

---

**Committee E13 on MOLECULAR SPECTROSCOPY AND CHROMATOGRAPHY**

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

10:40 am – 12:30 pm EDT

October 28, 2005

Virtual Meeting

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

**II. Review of Agenda and Minutes.** The agenda was approved without change.

Minutes for the October 14, 2005 meeting were not available.

**III. Attendees:**

Mohan Cashyap, GSK  
Tony Davies, Waters  
Maren Fiege, Waters  
Ronny Jopp, NIST  
Joe Koury, ASTM  
Gary Kramer, NIST

Peter Linstrom, NIST  
Dave Martinsen, ACS  
Mark Mullins, SSI  
Alex Mutin, Shimadzu  
Alexander Roth, NIST  
Karen Wilson, ASTM

**IV. Core developer Report**

1. Sample assignment and inheritance was clarified, and there is agreement on how those should work. As an example, in LCMS, is a new “sample” required for every spectrum, with the need to define a virtual sample for each MS? The group decided that this was not necessary. Rather, the appropriate sample properties for each step are inherited from the sample for the experiment. Burkhard Schaefer is preparing examples and use cases.
2. In this way, only the starting sample needs to be declared. Subtracting or summing spectra can be done with the existing reference mechanism.
3. New parameters are needed in referencing to accommodate extracted spectra, to describe what was done, e.g. baseline subtraction. The same mechanism could be done for derivative spectra.
4. Can this mechanism handle chromatography, where you only want to store the peak table? This could be done in two ways: the peak table could be considered a child of the chromatogram; or the peak table could be defined as a single experiment. The mechanism to use should be decided in the technique definition. This problem exists in JCAMP-DX, where both are allowed.
5. In MSMS, one might want to store all of the metadata of the 2<sup>nd</sup> step. In that case, you would need to consider a second experiment step.
6. These issues require further discussion by the core group.
7. Schema bugs within XMLSpy have been fixed on Sourceforge. Bugs in VisualStudio still need to be addressed. These are being addressed offline.

**V. AnIML Requirements Specification Review**

1. Section 3.1: Is there a requirement to store derived data, or simulated data? Is there any difference between that and measured data? There may be a need for additional parameters for simulated data. In that case, the additional metadata might need to be captured in an extension. At the end of paragraph 1, a note will be added to include derived data sets, and simulated data.
2. Is the term “flexible” appropriate? It might be kind of hard to verify.
3. Tony Davies will reorganize the format, to give the REQ first, followed by the explanation paragraphs. This document contains the project requirements as well as the specification. A new document, called the Scope, has been added to the list of documentation.
4. It was noted that the specification and an implementation guide are usually separate documents. There was some discussion about including a document which tells how to test against it. This second document, describing what the software needs to meet, will come after this specification documents. Some were uncomfortable with how such a test set could be described, especially in cases where there might be 2 ways to handle derived data. But eventually, when a new AnIML technique is defined, there will be a need to determine if it meets the requirements. In general, though, this will be done through round robin testing. All agree.
5. Section 3.2: good
6. Section 3.3: good
7. Section 3.4: Basic business requirement based on the guidelines for carcinogens. Longest requirement is 95 years for copyright, 30 years for chemicals in general, 60 years for a known carcinogen. In addition, there is a requirement that data must be preserved for as long as a compound is still sold.
8. Section 3.5: “must” will be modified to appropriate usage.
9. Section 3.6: substitute something for “conformant”, which is not a word. Once a document is certified as conformant, will all the necessary files be there? All of the history, test suites, should be maintained. There should be a list, or a site, which contains the “approved” schema files. We may need certificates to ensure that these files are the approved versions. In order to take hold and grow, we need this location to be available long term. This means that there is a requirement for location and access of files. These files should be available as a standalone package, even if one is not online. However, there is much that is outside of our control.
10. In the context of reprocessing data, can we store, the rework the data? Is this a separate requirement? Can be handled in 3.14. Whether or not the data can be reprocessed is an application issue, not part of the format. But in order to reprocess, enough metadata must be stored to allow the reprocessing to take place.
11. Section 3.8 – leave it in. It doesn’t hurt
12. Section 3.9 – good
13. Section 3.10 – The requirement is good, but the reasoning behind it is not. However, in a 96- or 384-well plate, each well might be a different customer, with different permission. But independent of well plates, automation. The first sentence will be removed.
14. Section 3.11 – “The AnIML project cannot”
15. Section 3.12 – take chemometrics out. It’s not a technique, but may encompass multiple techniques. The text will be changed to include “other techniques as those 3”.
16. Section 3.13 – change “software independence” to “PC platform independence, and make this a high priority requirement.
17. Section 3.14 – ok
18. Section 3.15 – this is telling software developers what to do. We should be telling them how the AnIML files are written. But the thrust of this section is not about how the files are written.

Rather it is to indicate that if only 2/3 of the AnIML file can be understood, then the other 1/3 should be ignored. Tony will recast this section as follows: “An AnIML file containing multiple techniques should be able to be read by software which understands only some of the techniques.”

19. Section 4.1: ok
20. Section 4.2: ok
21. Section 4.3: ok
22. Section 4.4: ok
23. The changes discussed in today’s meeting will be made and the resulting document circulated prior to the EAS meeting in November.
24. We need to find out from ASTM what to call the document. Could be a specification. Since it has some references, it could be guidelines. The more we can push these documents out, the better. Once available for ballot, people will come out of the woodwork.
25. The 2<sup>nd</sup> document, the project scope, is nearly ready and should be available for EAS: Draft Project Scope for AnIML 1.0.

**VI. Glossary and Definitions:** Another document we need is a glossary of AnIML specific terms, and terms and definitions we are using which are already defined by ASTM. We shouldn’t redefine something which is already defined somewhere else. Gary Kramer proposed a starting list for terms we need to define. Suggestions for additional terms will be welcomed. These terms should appear in a glossary, and perhaps could be explained in footnotes the first time they appear in the text.

1. AnIML Core
2. AnIML Technique Definition Document
3. AnIML Data Document
4. Vector Element – is this unique to AnIML?
5. Name Value Pair – this may be common enough that just need to identify the source of the definition we are using
6. Page
7. Experiment Step
8. Page Blueprint
9. AnIML Technique Extension Document
10. Module
11. Sample
12. Instance Document
13. Anything else with AnIML in front of it (should we put “AnIML” in front off all our terms, to give the context. E.g., AnIML Vector Element, AnIML Page, etc.)
14. **ACTION ITEM: Review this list, and come up with additions.**

**VII. Agenda for EAS Meeting:**

1. Review edited version of the AnIML Requirements Specification
2. Initial review of the AniML Project Scope Document
3. Discuss how to proceed to publish the standard
4. Work on Definitions
5. Developer group is nearing stage for round robins. Once the schema problems are fixed, this effort should move forward. Because of additional changes, we may need to start over and recheck UV/Vis, and then restart IR, since too many of the existing names were ignored.

- VIII. Naming and Design Recommendation:** There is a need for recommendations on conventions for naming the various objects within the AnIML environment, to avoid problems from the ANDI and NetCDF world, where there are many different ways to write the word “title”. Alex Ropp and JK are examining some naming and design recommendations (NDRs), and will prepare a specific recommendation to use in AnIML. Several NDRs are available. We don’t need to be consistent with them (we can’t, since they are not consistent with each other). But we need to have one for AnIML to increase interoperability when the standard is deployed, and other groups begin to write their own extension files. Some of these recommendations could be specific enough to check algorithmically. Some are more policy-related. The recommendations should be reviewed in a future business meeting.
- IX. unitsML:** Gary Kramer reported on ThermoML with respect to adoption of unitsML. At the current time, ThermoML is not using unitsML, but if unitsML could be shown to give an advantage, they would adopt it. ThermoML is currently in a 3-month review period, and cannot change anything at this time.
- X. Next meeting:**           November 13, 2005 3:00-5:00 pm EDT, EAS, Somerset, NJ  
                                  November 14, 2005 5:00-7:00 pm EDT, EAS, Somerset, NJ
- XI. Adjournment.** Motion to adjourn. Second. Carried unanimously at 12:35 pm EDT.

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