

Analytical Data Longevity and Management for Regulatory Compliance using AnIML Format and Cloud Technologies

John Gibbons¹; Viktor N Iassinskii¹; Burkhard Schaefer²

¹AB SCIEX, 71 Four Valley Drive, Concord, ON, L4K 4V8 Canada; ²BSSN Software GmbH, Darmstadt, Germany

ABSTRACT

The focus of our work is the creation of a 21 CFR part 11 compliant electronic records solution for the long term storage, archival and management of instrument data. The AnIML data format was able to accommodate all the required data including chromatograms, spectra, LC and MS methods, results as well as all associated audit trails.

INTRODUCTION

In order to comply with FDA regulation (21 CFR Part 11), electronic records produced by analytical instruments and data analysis applications are required to be stored for a very long and indefinite period of time. This poses significant challenges for the organizations as most of the instruments store data in proprietary formats. This requires to maintain a large number of potentially obsolete software and computer systems to be able to read that data, which is impractical. While instrument vendors strive to maintain backwards compatibility between versions the evolution of software platforms and technology continue to make it more challenging and costly. The question that arises from all this is, "what is your plan B?" A human readable, non-proprietary data format is one way to ensure your data will be available when an auditor asks to see it. But do they have the features and tools for you to build a compliant solution.

Non-proprietary data formats have been around for a long time from JCAMP-DX and ANDI-MS to mzData and mzXML. These formats were designed to be universal, however it became apparent that agreement on a single format would be elusive. The inherent advantage with the introduction of XML (eXtensible Markup Language) based formats is the human readability of them. This is necessary to remove dependencies and to improve interacting with the data while not taking on a significant burden to sustain the tools.

Once your data is safely archived your next problem comes to light. Over time, the data grows, project after project, year after year until terabytes and terabytes of data on rows and rows servers are needed to keep it all. That's when your IT department starts to complain about the cost associated with storing all this data "in house." The cost of sustaining all these servers over the long term becomes unbearable and lower cost alternatives are needed. That's where the "cloud" come in. While data security is still a high priority, cloud storage becomes a realistic alternative to massive onsite server rooms.

Regulations

From: OECD SERIES ON PRINCIPLES OF GOOD LABORATORY PRACTICE AND COMPLIANCE MONITORING

7.2 Records and Materials to be retained

- Records to be retained include paper records, photographs, microfilms or microfiches, computer media, dictated observations, **recorded data from automated instruments**, or any other storage medium containing the data generated in the conduct of a non-clinical health or environmental safety study.

7.3 Indexing

- The Principles of GLP require that records and materials retained in the archives be indexed so as to facilitate orderly storage and rapid retrieval.

7.6 Retention Period

- It is strongly recommended that records and other sustaining material associated with such safety studies be retained for as long as regulatory authorities might request GLP audits of the respective studies.

8.1 Decision to Retain Records Electronically

- The decision to retain records in electronic form has important implications. The long-term retention of electronic records may influence the choice of storage medium since deterioration of storage media can lead to permanent loss of records. Computer technology is developing rapidly and devices capable of reading storage media in common use today may not be available in the future. **Electronic records should be stored in a format that is readable for the duration of the applicable record retention period.**

Regulations cont'd

8.3 Defined Archive Area on a Computerized System

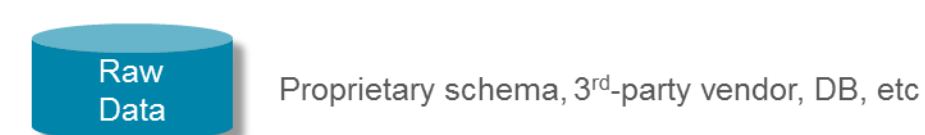
Electronic records may be moved from the production part of a computerized system to a **discrete, secure archive area** on the same computer system (physically separated, e.g. file record systems), or explicitly marked as archived (logically separated, e.g., database record systems). Records should be "locked" such that they can no longer be altered or deleted without detection. Records archived in this way must be under the control of a designated archivist and be subject to equivalent controls to those applied to other record types.

METHOD

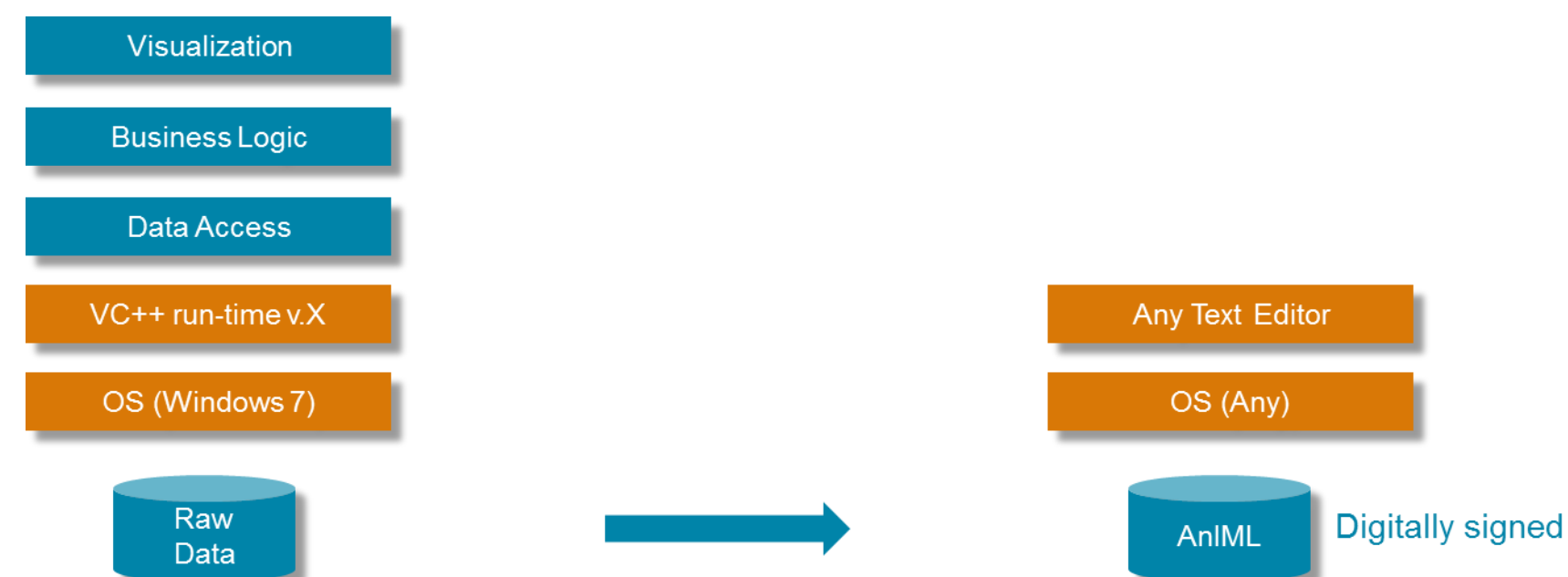
AnIML stands for Analytical Information Markup Language, it is an upcoming ASTM XML standard developed by a consortium of industry, academia, vendors and government bodies. AnIML has the ability to capture data from multiple analytical techniques capturing raw data, meta data (methods and audit trails) and processed data and results all secured with audit trail and digital signatures. Being XML based AnIML has the benefits of having a large choice of tools available to read the files, they are human readable and not tied to a specific implementation or library.

Current Technology

Visualization	Proprietary + 3rd-party modules (i.e graph control, grid, etc.)
Business Logic	Proprietary + 3d-party modules (math libs, domain specific algorithms, etc.)
Data Access	Proprietary + 3rd-party modules (data access libraries, ORM)
Compiler run-time libs (i.e VC++ v.X)	Compiler IO and UI libs (WPF, MFC, WinForms, etc.) or Open Source libs
OS (Windows 7, 32 or 64-bit)	OS Vendor or Open Source

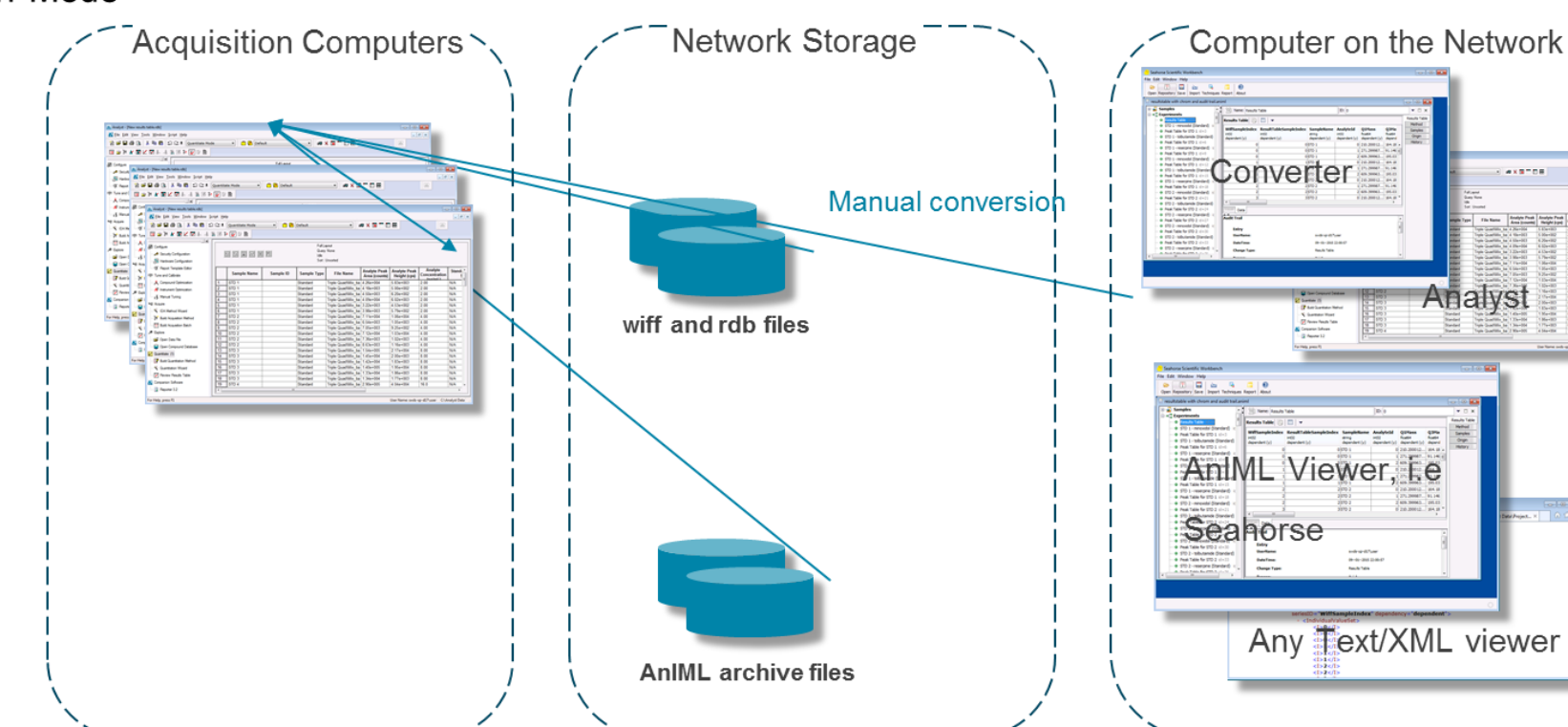


AnIML DATA LONGEVITY STRATEGY

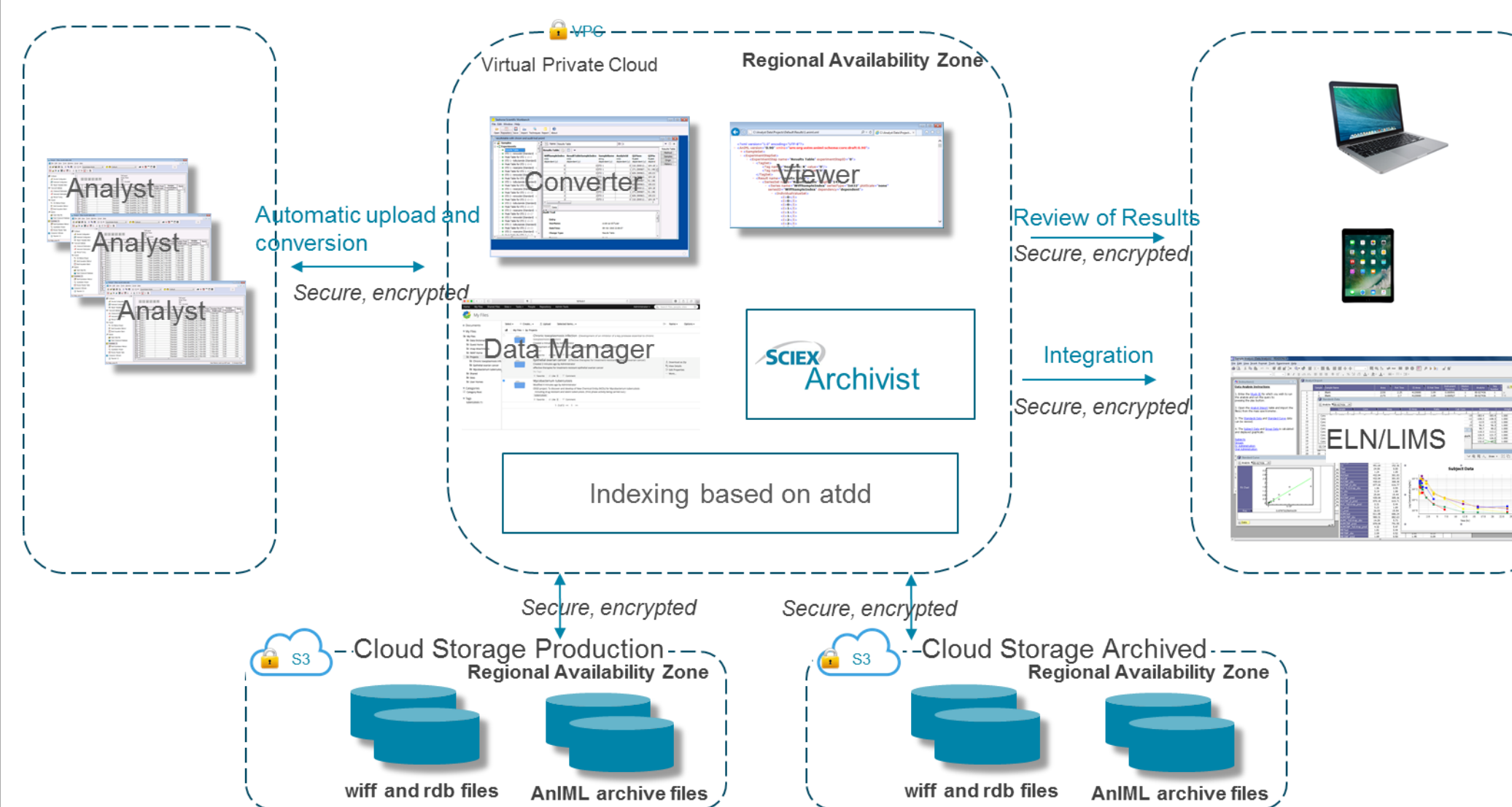


Phase 1: Local Deployment

- Converters
- Viewer
- Manual conversion
- Single Files or Batch Mode



Cloud Archiving Architecture



Cloud based System

AWS Example

- Data Durability.** Data is redundantly stored across multiple facilities and multiple devices in each facility with 99.99999999% of durability.
- Data accessibility.** S3 Standard is designed for 99.99% availability
- Security** S3 supports data transfer over SSL and automatic encryption of data once it is uploaded. Configuration of bucket policies to manage object permissions and control access to the data
- Scalability.** Virtually unlimited in terms of infrastructure.
- Centralized** storage with Global access.
- Indexing.** Availability of tools with powerful infrastructure behind.

CONCLUSIONS

An analytical data longevity and management solution for regulated environments can be achieved using the AnIML data format. The AnIML format provides a human readable solution free of constraints of operating systems, proprietary software and software versions. This solution is scalable and suitable for local or cloud based archival. AnIML's audit trail and digital signatures provide the required security measures to ensure data integrity.

TRADEMARKS/LICENSING

AB Sciex is doing business as SCIEX.

© 2017 AB Sciex. For Research Use Only. Not for use in diagnostic procedures. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Document number: [RUO-MKT-10-5957]