

---

**Committee E13 on MOLECULAR SPECTROSCOPY AND CHROMATOGRAPHY**

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

10:35 am – 12:30 pm EST

Friday, May 13, 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. Minutes from the April 29 meeting were approved as distributed.
- III. Attendees:**
- |                                      |                           |
|--------------------------------------|---------------------------|
| Michael Boruta, ACD                  | Peter Linstrom, NIST      |
| Mark Bean, GSK                       | Dave Martinsen, ACS       |
| Stuart Chalk, Univ. of North Florida | Charlie Manfredi, Agilent |
| Jamie Duckworth, ThermoElectron      | Mark Mullins, SSI         |
| Maren Fiege, Waters                  | Anh Dao Nguyen, NIST      |
| Gary Kramer, NIST                    | Ed Svastits, CH2M         |
| Corice Leonard, ASTM                 |                           |
- IV. AnIML Website Redesign.** The website has been unavailable. Mark Bean will move the site to a more stable area. Please review the site and send email to Mark Bean and Dave Martinsen. Stuart Chalk offered to have someone work on the redesign.
- V. The extension for AnIML Technique Instance Documents.** As an assignment from the previous meeting, Burkhard, who was not present today, sent a power point presentation to discuss the extensions for AnIML files. The issue is whether or not .atid is a good extension for the AnIML Technique Instance Document. The slides discussed the meaning of “instance” in general computer science terms, and within the XML community. It could be argued that a technique definition is an “instance” an XML sense, but that there will be some confusion in computer science terms. A suggestion was made to use a more neutral extension to avoid confusion. Several recommendations were made, including .ani, .aml, .animl, .atid, .atd, and .animltd. After some discussion, Maren Fiege proposed the extension .atdd, for AnIML Technique Definition Document. There was a consensus on this extension.
- Action: .atdd will be the extension for AnIML Technique Definition Documents.**
- VI. URLs for Schemas.** Mark Bean agreed to propose permanent URLs for AnIML schema, at least during the development period, so that these can be referenced in the various XML files under development. In this way, XML files will not break when they are moved from one environment (a person’s personal computer) to the AnIML development site, and back to another person’s personal computer. Mark was not able to complete this task, but will do this by the June 17 meeting.
- VII. Formal AnIML Requirements Document.** There was no activity to report in this area. It was noted that the project plan needs to be updated. Without formal requirements, it is difficult to move forward significantly.

**VIII. Core Developer's Meeting.** Mark Bean reported that progress has been made in the core developer's meetings, with a diversity of experiences and perspectives among the different participants. The group discussed the handling of hyphenated techniques, including LCMS and GCMS. In those techniques, it is common to sum spectra across the chromatographic peak or a time window, collecting perhaps one scan per second for 400 scans, and performing the sum as an outside process. Background spectra may also be measured. The question was raised as to what should be stored in the AnIML file: the raw spectra, the sum of the spectra, with or without background subtracted, etc. Furthermore, should the entire collection of spectra be stored in a single AnIML file, or should each spectrum be stored in a separate file.

In the subsequent discussion, a number of points were raised:

- In some cases, it might be desirable to retain the raw data, so different transforms could be applied later on
- Both raw data as well as processed data could be stored. In some cases, the instrument processes the data on the fly, and only records the processed data. In that case, only the processed data would be available for export to AnIML
- Both separate files for each spectrum, and multiple spectra in a single file could be accommodated within the scope of the AnIML definition
- In some cases, annotations are used to mark certain data points. Should these be accommodated?

While many different possibilities are allowed, we need to think about what usage to recommend in the documentation.

The Core Developer Group also noted that they have received nothing from the technique-specific task groups. They wondered if there were any reports. Stuart Chalk noted that the IR and UV/Vis group had not met. Gary noted that the Spectroscopy group, headed by Ken Busch, had definitions of many acronyms used in MS, which could be used in the technique definition file.

**IX. AnIML Example.** In keeping with the pattern for reviewing AnIML techniques and sample files, Anh Dao, working at NIST, has been using Jamie Duckworth's files, and Maren Fiege's technique definitions to create some practical examples. In this way, missing items, or items in the wrong place in the hierarchy, can be uncovered and reported back to the Core Developers and the technique task groups. For the June 17 meeting, Anh Dao will focus on the IR technique. For this meeting, she will review the synthetic UV/Vis spectrum of holmium oxide in perchloric acid.

It was noted that the xml files distributed by Anh Dao would not work in a generic environment because of references to files in specific directories on Anh Dao's machine. This highlights the need for a standard web-accessible location for schema files for use during the development cycle. This was mentioned earlier in the minutes, and will be addressed by the June 17 meeting.

A number of discussion points were raised during Anh Dao's presentation of her example AnIML file. The discussion is captured in the following question/answer (Q/A) format:

Q. A sampleID and a sampleName are both required parameters. Are both needed?

A. The consensus was that both are required. The sampleID is used for machines, and could, for example, store a bar code. The sampleName is intended as a more human-readable identifier. As a general philosophy, it is desirable to have enough parameters so people can operate within their normal environment. They can choose to use only those parameters they need to use.

Q. How should sample concentration be specified?

A. Anh Dao used a string to define how the sample was prepared. The discussion noted that a wide variety of parameters are user-interpreted, some are common usage, but even common usage may be vendor dependent. We should use ASTM definitions where they exist, and should reference the appropriate ASTM documentation when we do so. It was noted that Anh Dao had added ASTM references to her documentation.

Q. How should the sample in this example be defined: holmium oxide, or holmium oxide solution?

A. The sample is holmium oxide, and the solvent could be defined as another sample with the role of solvent. However, Anh Dao noted that in AnIML, only one sample description is allowed, so both holmium oxide and the solvent cannot be defined. Maren Fiege noted that one category named substanceDescription is allowed; it has been suggested that this be extended with a number. Jamie Duckworth noted that the substanceDescription could go from 0 to infinity. The decision was made to refer this question back to the Core Developer Group for more discussion and resolution. It was noted that there needs to be some place to define a substance as either a solvent, or sample, or background. This is intended to be accommodated in the sample role.

Q. For melting point, both min and max are defined. What if we don't know whether my measurement is the min or max?

A. Min and max were defined to allow the melting point or boiling point range to be reported in the file. If only one value is known, the min value should be used for melting point. For boiling point, is the max value to be used if only one value is known? We need to define the convention to be used.

Q. With respect to chemical substances, is CML allowed?

A. CML, MOLFILE, SMILES are now allowed, InChI is to be added.

## X. Miscellaneous.

- **UnitsML.** Gary Kramer noted that progress is being made on UnitsML at NIST. The goal is to provide both a human- and computer-readable database, and the ability to look up specific units based on ASCII identifiers. Different systems using UnitsML could share information using these ASCII strings. Some "units" are really unitless. These could be handled as a ratio of units, or has having a unit of 1. SI units are scheduled to be completed by October 1, and NIST is trying to get people from other ML's to comment. Derived units will be included. Physical constants are not supported in UnitsML; these are handled on another NIST site.
- **LIMS Institute.** Instead of running meetings, the LIMS Institute is becoming more of an educational organization. They plan webcasts of about 40 minutes, and wondered if ASTM E13.15 would like to make a webcast on AnIML. The consensus was that this is a good idea, but may be somewhat premature.

## XI. Agenda Items for Next Virtual Meeting.

- Website comments and redesign
- Review of formal AnIML requirements document

- Technique definition documents for IR

## **XII. Action Items:**

- Web page redesign
  - All should review, send comments to Dave.
  - Stuart Chalk and/or graduate student will look at website add comments for redesign.
- Mark Bean, Alex Mutin – complete formal requirements document, circulate prior to next virtual meeting
- Tool developers – post tools and documentation on sourceforge – does anybody have anything to post?

## **XIII. Future Meetings:**

- Working Group Teleconferences
  - June 17, 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

## **XIV. Adjournment**

Motion. Second. Carried unanimously at 12:33 pm EDT, at which time the virtual meeting evaporated.

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