

AHG on Genomic Information Representation (m41947)

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

M. Mattavelli (EPFL)

Joint AHG with ISO/IEC TC276

Mandates

- 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.
- To make available on line the MPEG genome data to be used for core experiments (N16726).
- To carry out the core experiments described in document N17143.
- To contribute to the editing and to the revision of the CD of Part 1 and Part 2, and for the WD of Part 3 and Part 4.
- To collect requirements for specifying the format of metadata to be included in Part 3 : MPEG-G metadata and APIs
- To finalize the collection and definition of test item descriptions and bytestreams for Conformance testing

AHG Activity

- Mainly focused around:
 - Editing the 2 CDs for Part 1 (Transport and Storage of Genomic Information) and Part 2 (Compression of Genomic Information)
 - Running the active Core Experiment:
 - CE5 entropy coding
 - No intermediate ad-hoc meeting
- Collecting feedbacks on Part 3
- Coordination activities with TC276 and other ISO committees:
 - Presentation of Torino and Macau output documents to the plenary TC276/WG5 meeting on the 29th-30th November in Rome.
- Further definition of list items for conformance
- 13 Input documents at this meeting

Summary of CE 5 “Entropy Coding”

- 1 input document on entropy coding for readnames and entropy coding:
 - Initial results show possible improvements, but more extensive and consistent results are needed.
- Conclusion/Consensus:
 - Continue CE5 with read descriptors, read name descriptors and QVs:
 - for new test items with the new descriptors of multiple alignments
 - for achieving higher compression or lower computational complexity
 - for documenting and preparing the reference encoding procedure

Joint work with TC276/WG5

- Presentation and discussion of all outputs documents from Torino and Macau MPEG meetings (all available to TC276/WG5 delegates):
 - Part 1 CD Transport and File Format
 - Part 2 CD Compression of Genomic Information
 - Part 3 WD APIs and Metadata
- All output documents have been approved by TC276/WG5

TC276/WG5 Resolutions on MPEG-G

- **“ISO/TC 276/WG5 congratulates the joint working group on genomic information compression and storage together with and under the lead of ISO/IEC JTC1/SC29/WG11 MPEG for its successful achievement of developing ISO/IEC 23092 Parts 1 and 2 to Committee Draft stage. WG5 supports the plan of the joint working group to reach FDIS stage of ISO/IEC 23092 Parts 1 and 2 within ISO/IEC JTC1/SC29/WG11 MPEG by October 2018. ISO/TC 276/WG5 experts regard all parts of ISO/IEC 23092 as important components of the standard ecosystem in biotechnology and the life sciences in general.”**

TC276/WG5 Resolutions on MPEG-G

- The current drafts of ISO/IEC 23092 parts 1, 2 and 3 have been distributed within ISO/TC 276/WG5 and all WG5 experts can provide comments on the current documents to the joint working group to support the further development of the documents.
- The contributors to the joint working group with MPEG from ISO/TC 276/WG5 are Petr Holub (BBMRI-ERIC), Jörg Geiger (Germany), Yong Zhang (China, WG5 co-convenor) and Martin Golebiewski (Germany, WG5 convenor).

Input document review

- Most of inputs reviewed during the Sat. and Sun. AHG meeting (except those on Part 2)
- Comments from NB suggesting changes to Part 1:
 - All editorial and most technical changes approved for inclusion in the next version
- Suggested changes to Part 2 for inclusion in the study document:
 - Will be discussed during the week
- Part 3 proposed changes:
 - Editorial and APIs simplifications approved for Part 3 CD
 - Extension mechanism better formalized
 - Further work on APIs expected during the week
- Status of Reference SW reviewed
- White paper improved with new content and pictures
- Proposed input for Part 5 approved for WD status

Recommendations

- Continue technical and editorial work on the CD of Part 1 and the study of part 2 during the week
- Continue CE5 to provide further technical input to Part 2 study document.
- Focus the work on metadata and APIs definition for promotion to CD at this meeting
- Continue the integration work on Part 4 (Reference SW for both Part 1 and Part 2) before promoting it to CD level
- Continue the work on identifying conformance test items