
Committee E13 on MOLECULAR SPECTROSCOPY AND CHROMATOGRAPHY

Minutes for E13.15 Subcommittee Working Group

10:35 am – 12:30 pm EST

Friday, June 17, 2005

Virtual Meeting

I. Introductions and Welcome: Gary Kramer, NIST, E13.15 Chair at 10:35 am EDT

II. Review of Agenda. The agenda was approved as circulated. The Minutes for the May 13 meeting were amended as follows:

- Section 3: last paragraph: should read article in Spectroscopy.
- Section 9: UV/Vis based on Anh Dao's UV/Vis definition, sample is from NIST.

The minutes were approved as amended.

III. Attendees:

Tony Davies, Waters

Mohan Cashyap, GSK

Michael Harzheim, Waters

Gary Kramer, NIST

Robert Lancashire, Univ. West Indies

Richard Larsen, JASCO

Peter Linstrom, NIST

Dave Martinsen, ACS

Alex Mutin, Shimadzu

Mark Mullins, SSI

Anh Dao Nguyen, NIST

Burkhard Schäfer, Univ. Kaiserlautern

Carolyn Sheahan, ASTM

IV. Location of Schema. It was noted that other ASTM groups, in addition to E13.15, have problems with location of schemas. We don't have a draft of the standards documents; the IT staff at ASTM cannot create an ASTM location for the schema without an understanding of how it will be addressed in the text of the Standard. Mark Bean will be proposing a directory on SourceForge. Questions were raised as to whether it would be useful to have it on the IUPAC site. Discussion concerned the intellectual property rights asserted by ASTM, since the work is being developed under the auspices of ASTM. It was noted that if people can't easily access the schema (e.g., if it is "protected" on the ASTM site), or if they can't load it on their local system, they won't use it. For the time being, the schemas will be stored on SourceForge as the canonical location, but will be mirrored at IUPAC.

V. AnIML Website Redesign. The new website has not yet been moved to a more stable location for review. However, it has been available and one comment has been received. Burkhard Schäfer has been examining some simple open source content management systems to make the addition of new HTML pages within a template environment more straightforward. Dave Martinsen will follow up with Stuart Chalk, to see if he has had anyone "simplify" the HTML encoding, with the goal of having the new site deployed in the next month.

VI. Formal AnIML Requirements Document. There was no activity to report in this area. Alex Mutin reported that the transcript of the worksheets from the organizing meeting in September, 2002, has been

prepared in Word, and will be distributed to the task group working on the requirements document. Tony Davies will be added to that group.

VII. Core Developer's Meeting. Burkhard reported that the core group met and discussed two main issues – multiple analyses and multiple techniques. The group is investigating whether the concept of analysis needs to be added to the AnIML core. There are some points in favor of adding the concept, but it appears as though all requirements for multiple analyses can be accommodated within the current structure. They are looking at the concept of a page group, which might be useful for multichannel experiments, where, for example, multiple wavelengths are monitored in a chromatographic analysis. The page group would allow peaks to be linked back to where they came from. This new concept would also work with the existing technique definition model. The core group is trying not to introduce new concepts, preferring to keep the number of xml elements which the parser needs to understand to a minimum. Maren Fiege and Mark Bean met in Baltimore and worked on creating some example files to prove the whether this new concept is right or wrong. These will be evaluated at the next meeting of the core group. The following points were raised in discussion:

- The difference between this new page group element and the existing pageSet element was discussed. The pageSet is a container with no meaning. It is just used to bring multiple page elements. A page group is a way of pointing to pages that belong together.
- It is desirable to have a simpler structure, rather than “fifteen levels in the hierarchy” to define multiple analyses, or hyphenated techniques.
- A concern was raised about how these files would be generated – on the fly or post-measurement. It was noted that they would need to be generate post-measurement, since some of the data would not be available until the experiment had been completed.
- AnIML doesn't work well in real time. But how much burden will be put on the user to enter additional information? That is up to the vendors. In a validated environment, the software will make sure that the required information is collected.
- How can the core be validated, except through the generation of example files in particular techniques? But how can the technique definitions be created until the core is finalized? Vendors are already developing systems around the existing documents. Theoretically, the standard could be validated, but that may well result in a standard that is unusable. The process of development of this standard is not ideal. Rather, encoding real examples will turn up problems in either the core definition, or the technique definitions, or both, and refinements will need to be made. Both theoretical validation (as far as possible), as well as practical examples, must be included.

VIII. AnIML Example. Anh Dao presented an example of IR data, taking files from Maren. Tony noted that some of the elements added in the technique files were already in the core schema. For example, title should be in the core, not in the technique. Number of sample scans and length of time of collection, should also be in the core. Group members are encouraged to examine the sample files prepared by Anh Dao and think about what should be included in the core, and what should be in the technique definitions. Discussion will continue in the next meeting, but discussion on the SourceForge Developer List is also encouraged.

A discussion of legacy files concluded that there should be an examination of JCAMP files, and an attempt to convert these to AnIML format. Sample files can be found on the JCAMP site. Some things to consider are the handling of integer values, and recording data as either absorbance or transmittance values, although the latter is a recommendation that perhaps is best left to the domain experts.

IX. Miscellaneous.

- Anh Dao Nguyen will be leaving Gary Kramer's group in September, so would like to write up the work she has done on UV/Vis, IR, and chromatography, for publication in JALA and/or other publications. Multiple publications are desirable, since a single article covering all techniques would not attract the attention of any of the individual experts. Co-authors are welcome. This needs to be done before Anh Dao leaves the group.
- **UnitsML.** A presentation to the group is needed to discuss how this fits into AnIML. This will probably not fit into the next meeting, but sometime in the future.
- **LIMS Institute.** The LIMS webcast mentioned in the last meeting is moving forward. Several AnIML people will be involved: Maren Fiege, Robert Lancashire, Richard Larsen. (It was noted in the July 7, 2005 meeting that this webcast is being organized by John Joyce, and is not related to the proposed training class from the LIMS Institute.)
- **Upcoming Meetings/Papers.**
 - i. In January, 2006, around the 23-25, two conferences are being held in San Diego, a laboratory informatics conference, and an ELN meeting. There is a possibility for some joint training/presentation that could involve both of these conferences.
 - ii. The deadline for submission of sessions to PittCon is nearing. Is there interest in having a session on analytical data there?
 - iii. Tony Davies will be presenting a paper at the ACS meeting.
 - iv. John Joyce has an article in SCI.

X. Agenda Items for Next Virtual Meeting.

- Report from Core Group
- Review of formal AnIML requirements document
- Discussion of technique definition documents for IR and sample IR AnIML file

XI. Action Items:

- Web page redesign: The web team will prepare for deployment of the new design
- Mark Bean, Alex Mutin, Tony Davies – complete formal requirements document, circulate prior to next virtual meeting
- Tool developers – post tools and documentation on sourceforge – does anybody have anything to post?

XII. Future Meetings:

- Working Group Teleconferences
 - July 7, 2005, 10:30 am – 12:30 pm EDT, pending teleconference availability
 - July 22, 2005, 10:30 am – 12:30 pm EDT
- Business Meeting: EAS, November 14-17, 2005, Somerset, NJ
- Business Meeting: PittCon, March 12-17, 2006, Orlando, FL

XIII. Adjournment

Motion. Second. Carried unanimously at 12:30 pm EDT.

Submitted by David Martinsen, ACS, ASTM E13.15 Secretary