

Oct. 08, 2018

Data Exchange in the Laboratory of the Future

A Glimpse at AnIML and SiLA

Open data formats such as AnIML and communication protocols such as SiLA help to connect up a digital laboratory. They provide planning, execution, and documentation of experiments and processes. The result is a unified and complete data package that is ideally suited as a basis for data analysis and other uses.

When talking about the laboratory of the future, one often thinks of data. One is looking for ways to make data easier and better to use. Modern data analysis, machine learning and artificial intelligence methods promise a lot, but require a well maintained and easily accessible data foundation, as well as an optimal data flow between all involved systems.

But how does that work? Standardized data formats and communication protocols are important building blocks for such a digital infrastructure. Using them cleverly enables the seamless integration of devices and software systems. This creates more efficient processes, seamless data flow and improved data integrity.

This article presents two initiatives: the AnIML data format for cross-vendor analytical and biological data representation and the SiLA initiative, a communication protocol for the integrated laboratory. It also looks at how standards can be used to build a sustainable infrastructure for using and analyzing data.

The AnIML Data Format

AnIML stands for Analytical Information Markup Language. It is an XML-based file format for data from analytical chemistry and biology. It facilitates the storage of data from different measurement techniques and manufacturers in the same format. Not only results, but also raw data, information on samples, devices, methods, workflows, and audit trails can be found here. AnIML makes it possible to describe and document simple and complex laboratory processes. During development, special emphasis was placed on ease of implementation and low total cost of ownership.

Description of Experiments

Laboratory procedures are described in so-called “Experiment Steps”. An Experiment Step is structured differently depending on the measuring technique. It can consume and produce samples.



An Experiment Step can also record the results of another and continue processing it. So, an Experiment Step can represent a physical process or data analysis or manipulation. The devices involved in this Step, their settings, the software used and the executing user are also stored. Raw and result data are also included. To save space, redundant data from several Experiment Steps can be summarized by templates.

A simple example would be a UV spectrum. There is a single Experiment Step that contains the spectrum itself as well as the method and configuration data of the spectrometer. This Experiment Step references the measurement sample it consumes. If the same sample is measured again, it would be possible to make further Experiment Steps on this measurement sample, also with other measuring techniques.

It becomes somewhat more complex with chromatography, e.g. HPLC. Where there is only one Experiment Step for describing the injection. This is linked to additional

Experiment Steps to document the work and results of each connected detector. Thus, each chromatogram or spectrum represents another Experiment Step.

A peak table does not directly consume a physical sample. Instead, it consumes a chromatogram, thus representing an “Experiment on Data”. The method describes the settings of the peak finder.

AnIML is generic, so it is suitable for many different measuring techniques. The concept of “Technique Definitions” describes how an Experiment Step is constructed for a specific measuring technique. So the format is easily expandable. Existing software can even be applied to future measuring techniques. Table 1 shows some already implemented measuring techniques.

The Origins of AnIML

The AnIML format comes from the ASTM Subcommittee on Analytics Data (E13.15). Thanks to a broad group of representatives of end users, manufacturers, government institutions as well as research and teaching, a broad spectrum of requirements arose. The ASTM process ensured that everyone involved met on equal terms and had the same opportunities to shape the standard. This guaranteed that the requirements and interests of both manufacturers and users of different industries were equally respected. This also explains why AnIML is accepted by manufacturers.

AnIML in the Regulated Environment

To describe AnIML completely would go beyond the scope of this article. Nevertheless, some interesting aspects should be mentioned here. The format provides support for audit trails and digital signatures. These simplify use in a regulated environment, where regulatory frameworks place increased demands on data management. Audit trails can record and document any changes to the data. Digital signatures provide evidence that data originates from a particular author and that they have not changed since their original generation.

XML makes AnIML text-based. Thus, it can be read without any application software if necessary. The combination of readability of the XML format and demonstrable integrity with digital signatures makes AnIML an ideal format for long-term archiving, especially in the regulated environment.

Access as a Recipe for Success

Standards only prevail on the market if they are easy to use and offer favorable total cost of ownership. With XML, AnIML uses a proven and widely accepted base

technology and blends seamlessly into a large ecosystem. This results in some interesting advantages. Vendors are independent of proprietary libraries or implementations. They can use the appropriate framework for their application. End users can use many existing XML tools. Even tools originating outside the laboratory world can suddenly handle lab data, even if they were not explicitly developed for it. Think of report generators or data analysis tools. XML is easy to learn. So it is not surprising that some manufacturers were able to implement an AnIML support within a few days without prior knowledge.

Support from the Manufacturer

Another criterion by which a standard must be measured is its adoption. Here, the acceptance by manufacturers is crucial. Ultimately, a standard can only be used if the appropriate tools are available. AnIML is currently supported by over 25 manufacturers. For example, Agilent uses AnIML in its ECM system. Sciex relies on AnIML for its long-term archiving solution. LabWare, the largest LIMS vendor, provides its customers with AnIML for its device interfaces. BSSN Software offers AnIML tools as well as converters for more than 150 device models from different manufacturers. Due to the excellent availability of tools, AnIML is becoming increasingly popular among end users.

SiLA

SiLA focuses on the development of standards for communication with devices and laboratory software. So devices can be controlled, but also software can communicate with each other. SiLA relies on proven web service technology, providing a broad ecosystem of tools for implementation across multiple platforms. By using SiLA, both devices and software systems can expose their services to the network and be integrated. Such a service approach eliminates the need for the less robust exchange of job lists and results, using files in transfer directories. Instead, SiLA facilitates interactive communication, status queries and reaction to events. Since files are never stored for communication and later collected, it is much easier to ensure data integrity and traceability.

The Origins of SiLA

SiLA originally came from the field of laboratory automation and high throughput screening. Initially, devices such as balances, plate readers, pipettes and robotic arms were supported. SiLA then took the step into analytical chemistry and biology, starting with balances continuing through to complex systems such as Chromatography Data Systems (CDS). A standard for process management systems allows complex processes or individual samples to be delegated to subsystems. This

mechanism can be used for the standardized connection of CDS systems to LIMS or ELN.

The New Version SiLA 2

The SiLA consortium is currently working on SiLA 2, the next generation of the protocol. In the new version, the technical substructure has been modernized. Manufacturer-independent instruction sets are simplified by the feature concept. In addition, tool support has been improved to allow even easier implementation.

Adoption of SiLA

So far, SiLA has mainly proven itself in small and large installations in the pharmaceutical and academic environment in Europe. Numerous manufacturers are actively involved in the further development and use of SiLA in their products. The SiLA homepage provides an overview of the available drivers for about 200 device models.

Data and Communication Belong Together

SiLA and AnIML pursue different goals. As a communication protocol, SiLA enables the controlling of devices and software systems via web interfaces. AnIML, on the other hand, represents an XML-based data format for describing analytical and process data.

Both are important. If one uses a universal file format, one gets a complete documentation of an experimental workflow. Any exchange of data between systems, however, would have to be accomplished by moving files. This would not be reliable. However, if one had only one standardized communication protocol, although one could monitor devices and systems, the resulting data would be stored in proprietary silos.

The combination of SiLA and AnIML represents a powerful approach: SiLA controls the instrumentation, and transports resulting data in the AnIML format. This results in a complete data package after completion of the experiment, which includes all process steps as well as raw and result data. Both projects have been working together for several years to promote a unified and open ecosystem.

Data analysis

Laboratory tests are always triggered by an immediate need: one needs to examine a production sample for batch release. One takes a water sample to monitor the function of a treatment plant or operates a pH sensor to monitor a bioprocess. However, it becomes more interesting when we use this data to achieve further

goals through skillful analysis and interpretation. For example, one could discover relationships between process parameters and contaminants, perform trend analysis or generate statistics on the raw data from multiple clinical trials.

In order to allow such analyses, a large set of data must be available. This must be well structured and easily accessible. Only then can different analysis tools be applied to it. Also, the techniques for machine learning or the methods of artificial intelligence can be supplied with data.

Here an AnIML-based Data Lake can serve well. Relevant elements are extracted from the data collected by experiments and provided to an evaluation layer. Then suitable tools can draw from and operate on this layer. Since the rules for data extraction can be changed at any time, the available data can be supplemented if necessary. This results in a future-proof infrastructure that allows maximum value from the data.

Summary

Standards such as AnIML and SiLA clearly promote inter-operability. By using them, the data from different experiments can be uniformly collected, stored and evaluated. This allows the data of different devices and measuring techniques to be used with a uniform set of tools. You no longer need separate interfaces for each device type to perform a LIMS connection. Also, you no longer need separate software to use the data of a device. This reduces the overall costs of integration and data management. Also, the perspective of being able to analyze collected data more easily and use it across different organizations makes AnIML and SiLA worth more than a short look.

Author

Burkhard Schäfer

Contact

Burkhard Schäfer

BSSN Software GmbH

Darmstadt, Deutschland

Weitere Beiträge zum Thema

Contact

BSSN Software GmbH

Darmstadt

Deutschland