



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

April 13, 2007

Virtual Meeting

I. Introductions and Welcome: Gary Kramer, E13.15 Chair called the meeting to order at 10:35 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
Mark Bean, GSK
Maren Fiege, Waters
Steve Hoffman, BMS
Gary Kramer, NIST

Peter Linstrom, NIST
David Martinsen, ACS
Jamie McQuay, Scimatic
Dale O'Neill, Agilent
Karen Wilson, ASTM

III. Next Virtual Meetings: May 15, 2007, 10:30 - 12:30 EDT

- UV/Vis Technique Group: April 17, 2007, 10:30 am - 12:30 pm EDT.
- Chromatography, April 25, 2007, 10:30 am - 12:30 pm EDT.
- The next business meeting will not be held at EAS, since none of the usual attendees normally attends EAS. The meeting does not have to be face-to-face, but the IUPAC meeting in Torino, Italy in August is a possibility. Gary Kramer and Robert Lancashire have submitted papers. Dave will check with Tony and Robert to see if IUPAC has scheduled a business meeting.
- PittCon 2008: Gary submitted a proposal.

IV. Minutes:

- Minutes from the March 16, 2007 meeting are not yet completed.
- Reminder from Joe Koury: Minutes from the business meeting at PittCon 2007 are needed.

V. AnIML Wiki:

- Mark Bean made a mockup of our web page as a wiki. This is a prototype, not yet public. But it is a test to see if we could maintain our web site as a wiki, rather than with special software. The url is <http://www.animlwiki.org/animl/wiki/default.aspx?AspxAutoDetectCookieSupport=1>. The same software is used as the developer wiki. Mark is currently copying information over. There was some discussion about a web site wiki: a) Would everyone be able to edit? We could set whatever policy we wanted. Probably the committee members would be allowed to edit, all others to view; b) Mark will work with Dave and Patrick to decide if and how to incorporate the wiki; c) The benefit of the wiki is that anybody can edit, so we don't need to rely on a single person to keep things up to date. But if Mark was no longer able to provide the wiki site, someone would need to migrate it.

VI. Pittcon 2007:

- The Pittcon presentations will be posted on the website. Speakers should put them on the wiki, or send to Gary.

VII. Reports from subgroups

- **UV/Vis Task Group:**

- The group has been working on the UV/VIS technique document. They are ready to examine the Fourier Transform variables. These are present already, to some degree. These will be discussed at the next meeting on Tuesday, 4/17. Other than that, they are awaiting resolution of the issue of indexing techniques. This is not preventing progress right now, but will be very soon. Note that the indexing question is delaying development of the validator software.

- **Core Group:**

- The core group discussed indexing at their last meeting. Two possibilities exist - each technique could be required to define its own structure for indexing, or there could be a global indexing structure. The first approach has the advantage that both the data and the indexing are defined in the same place. On the other hand, a global technique avoids the repetition required to store the indexing structures in all of the technique definitions. The concept of XML modules was discussed, as a mechanism to write the indexing rules once, and then include the same module in each of the techniques. This makes the AnIML core simpler, but makes it harder to write the techniques.
- A concern was expressed that using the indexing approach, AnIML doesn't reflect the hierarchical relationship between parent and child

mass spectra in MSMS, for example. If the semantics of the experiment aren't captured, the implementation of the standard could be discouraged. It would be nice if you could create a structure which reflected the reality. On the other hand, there is a problem when there is no parent-child relationship. For example, consider LC/UV. There is no LC data, only the UV data collected on the effluent of the LC. An assumption is made that the UV data is strictly related to the separation step. However, there may be some kind of splitter in place, which is doing some kind of filtering. It was noted that there could be an infinite variety of special experiments of this nature, and that we need to target the most common techniques, and provide a structure to handle those 80%. Someone who runs the nonstandard experiments will need to develop extensions for the special techniques.

- It was pointed out that chromatography is a separation method, and produces no data. In order to collect data, a detector needs to be included, and perhaps multiple detectors. Some detectors are so common that they could, and perhaps should, be added to the chromatography technique. These are somewhat different than hyphenated techniques. Hyphenated techniques do have a parent-child relationship, but in chromatography, there is not. Rather the parent-child relationship is between the sample injection and the detector. Perhaps one way to view experiments is that there are analytical techniques, which are defined by a detector collecting data, and structuring techniques, which operate on the sample in some fashion but do not themselves generate data. Fundamentally, there is no difference. The purpose of the structuring technique is to provide indexing variables. Chromatography could be used to structure the experiment by creating a timeline, but also has a set of associated metadata. It was noted that if, in an LCMS experiment, you stored the index value with the child, rather than with the LC parent, which has no data, it would follow the way vendors currently store indexed mass spectral data. Rather than a giant matrix, it really is a sequence of experiment steps. The parent contains no data; rather it just sets up the indexing. In MS, m/z and time are stored together. It was noted that this is exactly what the core group has been arguing about, and is seeking to resolve.
- **Chromatography Working Group:**
 - The chromatography group is working on quantitation and peak finding. One question they have is whether quantitation and peak finding belong to chromatography, or to all techniques. It was noted that different techniques handle these things differently. For example, peak skimming and tailing are well supported in chromatography, but are rarely used in NMR. It is basically integration software, but the

way the software is applied is not the same. It was noted that Sciex uses the same peak finding software for chromatography and spectra. It was also noted that there are many ways to locate the top of the peak, and there is some controversy about the most appropriate mechanism. There are customary ways that chromatography does it, customary ways for NMR, etc. Just in case of UV/Vis and NIR, there are at least four ways to locate the top of the peak. There are advantages and disadvantages to each. What AnIML should do is capture the customary way people do it. Therefore, the quantitation method should belong to the technique. Dale O'Neill noted the issue is that this should be put in the methods area, so it doesn't need to be duplicated in every one of a series of injections. Therefore, it should not be in the experimentStep, but rather in the experimentStepSet. So far, they have defined the separation technique. They are now working on method quantitation, which would go into the template area. The third item is documentation. They are wondering what the best way is to send out documentation so it's easy to review and make comments. Mark Bean suggested that what doesn't work is the way the UV/Vis method is currently being handled - the documentation is always 2-3 months behind. They are trying to work on the XML and use the XML to create the documentation. Gary Kramer noted that this is backwards in UV/Vis. The group should first generate the text, and then generate the XML. The reason for using the XML is that they are working from documents created by Alex and Ronny. Burkhard created a program to generate text from the XML, but the tool isn't available to everyone. Mark Bean offered to will take documentation file and put into the AnIML wiki. Others could then view/edit, and everybody will always see the same version. He noted that Stuart Chalk has been talking about some mechanism to convert a wiki environment into technique definitions. Dale noted that he just uses XMLSpy to lay out the different pieces of information for discussion and display, rather than creating a full schema.

VIII. Miscellaneous Discussion:

- Peter Linstrom circulated a document entitled *Indexing Techniques and Representations of Hyphenated Techniques*, in which he presented an argument that references should be by value, not by index. The reference could be a floating point value, such as the time of the spectrum, in which case there is no data in the chromatogram. He feels that this is the way it should be. The references were made not into a keyref, but an image or index. Peter suggests that this should be converted to a value, and referred to page 2 of his document. If the reference has units, such as time, the units need to be specified. Maybe units could be moved into the ParentDataPointReference element, but we don't need a separate indexing table. If the real index is, for example, 5 seconds, we want that to be stored. If it is stored with the child,

then you don't need an index table, but it needs to refer to a parent. That is, a parent not in the sense of XML or SQL, but rather, for example, a point of reference to where a mass spectrum came from. This is in an analytical chemistry mode, rather than an XML mode.

- Gary Kramer noted that the problem is that you can add vectors, but you can't add indexing vectors. An extension can not define a new indexing variable. Peter countered that if you are thinking of adding an indexing variable, you are probably doing something wrong.
- Mark asked about an instance where data for different types of data were acquired at independent data rates. Peter said each could be treated as a child. But if you want the data to be organized by temperature rather than time, you would need to make a new experiment step.
- Gary's document shows simple examples with GCMS, then a more complex example using a Faraday cup detector. All of the experiments are put together as either as parent-child, or as parallel steps. Examples include LCMS and UV/Vis with diode array detector. Time values go into the time vector, with either the mass spec, or the diode array. There is no guarantee in the parallel experiment, that the two different techniques collect data at the same times. But since the indexing value is with the data, both are handled independently.
- Gary noted that there doesn't need to be one right way to do this. With indexing techniques, there may be optimal ways to do it. But how you put it together, and what you decide are primary indexing variables, is up to the way the developers see the way the techniques fit together. The user may not agree with the way the vendor sees the experiment. But the indexing variables will tell the software how the data are structured.
- Mark wondered what the parent would be. Peter replied that the parent may contain data, but it may not. Consider an FID chromatogram. The FID is collecting data, giving intensity vs. time. If there is a splitter connected to a diode array, two sets of data result, spectra from the diode array, and an FID curve. Note: Indexing vector doesn't always need to be time. It could be temperature. Take spectrum, or data point measurement, at specific temperature points.

IX. Adjourn: 12:30 pm EDT.

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