



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

November 3, 2006

Virtual Meeting

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

II. Attendees:

Mark Bean, GSK
David Farrusseng, CNRS
Maren Fiege, Waters
Patrick Gleichmann, NIST
Joe Koury, ASTM
Gary Kramer, NIST
Richard Larsen, JASCO
Dave Martinsen, ACS

Dale O'Neill, Agilent
Jim Olshefsky, ASTM
Kordian Placzek, NIST
Alexander Roth, Fachhochschule
Wiesbaden
Burkhard Schäfer, BSSN
Karen Wilson, ASTM

III. Next Meeting: December 8, 2006, 10:30 - 12:30 EST.

IV. Minutes: Several modifications were suggested for the minutes of the October 13 meeting:

- a. Under item IV.d: Jim Powers is from Wyeth-Ayerst
- b. VIII bullet 1: change “when created” to “when created, and/or extended”
- c. A question was raised as to whether we decided we needed two validators. It was noted that there were two different requirements. The speed/efficiency is a requirement for AnIML instance documents. Technique Definition validation needs to be thorough, but not necessarily speedy. No change was needed for the minutes on this point.
- d. Another discussion ensued regarding validation in the context of technique definitions that were themselves extended. Some extensions might not be able to be combined with others, some extensions may have parameters of the same name but different type, some extensions could be combined with the wrong technique. These points were not part of the original discussion reported in the minutes, and so no change was needed to the minutes for this point.

A motion was made to approve the minutes as amended. Second. The motion carried.

V. Core Developers:

Maren Fiege reported that the core group started creating a sample AnIML file. Mark Bean provided data with both chromatograms and spectra. These are being transformed manually into AnIML, but the task is not yet complete. There are some UV/Vis examples and IR examples as well. Mark Bean has been using a flat data structure at GSK, and this is what the group is using as a starting point. It has been fairly straightforward to do the conversions, although the hierarchy from chromatograms to spectra has not yet been addressed.

VI. Technique Committees:

- a. **UV/Vis:** The group has been discussing parameter and sample roles, and noticed that quite a few roles share the same properties. They have also discussed how to capture the solvent parameters. They are ready to look at the UV part next, particularly how to structure the results data. Burkhard is working on cleaning up the documents.

VII. The AnIML Validator: Patrick and Kordian have worked hard on the document, which was distributed by Gary in the meeting invitation. The document has been divided into several parts. These cover how to determine when an AnIML is valid, validating documents without a technique document, making all reference ids unique. There is some syntax in the document that will need to be changed. The use of arrows and periods will be changed to “/” and “@”, as used in XPath.

Patrick noted that the program that determines whether an AnIML document is valid will first check all of the bounds, and then will check the dates in the time stamp and in the log. In this way, it can be determined if someone changes the document without changing the log entry.

A question was raised in relation to knowing how many bits are in the AnIML data. This could be done by checking the length of the base64 string, to see if the string length matches the startOffset-endOffset range. Burkhard noted that vectorSetLength is the maximum. It can't have holes. The Cartesian product of all independent vectors is the primary key base.

Section 2.2 discusses testing the AnIML Instance Document against the AnIML Technique Definition. Any extensions must also be checked to make sure they are valid with the technique definition. There must also be a check to make sure all required components are available.

It was suggested that section 3.2.3 is not needed in this requirements document. In fact, sections 3 and 4 are more like design documents, and these will not be included in this document.

Patrick noted that the software will be created as a framework to allow integration with 3rd party applications. A demonstration will all be created, incorporating the different software components. Gary Kramer noted that in the past, it was problematic for other to load and run the distributed software, because their software environment was not identical to the development environment. We need to make sure all required components are generally available, and let people know which components might require registration.

A discussion ensued on the idea of minimal validity. Are there different levels of errors and/or warnings, or is there just pass/fail? It was noted that in some cases it may not be possible to validate a digital signature, since the user of the file may not have access to the appropriate repository - it may be internal to the company that wrote it. If the file is distributed outside that environment, the digital signature may not be able to be validated.

Whenever a condition is found that doesn't match the standard, an exception needs to be noted. Perhaps a table which says *a*, *b*, and *c* were tested, and were found to be valid. But *d* could not be validated, because some component was not available. This could lead to a case where an external user may check an AnIML instance document and find it to be valid, where an internal user may find the same document to be invalid. If there are components which can't be checked, then we can't really say that a document is valid.

This may be a problem of terminology. The designation "fully valid" was suggested to indicate that all core, technique, and extension files are available, and that the instance document is valid across all files. If some of the extensions are missing, then the instance document could be valid, but not "fully valid". It was noted that XmlSpy uses a similar approach. When it can't find special schema, it issues a warning that the schema does not appear to be valid by itself, because it couldn't find a certain schema document. This is not exactly the approach we want, but we could use it as a model.

VIII. Other meetings:

- a. UV/Vis: Dale O'Neill wants to be included on the list for this task group.
- b. The next Core Group will be the coming Tuesday.
- c. The next UV/Vis group will be next Thursday.
- d. E13.15 will have both a working group meeting and a business meeting at EAS. The working group meeting will be Sunday from 3-5 pm. The cost for a phone is \$95. Joe Koury will see if he can arrange for a phone.

IX. Other Promotional Material: Maren noted that the October/November 2006 issue of Scientific Computing World has an article by Siân Harris on standard data formats. He covers HUPO, but does not mention AnIML.

X. Adjourn: 12:40 pm EST.

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