

AHG on Genomic Information Representation (m49773)

ISO/IEC 23092 (MPEG-G)

M. Golebiewsky (HITS), J. Delgado (UPC),

M. Mattavelli (EPFL)

Joint AHG with ISO/IEC TC276

Mandates

1. To disseminate the information and coordinate the activities jointly carried out by ISO/IEC JTC 1/SC 29/WG 11 and ISO TC 276/WG 5 to other ISO TCs, GA4GH and other organizations.
2. To discuss and coordinate a collaboration with GATK development team to support MPEG-G as native format.
3. To discuss and coordinate WG 11 contributions to relevant bioinformatics conferences and events.
4. To contribute to the editing and to the revision of FDIS of Part-3 and for the draft text of DIS of Part-4 and Part-5.
5. To finalize the collection and definition of test item descriptions and binary streams for Conformance testing.
6. To revise requirements and call text for Part 6 Coding of Genomic Annotations, to identify use cases and requirements for MPEG-G Version 2.
7. To collect test data to be used for Part 6 call, for the evaluation of the performance of proposed technologies and for future standardization activities.
8. To investigate anonymization methods for genomic data and understand how MPEG-G can support their implementation.

AHG Activity Overview

- **Work mainly focused on:**
 - Development of Part 4 reference software for both Part 1 and Part 2
 - Generation of bitstreams for Part 5 Conformance
 - Reviews of input documents
 - Revision of Part 6 Call for proposals and related documents

AHG Activity Overview

- **Coordination activities with TC276 and other ISO committees:**
 - Meeting in Tokyo in June 2019
 - Document presented and discussed
 - Great interest and support for the call for Part 6
 - Received feedback and proposition of revision for the call text and requirement document

AHG Activity Overview

- **Dissemination and promotion of MPEG-G:**
 - Special session on Omics Data Processing and Analysis organized at IEEE Data Science Workshop 2019. The Conference was held in June 3, 2019 in Minneapolis USA
 - Special session on Omics Data Compression organized at ISMB/ECCB'19 jointly with EBI. To happen in Basel on July 25th. The special session to cover the current standards for genomic information representation as well as compression efforts on other omics data types.

AHG Activity Overview

- **8 Input documents:**

MPEG number	Title	Source
m48100	Proposed Updates to the MPEG-G Genomic Information Database	Jan Voges , Idoia Ochoa , Mikel Hernaez , Giorgio Zoia , Marco Mattavelli
m48850	Report on the generation of Conformance bitstreams	Giorgio Zoia
m49078	Additional test items for part 6 call	Paolo Ribeca, Giuseppe Codispoti, Massimo Ravasi, Claudio Alberti
m49167	Survey of the regulations of genomic information in different countries, demands of security and privacy and technological tools, for the application guide for MPEG-G - Genomic Information Representation	Itaru Kaneko - Nagoya City University , Emi Yuda - Tohoku University Graduate School of Engineering,
m49171	Possible demand in privacy-related functionalities of MPEG-G and some thought of the coding technologies for the intermediate exchange format of anonymized genomic information based on MPEG-G	Itaru Kaneko - Nagoya City University , Emi Yuda - Tohoku University Graduate School of Engineering
m49345	Revised Draft Call for Proposals for Coding of Genomic Annotations	Martin Golebiewski, Marco Mattavelli
m49353	MPEG-G Best Practices Deployment Guide	Dunling Li, Claudio Alberti
m49439	Considerations for the Call for Proposals for Coding of Genomic Annotations	Jaime Delgado , Daniel Naro , Mikel HernÃ¡ez , Idoia Ochoa , Raymond Krasinski , Patrick Cheung

AHG Activity Overview

- Addition of new test data for:
 - Main MPEG dataset
 - Dataset for Part 6 Call
- New inputs for data protection and best practice document

AHG Activity Overview

- **Progress of Part 4 Reference Software:**
 - Part 1: about 96 % completed
 - Functionality still missing (selective access, raw reference: expected to be completed by the end of the meeting, offset to be completed within DIS editing period)
 - Part 2: about 96 % completed
 - 29 conformance bitstreams are correctly decoded
 - New bitstreams of MPEG-G will be under test soon
 - Missing modules: 3 types computed references (not completed, to be completed within editing period)
- **Reference SW repository**
 - Migration to MPEG official repository to be completed with DIS submission

- **Status of Part 5 Conformance**

- Conformance of Part 1

- Bitstreams available and cross checked **24**
 - Expected to be available after editing period **2**

- Conformance of Part 2

- Bitstreams available and cross checked **45**
 - To be done within editing period **10**

- **Part 6 related input documents:**
 - Issue the Call at this meeting, but with a 6 month period for answering
 - Encourage type 1 solutions according to ISO rules
 - Remind to provide IP declarations statements when answering the call
 - Clarify “shall” and “should” meaning to make them clear to submitters not familiar with standardization formalities

Recommendations

- Complete the reference SW development work and promote Part 4 to DIS stage
- Update the database with new test items as suggested by input proposals to this meeting
- Complete the verification of available bitstreams and promote Part 5 to DIS
- Issue a corrigendum for Part 2
- Finalize the call for contributions and related support documents for ISO/IEC 23092-6 Annotations

- Internal MPEG-G Genomic Information Database (Jan Voges)
- MPEG-G Genomic Information Database (Jan Voges) (Public)
- Text of ISO/IEC DIS 23092-4 Reference Software (Massimo Ravasi)
- Text of ISO/IEC DIS 23092-5 Conformance (Giorgio Zoia)
- Requirements for 23092-6 Coding of Genomic Annotations (Paolo Ribeca) (Public)
- Call for proposals for 23092-6 Coding of Genomic Annotations (Paolo) (Public)
- Evaluation procedure for the answers to the Call for proposals for Part 6 (To be assigned)
- Text of Cor 1 of 23092-2 Coding of Genomic Information (Editors to be assigned)