

AHG on Genomic Information Representation (m46633)

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 MPEG contributions to relevant bioinformatics conferences and events.
4. To contribute to the editing and to the revision of the FDIS text of Part-1, Part-2, and DIS of Part-3 and for the CDs 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 identify work items, use cases and requirements for MPEG-G Version 2.
7. To collect test data for future standardization activities.
8. To investigate anonymization methods for genomic data and understand how MPEG-G can support their implementation.

- **Work mainly focused on:**
 - Development of Part 4 reference software
 - Generation of Conformance bitstreams
 - Revision of Part 3 DIS text and analysis of the NB comments received with DIS ballot results
 - Reviews of input documents

AHG Activity Overview

- **Coordination activities with TC276 and other ISO committees:**
 - No meetings after Potsdam in December 2018
 - Next meeting Tokyo in June 2019

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 is scheduled from June 2-5, 2019 in Minneapolis USA
 - Special session on Omics Data Compression organized at ISMB/ECCB'19 jointly with EBI. To happen in Basel 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

- **11 Input documents, 26 participants AHG meeting:**

MPEG number	Title	Source
m46657	Study on ISO/IEC DIS 23092-3	Junaid Ahmad
m46661	Proposal for inclusion of single cell datasets for MPEG-G benchmarking and future development of MPEG-G.	Tom Paridaens , Jan Fostier, Yvan Saeys
m46665	Study on ISO/IEC DIS 23092-3	Claudio Alberti , Giorgio Zoia, Daniele Renzi, Giuseppe Codispoti, Massimo Ravasi
m47166	Current status of MPEG-G reference software implementation	massimo.ravasi@epfl.ch, dnaro@ac.upc.edu, voges@tnt.uni-hannover.de, junaid@jnomics.com, tom.paridaens@ugent.be
m47220	Study on ISO/IEC 23092-5 CD	Giorgio Zoia
m47222	Draft text proposal for ISO/IEC FDIS 23092-3 Genomic Information Metadata and APIs	Jaime Delgado , Daniel Naro , Silvia Llorente
m47381	Proposed additional formats to be considered for future standardization activities in genome information representation	Giuseppe Codispoti, Claudio Alberti, Marco Mattavelli, Massimo Ravasi
m47411	MPEG-G Best Practices Deployment Guide	Dunling Li , Claudio Alberti
m47435	Report on the generation of Conformance bitstreams	Giorgio Zoia , Daniele Renzi
m47841	Survey and comparison of the demands of security and privacy in different countries and technological tools relate to MPEG-G - Genomic Information Representation	Itaru Kaneko, Emi Yuda
m46660	Proposed Updates to the MPEG-G Genomic Information Database	Jan Voges , Idoia Ochoa , Mikel Hernaez , James Bonfield

- **Part 3 related input documents:**
 - Result of ballot and about 275 NB comments
 - 3 Study documents of current DIS

AHG Activity Overview

- **Progress of Part 4 Reference Software:**
 - Part 1: about 90 % completed
 - A few functionality still missing (not major development needed)
 - Part 2: about 90 % completed
 - 18 conformance bitstreams are correctly decoded
 - New bitstreams of MPEG-G will be under test soon (multiple alignment)
 - Missing modules: computed references (not completed) , multiple alignments (not tested), multi parameter set support (not implemented)
- **Reference SW repository**
 - Migration to MPEG official repository soon to be completed

- **Status of Part 5 Conformance**
 - Conformance of Part 1
 - Bitstreams available **0**
 - Expected to be soon ready **25**
 - Conformance of Part 2
 - Bitstreams available and cross checked **18**
 - Expected to be soon ready **28**
 - Need to be generated by July **11**

- **Part 6 related input documents:**
 - M47381: Proposed additional formats to be considered for future standardization activities in genome information representation
 - Analysis of annotation formats that could be included in MPEG-G files in compressed form and that provide a normative link to MPEG-G compressed sequence data/metadata.

Recommendations

- Finalize answers to NB comments on Part 3 DIS ballots
- Promote Part 3 Metadata and APIs to FDIS
- Promote Part 4 Reference Software to DIS
- Promote Part 5 Conformance to DIS
- Proceed with the reference SW development work and update the plan
- Update the database with new test items as suggested by inputs proposal to this meeting
- Finalize a call for contributions on Part 6 Annotations

- Internal MPEG-G Genomic Information Database (Jan Voges)
- MPEG-G Genomic Information Database (Jan Voges) (Public)
- Test data for future standardization activities in the area of genomic information representation (Jan Voges)
- Text of ISO/IEC FDIS 23092-3 Genomic Information Metadata and Application Programming Interfaces (APIs) (Jaime Delgado)
- Disposition of comments on ISO/IEC 23092-3 DIS (Jaime Delgado)
- Text of ISO/IEC DIS 23092-4 Reference Software (Massimo Ravasi)
- Text of ISO/IEC DIS 23092-5 Conformance (Giorgio Zoia)
- Thoughts on future standardization activities in the area of Genomic Information Representation (Paolo Ribeca) (Public)
- Draft requirements for future standardization activities (Annotations) (Paolo Ribeca) (Public)
- Draft Call for technology (Paolo) (Public)
- MPEG-G Best practice and deployment (not public yet) (Dunling Li)
- Request of subdivision for Part 6
- Update Master Table of Terms and Definitions (Public)