

AHG on Genomic Information Representation (m42862)

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

M. Mattavelli (EPFL)

Joint AHG with ISO/IEC TC276

Mandates

1. To disseminate the information on the activities jointly carried out by ISO/IEC JTC 1/SC 29/WG 11 and ISO TC 276/WG 5 to other ISO TC and other organizations.
2. To discuss and organize an MPEG contribution to the GA4GH October meeting.
3. To discuss and coordinate a collaboration with GATK development team to support MPEG-G as native format.
4. To discuss and coordinate MPEG contributions to relevant bioinformatic conferences and events.
5. To contribute to the editing and to the revision of the DIS of Part 1 and Part 2, CD and DIS of Part 3 and for the CD of Part 4.
6. To finalize the collection and definition of test item descriptions and binary streams for Conformance testing
7. To identify work items for MPEG-G Version 2
8. To explore the possibility of organizing a WS on MPEG-G after the 124 MPEG meeting

AHG Activity Overview

- Technical work mainly focused around:
 - Providing and analyzing NB comments on CD of Part 3
 - Preparing the DIS text for Part 3
- Coordination activities with TC276 and other ISO committees:
 - Presentation of standardization status and collection of feedbacks at TC276/WG meeting In Beijing on the 14th of June
 - All output documents from Gwangju and San Diego approved by TC276/WG5

AHG Activity Overview

- Dissemination and promotion:
 - MPEG-GA4GH WS on “Workshop on Genomic Sequencing Data Compression” on the 3rd October in Basel during the GA4GH plenary meeting (Call can be released today)
 - Several contributions on MPEG-G technology presented to ISMB 2018 (July 6-10, Chicago)
 - MPEG-G features and standardization status presented at ECMA/ISO TC31 meeting in Geneva on the 25th June
 - Discussion on support of MPEG-G technology by GATK started
- Progress of reference SW integration:
 - Part 1, QVs, Read names, Parsing of bitstreams ,
 - Next steps: integration of CABAC, descriptors decoding

Part 3 - 5 input documents review

- 13 Input documents at this meeting, 6 on Part 3 editing, most inputs reviewed during the Sat. and Sun. AHG meeting.
- Comments from NB suggesting changes and corrections to Part 3: (M42919, M43542, M43543, M43546, M43548, M43652)
 - Review of the editorial and a few technical changes completed for about 50%, most processed NB comments approved for inclusion in the next version.
- Input for text of Part 5:
 - Good progress, but more editing will be done during the week.

Other input documents review

- MPEG-G database update: (M42920)
- Need of generic compressor for metadata text to be added to BIM for XML. A good candidate seems to be ISO/IEC 21320 (zip) (M43547)
- A new use case for the anonymization of health data in (M43376)
- A description of how to implement most used Samtools functions using MPEG-G files (M43836)
- A study of appropriate methodologies for robust encryptions of variants for MPEG-G files (M43545)

Recommendations

- Continue technical and editorial work on the DIS text of Part 3 during the week
- Progress with the editorial work and on identification and generation of conformance test items for Part 5 Conformance
- Finalize the participation plan of the joint MPEG-GA4GH WS in Basel on the 3rd of October
- Progress with the organization of the WS in Shenzhen (the Saturday after the October MPEG meeting in Macau)
- Continue the integration work on Part 4 (Reference SW for both Part 1 and Part 2)
- Continue the work on new use cases validation and the identification of new work items and corresponding solutions for version 2.
- Update the data base for new items and for conformance testing