data collection | MPEG\_VR\_ADC136\_mXXXXX.bit |
| 12 | Item 1.3.7 in the annotations data collection | MPEG\_VR\_ADC137\_mXXXXX.bit |
| 13 | Item 1.1.3 in the annotations data collection | MPEG\_AN\_ADC113\_mXXXXX.bit |
| 14 | Item 1.1.6 in the annotations data collection | MPEG\_AN\_ADC116\_mXXXXX.bit |
| 15 | Item 3.1.2 in the annotations data collection | MPEG\_AN\_ADC312\_mXXXXX.bit |
| 16 | Item 4.1.1 in the annotations data collection | MPEG\_AN\_ADC411\_mXXXXX.bit |
| 17 | Item 1.4.15 in the annotations data collection | MPEG\_EX\_ID1415\_mXXXXX.bit |
| 18 | Item 1.5.15 in the annotations data collection | MPEG\_EX\_ID1515\_mXXXXX.bit |
| 19 | Item 1.5.16 in the annotations data collection | MPEG\_EX\_ID1516\_mXXXXX.bit |
| 20 | Item 1.5.12 in the annotations data collection | MPEG\_HC\_ID1512\_mXXXXX.bit |

Table 1 - IDs of input data to be encoded and naming policy for encoded bitstreams

Notes:

1. In case an input item is composed of multiple files, proponents are required to submit a single encoded bitstream.
2. mXXXXX is the registration number of the MPEG input contribution related to the bitstreams submitted by the proponent.
3. All necessary reference assemblies are listed in the MPEG-G Genomic Information Database document [2].

# Evaluation Procedure

This section describes the test cases that will be performed to validate the fulfillment of the requirements listed in the Call for Proposals [3].

## General functionality

Every submission shall clearly state:

1. which data types are supported
2. if inconsistently formatted portions of the input data were skipped during the coding process

## Mapping statistics

This section describes the process to assess the coding efficiency of submitted solutions for mapping statistics.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Input Items** | **Test case steps** | **Reqs** |
| 1.1 | 01, 02, 03, 04, 05 | 1. Provide a list of encoded statistics and how they were computed 2. Calculate the statistics on the original file 3. Decompress encoded bitstream 4. Compare the decoded statistics with the computed ones | 4.1.1 |

## Quantitative browser tracks

This section describes the process to assess the coding efficiency of submitted solutions for quantitative browser tracks.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Input Items** | **Test case steps** | **Reqs** |
| 2.1 | 06 | 1. List available tracks 2. List available resolution levels per track 3. Decompress encoded bitstream at    1. all available resolutions    2. single resolution 4. Compare the decoded values with the input data for    1. all available resolutions    2. a single resolution 5. Compare compression/decompression    1. memory requirements    2. algorithmic complexity 6. Decompress coded information associated to any arbitrary genomic interval 7. Compute new resolution levels from compressed data 8. Decode metadata associated with each track | 4.2.1, 4.2.2, 4.2.3, 4.2.4, 4.2.5 |

## Variants

This section describes the process to assess the coding efficiency of submitted solutions for genomic variants.

The input items to be used for the assessment are: 07, 08, 09, 10, 11, 12.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Short name** | **Test case steps** | **Reqs** |
| 3.1 | Compression of metadata | 1. Decompress the metadata associated to the coded bitstream by separating information such as that contained in the VCF header from other metadata 2. Compare it with the test item header 3. Decode the other associated metadata | 4.3.2 |
| 3.2 | Compression of data | Decompress the coded bitstream by extracting the variant data associated to   1. The genomic interval (typically encoded in VCF as fields CHROM, POS) 2. The variant (typically encoded in VCF as fields ID, ALT, REF) 3. The variant likelihood and frequency for the population and samples (typically encoded in VCF as fields QUAL, FILTER, INFO and additional sample specific fields) 4. The phasing information 5. Additional information associated to the variant for the population and samples (typically encoded in VCF as fields INFO and FORMAT and additional sample specific fields) | 4.3.1, 4.3.9, 4.3.10 |
| 3.3 | Use of 23092 | Document how interoperability with other parts of 23092 is provided | 4.3.3 |
| 3.4 | Search variants | 1. List the available search criteria 2. Perform search on the compressed bitstreams according to the listed criteria 3. Display all the information related to the searched variants including    1. identifiers in terms of external variants databases    2. human readable description of the variant    3. localization of the variant on the reference genome    4. frequency and likelihood at population and sample level    5. additional information associated with the variant at population and sample level | 4.3.4, 4.3.5, 4.3.6, 4.3.7, 4.3.8, 4.3.9 |

## Genome functional annotations

This section describes the process to assess the coding efficiency of submitted solutions for genome functional annotations.

The input items to be used for the assessment are: 13, 14, 15, 16.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Short name** | **Test case steps** | **Reqs** |
| 4.1 | Compression of functional annotations | 1. Decompress the compressed bitstream 2. Compare it with the test item | 4.4.1, 4.4.2 |
| 4.2 | Use of 23092 | Document how interoperability with other parts of 23092 is provided | 4.4.3 |
| 4.3 | Search annotations | 1. List the available search criteria 2. Perform a search on the compressed bitstream according to the listed criteria | 4.4.4 |
| 4.4 | Identify annotations | Upon performing a search according to test 4.3 display all the information related to the searched annotations including |  |
| 4.4.1 | database information | 1. unique identifiers of the feature within one or more databases and the corresponding databases identifiers | 4.4.5 |
| 4.4.2 | human readable description | 1. human readable description of the feature | 4.4.6 |
| 4.4.3 | localization | 1. localization of the feature on the reference genome | 4.4.7 |
| 4.4.4 | links to other features | 1. other features associated to each feature | 4.4.8 |
| 4.4.5 | ontologies | 1. description in terms of biological ontologies and the corresponding database identifiers | 4.4.9 |
| 4.4.6 | Additional information | 1. additional information related to the specified annotation | 4.4.10 |

## Expression values

This section describes the process to assess the coding efficiency of submitted solutions for expression values.

The input items to be used for the assessment are 17, 18 and 19.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Short name** | **Test case steps** | **Reqs** |
| 5.1 | Compression of expression values | 1. Decompress the compressed bitstream 2. Compare it with the test item | 4.5.1 |
| 5.2 | Search expression values | 1. List the available search criteria 2. Perform a search on the compressed bitstream according to the listed criteria | 4.5.2 |
| 5.3 | Identify expression values | Upon performing a search according to test 5.2 display all the information related to the searched expression values including |  |
| 5.3.1 | database information | 1. unique identifiers of the feature within one or more databases and the corresponding databases identifiers | 4.5.3 |
| 5.3.2 | human readable description | 1. human readable description of the feature | 4.5.4 |
| 5.3.3 | localization | 1. localization of the feature on the reference genome | 4.5.5 |
| 5.3.4 | ontologies | 1. description in terms of biological ontologies and the corresponding database identifiers | 4.5.6 |
| 5.3.5 | expression values | 1. one or more numerical expression values | 4.5.7 |
| 5.3.6 | additional information | 1. additional information related to the specified feature | 4.5.8 |
| 5.4 | Use of 23092 | Document how interoperability with other parts of 23092 is provided | 4.5.9 |
| 5.5 | Matrices | 1. Require access to a set of expression values 2. Compare the decoded values with the input item values | 4.5.10 |

## Hi-C-like experiments

This section describes the process to assess the coding efficiency of submitted solutions for hi-C like experiments.

The input item to be used for the assessment is no. 20.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Short name** | **Test case steps** | **Reqs** |
| 6.1 | Compression of matrices at multiple resolutions | 1. List available resolution levels 2. Decompress encoded bitstream at    1. all available resolutions    2. single resolution | 4.6.1, 4.6.2 |
| 6.2 | Selective access | 1. List the available search criteria 2. Perform a search on the compressed bitstream according to the listed criteria | 4.6.3 |
| 6.3 | Additional information | Upon performing a search according to test 6.2 decode all the information related to the selected data. | 4.6.4 |
| 6.4 | Use of 23092 | Document how interoperability with other parts of 23092 is provided | 4.6.5 |

## Transport

This section describes the process to assess the efficiency of submitted solutions for the transport of coded genomic annotation.

The input items to be used for the assessment are: 01, 02, 06, 08, 13, 17, 19.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test case ID** | **Short name** | **Test case steps** | **Reqs** |
| 7.1 | Integrity | The proponents are required to document how and to what extent the solution meets the requirement | 4.7.1 |
| 7.2 | Protection | The proponents are required to document how and to what extent the solution meets the requirement | 4.7.2 |
| 7.3 | Accountability and traceability | The proponents are required to document how and to what extent the solution meets the requirement | 4.7.3 |
| 7.4 | Transparency | The proponents are required to document how and to what extent the solution meets the requirement | 4.7.4 |
| 7.5 | Streaming | The proponents are required to document how and to what extent the solution meets the requirement | 4.7.5 |
| 7.6 | Use of 23092 | The proponents are required to document how and to what extent the solution meets the requirement | 4.7.6 |

# Selection criteria

Criteria to rank technologies overall and per type of data will be:

1. requirements coverage
2. support for random access in terms of:
   1. types supported
   2. granularity of random access for each supported type
   3. flexibility in setting granularity levels for the different types of random access
   4. overall bitrate overhead vs. granularity
3. size of the produced bitstreams
4. support for transport in terms of:
   1. functionality and flexibility of use
   2. computational complexity
   3. overall bitrate overhead
   4. possible impact on transport latency (if relevant for the application scenarios)
5. use of technology specified in other parts of ISO/IEC 23092
6. interoperability with legacy formats

The proponents are required to characterize the solution on the points above and provide evidence (code, examples, statistics, etc.) of the results.

*Nota bene*: measured performance such as:

1. computational complexity
2. processing time
3. memory requirements

might be used to further rank proposals.

# Annex A – Template for the self-assessment of submissions

An Excel form is annexed to this document in order to enable respondents to self-assess the technology submitted as answer to this CfP.

# References

|  |  |
| --- | --- |
| [1] | ISO/IEC JTC 1/SC 29/WG 11, “N18648 - Call for Proposals for ISO/IEC 23092-6 Coding of Genomic Annotations,” Gothenburg, 2019. |
| [2] | ISO/IEC JTC 1/SC 29/WG 11, “N18645 - MPEG-G Genomic Information Database,” Gothenburg, 2019. |
| [3] | ISO/IEC JTC 1/SC 29/WG 11, “N18647 - Requirements for ISO/IEC 23092-6 Coding of Genomic Annotations Call for Proposals,” Gothenburg, 2019. |