



Address 100 Barr Harbor Drive
PO Box C700
W. Conshohocken, PA
19428-2959 | USA

Phone 610.832.9500
Fax 610.832.9555
e-mail service@astm.org
Web www.astm.org

Committee E13 on MOLECULAR SPECTROSCOPY AND CHROMATOGRAPHY

Minutes for E13.15 Subcommittee Working Group

10:30 am – 12:30 pm EDT

October 13, 2006

Virtual Meeting

I. Introductions and Welcome: Gary Kramer, E13.15 Chair called the meeting to order at 10:30 am EDT. The ASTM release form for recording the session was displayed on the screen. There being no objections, the meeting was recorded.

II. Attendees:

Michael Boruta, ACD
Stuart Chalk, Univ. of N. Florida
Maren Fiege, Waters
Patrick Gleichmann, NIST
Steve Hoffman, BMS
Joe Koury, ASTM

Gary Kramer, NIST
Peter Linstrom, NIST
Dave Martinsen, ACS
Dale O'Neill, Agilent
Kordian Placzek, NIST
Burkhard Schäfer, BSSN

III. Next Meeting: November 3, 2006, 10:30 - 12:30 EDT.

IV. Minutes:

- a. June 30 Minutes - a motion was made to approve as distributed. Second. The motion was approved.
- b. July 28 Minutes - a motion was made to approve as distributed. Second. The motion was approved.
- c. August 15 Minutes - the following changes were requested: Add the following attendees: Wolfgang Winter, Michael Boruta, Alex Roth. A motion was made to approve the minutes as modified. Second. The motion was approved.
- d. September 15 Minutes: The following changes were requested: Jim Powers affiliation is Wyeth-Ayerst. Alex Roth's current affiliation is Fachhochschule Wiesbaden. A motion was made to approve the minutes as amended. Second. The motion was approved.

V. Core Developers:

- a. Discussion of the relationship between spectra and chromatograms continued. The group decided that Mark Bean and Burkhard Schäfer would meet separately and determine if the current model satisfies the requirements. That

meeting has not yet taken place. Most participants had conflicting meetings, so the core meeting was shorter than expected. Another core developer meeting is scheduled for 17 October.

VI. Technique Committees:

- a. **UV/Vis:** The scheduling of another meeting of the UV/Vis group was discussed. 26 October was selected, at 10:30 am EDT. Joe Koury will send out invitations.
- b. **Chromatography:** Dale O'Neill has been meeting weekly with Agilent scientists, analyzing the chromatography technique definition. Once they get to a reasonable point, he would like to bring it their proposal to the full committee.

VII. AnIML Documents:

- a. Some time ago, Burkhard Schäfer agreed to write the AnIML documents if funding could be found. At the time, several had agreed to check for funding, but no follow-up had been completed. Burkhard will recirculate the proposal for creating the documentation, and Gary Kramer will check on possible contributors.
- b. Gary will discuss this project with ASTM to work out details of what needs to be done.
- c. If anyone else can offer funding to support the project, let Gary know offline

VIII. The AnIML Validator:

- a. Patrick Gleichmann and Kordian Placzek have been looking at Dominik's validator and had some questions. Before they move forward, though, Gary would like to have a formal list of specifications so they know exactly what the requirements are. There was some discussion about whether one validator was needed, or whether there should be two validators, one to validate technique definition documents against the technique schema, and one to validate the AnIML instance documents. We need to be sure that the technique definition documents are valid, to avoid the problems which have occurred in the JCAMP world. Several points were discussed:
 - i. Validating technique definition documents will only need to be done when new technique definitions are created and/or extended. This will be rather infrequent. The validation will be complex, but need not be particularly fast or efficient.
 - ii. Validation of instance documents could occur every time a document is opened. This procedure must be very efficient.
 - iii. The technique definition validator will also need to validate against the NDRs. This won't be needed for the instance documents.

- iv. It was noted that you can only validate that an AnIML instance document is valid if the technique definitions and extensions are also valid.
- v. The validator for instance documents should be done first. Since this group is creating the initial technique definition documents, we will be careful to follow the NDRs (which themselves are not yet complete).
- vi. While there might be some advantages to using Dominik's validator as a starting point, it turns out that because the structure of the AnIML core has changed over the two years since the validator was written, it will probably be better to start from scratch. Dominik's approach will certainly be taken into consideration.
- vii. In designing the new validator, Patrick and Kordian will work to make it extensible, so additional tests could be performed by adding a new module or modules.
- viii. The validator should be capable of being run through a GUI as well as in a batch mode.
- ix. There was some discussion of an execution time requirement for the validator. It is difficult to predict the time at this point, especially considering the potential size of AnIML files. However, a 30 second execution time was discussed as a reasonable target for files of "reasonable" size.
- x. There was discussion about whether a validation flag could be added to a file, so one could know whether the file needed to be validated or not. This might be possible, but could the flag itself be trusted?
- xi. There could be degrees of validation. To validate on the basis of schemas, syntax, and semantics could take quite a while. If one could select only schema and syntax validation, the validation would be quicker. The user could decide which degree of validation was important in the application.
- xii. The current vision of the validator is in terms of a hierarchy of validation. If one of the validation nodes was skipped, then all subnodes would be skipped as well. Syntactic validation means that the instance document conforms to the AnIML schema. It may or may not conform to the technique definition. That might be the next level of validation.
- xiii. The validator must also be able to determine whether an instance document is completely valid or not. This committee is not in the software business, but without such a tool, it is unlikely that the standard will be adopted.
- xiv. The source code for the validator will be openly available, on SourceForge.
- xv. Some suggested requirements for a valid document:
 - 1. Well-formed
 - 2. Syntactically correct
 - 3. Correct types

4. Must have all required elements
 5. Range checking
 6. Allowed values
 7. Cardinality
 8. Sample inheritance
 9. Digital signatures
 10. Modality
 11. Vectors
 12. Start offset, end offset
 13. Base64 encoding
- xvi. Patrick envisions the validator framework as something which could be integrated into 3rd party tools, with three output modules - GUI mode, terminal mode, and an xml document. It was suggested that a valid/invalid indicator might also be useful. The current validator has two levels of problems - errors and warnings. An error was used to indicate that a file in violation of a requirement, while a warning indicated that some best practice recommendation was not followed. One scenario discussed was an xml file which made use of specific user extensions, but the user extensions were unavailable. In xml, these would be ignored. But it might be useful if the validator issued a warning about missing extensions.
- xvii. Some other issues were discussed to help Patrick and Kordian understand the reasons for some of the decisions the AnIML committee had made in the past:
1. Sample and results data are kept in the same instance files because AnIML files are intended to be used as both an archive and an interchange format. It is important to keep the metadata with the results data.
 2. Certain elements, such as sampleStep, were not made optional because in regulated industries, it's essential to have an audit trail.
 3. There are a number of elements which are enclosed by a "step" container, but the audit trail itself is not. The reason for this is that set containers are used for digital signatures. You can sign the container, or the individual items inside. If you could only sign the items, those items could be cut and pasted elsewhere, and the signatures would remain valid. The containers make sure you haven't moved things around.
 4. Some standard schema types have been overridden. For example, float is replaced by base64 encoding. The XML float type defines an ASCII representation of an IEEE format 32 bit number. In some applications, people need to preserve all the bits in the measurement.
- b. At EAS - the technical working group meeting will be held on Sunday, Nov. 12, 3-5 pm (Library Room at Doubletree), and the business meeting on

Monday, Nov 13, 4-6pm (Library Room at Doubletree). This would be a good opportunity to go over requirements for the validator. Joe Koury will check on the possibility of phone conferencing for those who can't attend. Discussion could also include the definition of a valid document, and a scenario on how the validator would be used.

- c. Next step: Patrick and Kordian will begin to write up a requirements document and post it on the sourceforge site, as well as sending it out via email.

IX. Other meetings:

- a. Because of the proximity to PittCon, and the extent of the technical presentations there, there was a decision to bypass the ACS meeting. It is only three weeks later.
- b. David Farrusseng sent an email about the EuroCombiCat 2007 meeting, 22-25 April, Bari, Italy (<http://www.eurocombicat.org/>). Burkhard is considering attending that meeting.
- c. Burkhard hasn't heard back from Lab Automation on the status of his presentation.

- X. Other Promotional Material:** In response to Gary's suggestion, Burkhard will investigate writing a new paper on AnIML, taking into account more recent activities to update the JALA article.

XI. Adjourn: 12:35 pm EDT.

Minutes prepared by David Martinsen, ACS
ASTM E13.15 Secretary