



**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](mailto:service@astm.org)  
**Web** [www.astm.org](http://www.astm.org)

---

**Committee E13 on MOLECULAR SPECTROSCOPY AND CHROMATOGRAPHY**

**Minutes for E13.15 Subcommittee Working Group**

10:30 am – 12:30 pm EST

December 8, 2006

Virtual Meeting

**I. Introductions and Welcome:** Gary Kramer, E13.15 Chair called the meeting to order at 10:40 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:**

Stuart Chalk, U. of North Florida  
Tony Davies, GSK  
Maren Fiege, Waters  
Patrick Gleichmann, NIST  
Rolf Grigat, Waters  
Joe Koury, ASTM

Gary Kramer, NIST  
Dale O'Neill, Agilent  
Kordian Placzek, NIST  
Alexander Roth, Fachhochschule  
Wiesbaden  
Burkhard Schäfer, BSSN

**III. Next Meeting:** January 19, 2006, 10:30 - 12:30 EST.

Note: Next core meeting is Tuesday 12/12/2007; next UV/Vis on the 12/14/2006.

**IV. Minutes:** No minutes available.

**V. Other Meetings:**

- a. TOPCOMBI meeting:** Gary Kramer and Burkhard Schäfer have submitted abstracts.
- b. Lab Automation:** We do not have a podium spot, but Burkhard Schäfer will try to submit a poster. The AnIML committee members attending the meeting may try to get together for a meeting during or after the meeting.

**VI. Core Developers:**

The core group has succeeded in converting Mark Bean's data files to AnIML. It was complicated, but successful. Mark has since written a converter to convert his files into AnIML in a single step. The core group feels it is on the right track.

The core group expressed a concern about the best way to handle detectors for which no technique definition exists. There are many different types of detectors that could be used with chromatography. Does each detector type need its own technique definition? The AnIML committee has discussed techniques for UV and MS, for instance, but not ELSD. If someone uses ELSD as a detector, how can he/she store their data and metadata? The committee spent much time in discussion on this issue. Several options were considered. A primary consideration is to avoid some of the difficulties that have emerged in the JCAMP world, where an organization or individual could create arbitrary fields which would have no meaning to anyone else. One suggestion was to create a generic technique which could be used for those detectors which did not already have a standard technique definition, with rules about what could be stored in the generic fields. This generic technique could be used for a specific detector, in the interim period until a standard technique definition was created. There was concern that people might use the generic technique, even when a standard technique was available. It was suggested that a generic technique could be modified by anyone to create a new technique for a new detection method, but that the new technique be stored in a standard location. Others using that detector could also make use of the new technique. Eventually, the discussion led to a consensus reflected in the following motion:

**Motion:**

Add a Vector "Intensity" to the "Chromatogram" Page of the Chromatography Technique Definition(s). The purpose of the vector is to take intensity data from arbitrary detectors. The data type is "floating point". The vector is optional, dependent, and can have an unlimited number of occurrences.

The vector is only used if the chromatogram is materialized (and not derived from individual spectra on the fly).

The chromatography technique(s) would also hold a number of generic parameters describing the instrument, plus a parameter for the original unit the detector used.

Any parameters beyond that would need to be defined in an extension.

If there is a technique extension available for the detector technique, this is to be used.

The motion was seconded, and passed unanimously.

- VII. UV/Vis:** The UV/Vis document is in progress. The group is approximately  $\frac{3}{4}$  of the way through the document. The draft is on the CVS are on SourceForge.

- VIII. CVS vs. SVN:** A proposal was made to change from CVS to SVN, an alternate versioning mechanism now recommended on SourceForge. It includes a tracking capability, and could be useful if a wiki model were adopted for the group to create documentation on elements in the schema. It was noted that SVN has poor tool support, but can track files as they are moved around in the system, something which CVS cannot do. It is more advanced, but the tool support is limited. For more info <http://trac.edgewall.org/>. It was noted that we haven't yet reached the limits on CVS, and it would take some effort to switch to the new system. Perhaps we should continue using CVS until we see a more significant reason to switch.
- IX. AnIML File Size:** Dale O'Neill raised an issue about the potential size of AnIML files, especially for mass spectral data. These will be very data intensive. Even if a viewer wants to display just a small portion of the data, it will need to read and process the entire file, with an impact on both time and memory. Dale discussed a new container strategy, which will be part of the new Microsoft Vista OS, as well as Microsoft's XPS file specification. Using containers, the basic AnIML core would contain metadata and references to the vectors, with the vector data being stored in the containers. Applications would read the core, but would only read the data itself when needed. This would be much more efficient since the complete file would rarely need to be read into memory in its entirety. In the discussion that followed, the point was made that we need to be compatible with more operating systems than Microsoft. In addition, one of the primary goals of AnIML is to support the interchange of data. An application could first import the data into a local environment, which was optimized for reading and viewing. A viewer would not necessarily need to read the AnIML file itself. Burkhard pointed out that XML already has a capability for handling references, using XML entities. Using XML entities would require no changes at all in the AnIML schema. Normally, a parser reading an XML file would automatically resolve all entities, resulting in reading the entire data set and creating the large virtual file. However, the parser could be set to not resolve all of the entities, but just read in the entities themselves. The vector data, referenced by the entities, could then be read in as needed. It is an appealing strategy to increase the efficiency in handling large datasets. It was noted that there might be some advantage in documenting how the XML entities could be used, to let people know that this capability exists, and to guide them in their usage. Exploring some of the newer technologies may also be useful. The discussion on this area was cut short because of time.

**X. Adjourn:** 12:45 pm EST

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