A New File Format Grants User Written Software Access to all Kind of Acquired Data.

Hartmut VALLEN, Gabriel CORNEANU, Thomas THENIKL, Thomas DUSCHL
Vallen Systeme GmbH, Icking, Germany

Abstract. During the development of our first AE system in the mid 1980's, we decided to store AE data in a binary file format. Back then, this was the best solution since data could be accessed randomly, not only sequentially, and stored to disk at highest speed and compactness. Several tasks could read the data simultaneously for online analysis.

A bottleneck is caused by the complexity of managing rather different kinds of data, such as hit driven and time driven features, waveforms, labels, and others. This bottleneck makes it difficult to add new features and functions, and to give other parties' software read access to all kind of AE data in real time.

Meanwhile the development in PC technology as well as in data management presents some nice alternatives in storing and organizing acquired AE data. The most flexible way in overcoming the limitations of the former file format and to address the upcoming challenges lies in using a data-base engine that organizes and manages measurement data.

Trials have proven that the performance of certain data-base engines satisfied our expectations and requirements with regards to speed, safety and efficiency. We decided for the modern SQL Structured Query Language (SQL) database engine SQLite and the format SQLite3.

Since the release (R2013.0809) SQLite is implemented in our software with excellent stability and maturity. Online analysis of over 100 thousands hits per second is ongoing supported. There are many tools available free of cost, that lets one look into the structure and data of a SQLite data-base file. Most programming languages can access SQLite data-base files.

Users can now read and process all kind of data on an AE data file, as it is written by the data acquisition software, without any need of conversion, even during data acquisition. The paper will give some examples.

1. Introduction

The data acquisition unit of an acoustic emission (AE) system receives and measures AE signals from any number of AE sensors, converts the measurement results into digital data sets and transfers them to the memory of the PC that controls the data acquisition. From the memory, data is stored to a permanent storage device, usually a hard disk, and also used by online analysis software, that performs a number of steps like data rejection based on filter criteria, AE source location, clustering statistical data processing and data visualization in graphical and numerical form.

Only data stored on file can be processed during post analysis. The data format, written during data acquisition, determines important performance parameters, e.g. speed of

data storage and analysis in terms of hits and waveforms per second, accuracy, resolution and dynamic range of measurements and time stamps, data volume in terms of bytes per hit and waveform, the versatility of using AE data in many different applications and the accessibility of AE data by third party software.

2. Previous binary data format

Using the previous binary data format and a modern PC one can store, analyze and visualize the results of more than hundred thousand hits per second in real time. The binary format creates very compact data files, compared to readable text format, and allows for fastest random access to any data set on the file, e.g. to start an analysis process at any desired data set.

A. Requirements on the AE data format

The data format must be able to store and read different kinds of data, e.g.

- Hit driven AE feature data
- Time driven AE status data
- External parameter data
- System driven labels
- User entered labels
- Hit or time driven AE waveforms
- Features derived from AE-waveforms after data acquisition

Furthermore, the data format must allow for simultaneous access of data by different analysis tasks during data acquisition.

All these requirements lead to a considerable complexity that makes it difficult to extend the functional range and to implement new features and new kind of data, for example, external data from a weather station.

B. Limitations of the previous binary AE data format

The main limitations of the previous data format were:

- Either "True Energy" (from squared AE signal) or "Signal Strength" (from rectified AE signal) must be selected prior to data acquisition
- Only 8 external parametric inputs were supported
- Only 26 characters user-text per Label (unlimited numbers of labels) were supported
- Dynamic range of time stamp was limited to about 280 days at 100 ns resolution (70 days at 25 ns resolution)
- Same waveform record time for short and long hits

C. Consequences

A new modern AE data format would significantly shorten the efforts for the implementation of new functions and features and could accelerate the progress in AE technology!

The implementation of a new data format was one of our essential objectives over the past years. It is already implemented in the current software release R2013.0809 issued in August 2013.

3. Online data processing and data file types

Fig. 1 shows the interrelation between real-time processes and involved types of data files in previous software versions.

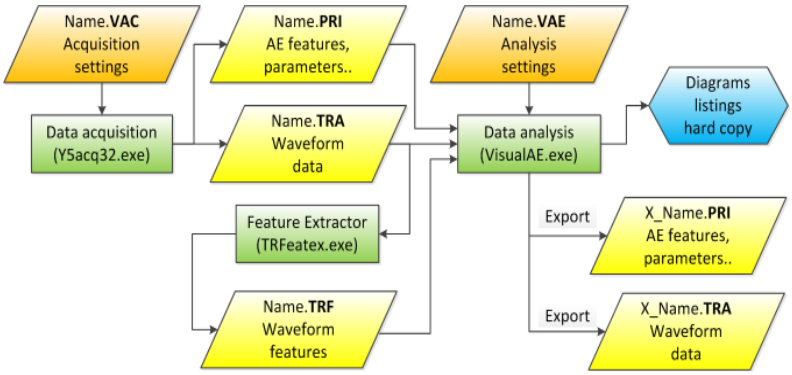

Fig. 1 Online data processing and data file types (previous versions)

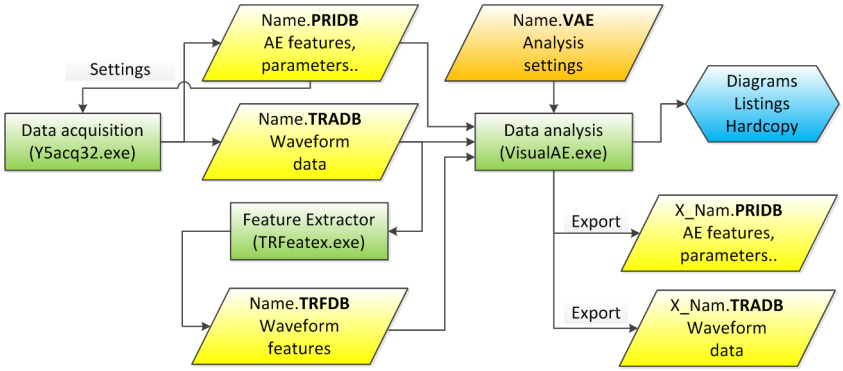

Fig. 2 Online data processing and data file types (new data-base version)

After installing the new software release all data processing output file extensions have "dB" appended to it. This is the main difference the user will encounter. Fig. 2 shows the changed file extensions in their context. Each file with extension PRIDB, TRADB and TRFDB contains a complete self-describing data-base structure which can be easily modified to match the requirements of new challenges. VisualAE, VisualTR and VisualClass are still able to read and process the previous file format as shown in Fig. 3, as well as the new format. For projects in which TRF files are exchanged among the partners, we recommend all partners either to change to the new file format at the same time or to postpone the installation of the new version until the project is completed.

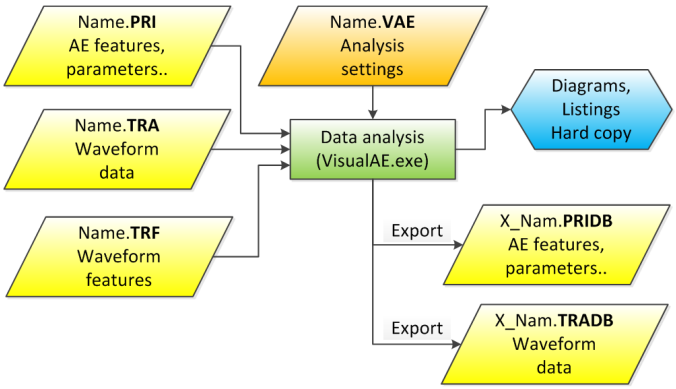

Fig. 3 In offline mode VisualAE can still process the previous binary format and export filtered data in new data-base format

For the convenient conversion of any number of data files acquired by a previous software version to the data-base format the conversion tool PRI2DB is provided.

4. Alternative data file formats

In the course of our analysis of the alternative data formats we selected two as most promising: SQLite and HDF5.

A. Library and data file format HDF5

HDF5 is a multipurpose data file format and a software library for the storage of all kinds of scientific data (see [2]). But we recognized two essential limitations with HDF5 in the tested version:

First the internal structure works with constant data set lengths which means that it would not become possible with HDF5 to store short compact waveforms of short hits and long waveforms of long hits.

Second, HDF 5 does not yet support the access of multiple simultaneously running analysis tasks to the file currently written by a data acquisition task.

B. Data-base engine SQLite and data file format SQLite3

According to [1] SQLite is the most widely deployed "Structured Query Language" (SQL) database engine in the world. The source code for SQLite is in the public domain.

SQLite offers following advantages over the previous format:

- The SQL language is very powerful, easy to learn and makes a self-describing data structure possible
- Access is easier, compared to HDF5
- There are a lot of free-of-cost tools available that let one examine the data structure and contents.
- SQLite is used in thousands of applications, means it is very stable and mature.

We found very few disadvantages of SQLite:

- SQLite requires more data processing power than the binary file, however, hundred thousand hits per second online data analysis performance is still possible when using a Dual Core CPU.
- SQLite data files are a little bigger than the previous binary files when saving the same content; partly due to database overhead, partly due to increased resolution of various fields (e.g. time). When deselecting unused data, files can be even smaller.

5. Benefits of SQLite in favor of AE technology

Data transparency: Data can be read and analyzed by proven and wide spread tools like Matlab without the need of a prior conversion. Free of cost tools make it possible to examine the structure and contents of internal tables. Fig. 4 shows a screen dump of "SQLite Expert" one of the mentioned free of cost tools. On the right of Fig. 4 you see a table showing the supported variables and their conversion rules to obtain values in physical units from the internally used integer format. All data measured are provided in tables of raw integer format as well as in tables of values converted in physical units.

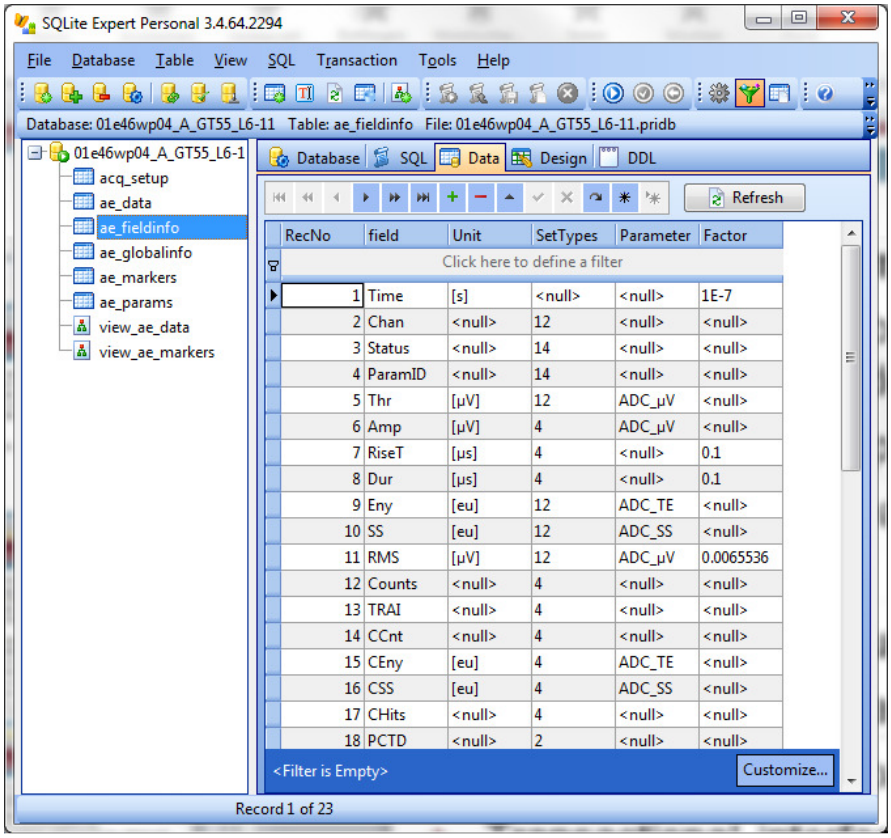

Fig. 4 Free of cost tool "SQLite Expert" shows structure (tables on the left) and content of one selected table (on the right)

Flexibility: Implementing new functions and features becomes easier, progress in AE technology becomes accelerated.

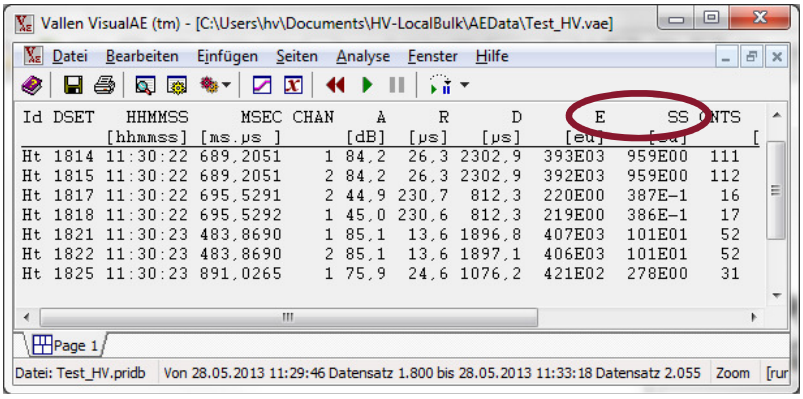

Fig. 5 True Energy (E) and Signal Strength (SS) can be shown simultaneously

Both, True Energy (E) and Signal Strength (SS) are acquired and can be shown simultaneously. Previously the operator had to select either E or SS prior to starting data acquisition (see Fig. 5)

Variable storage length of waveform data can be adapted to the length of a hit. Short hits, short waveform length, less memory occupation, long hits, long waveform length, more memory occupation. (Complete realization planned for 2014)

Self-describing names for Waveform features extracted by TR feature extractor. Previously, one had to assign one of the names FEx08 to FEx15 to one of 38 offered waveform features in the TR Feature Extractor program in order to get this waveform feature visualized in VisualAE. Now one can select a self-describing name, e.g. "T2_SR5" for "Spectral Ratio of time segment 2 and frequency segment 5" and gets this name offered by VisualAE for data visualizing or filtering.

Any number of external parameters can be supported. The current version supports 16 external parameters (PA0 to PA15, can be expanded to any desired number in near future).

New data integrity check: A manipulation of stored data during or after data acquisition will be discovered and indicated by VisualAE.

New way of handling time: The time stamps dynamic range is now unlimited. In the old file format the time stamp was limited to the range from 0 to 320 days. For the operator it is no longer necessary to distinguish between “absolute date/time of recording” and “recording time” for the start time reference. The operator can select either “absolute date/ time” or “relative time” in VisualAE.

Work with all frequency domain features: The new file format has no limitation for the number of records stored per dataset. All possible frequency domain features can be used simultaneously next ot the classical AE-features in an analysis. There is no more limitation to a maximum of 14 frequency domain features.

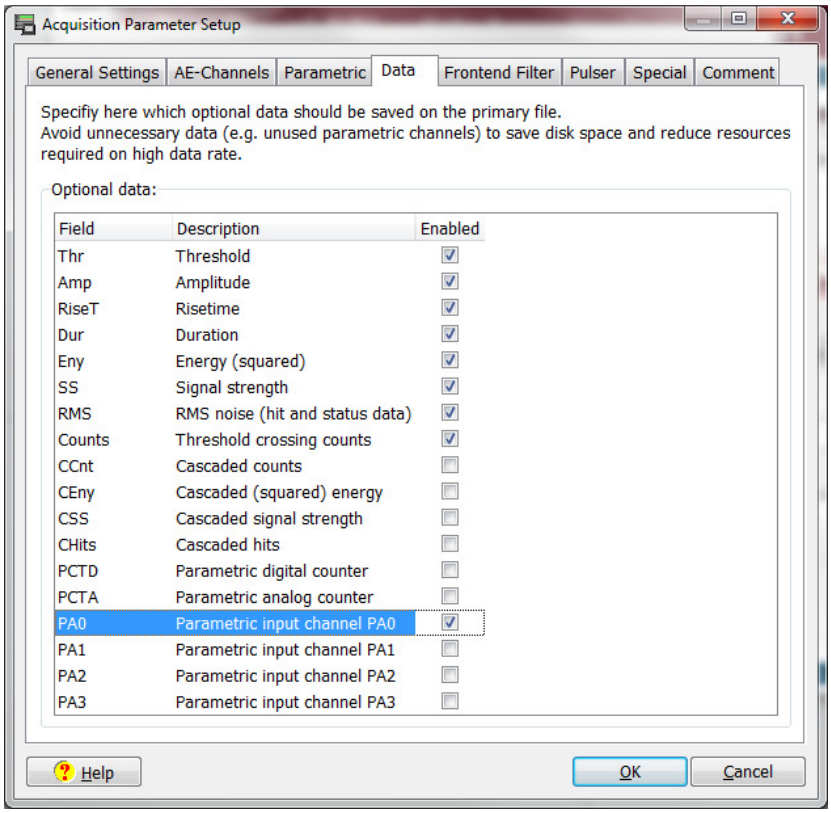

Fig. 6 Parameters not required can now be switched off from data acquisition in order to save space and processing power.

6. Compatibility with previous software releases

Profound changes in software are sometimes met with resentment. We wish to avoid this by making the transition easy and safe. As it can be seen in the sequence of Fig. 1 to 3, the user realizes the transition just by the modified file extensions PRIDB, TRADB and TRFDB. The "DB" in the extension gives the hint, that each file is a complete and transparent data-base.

VisualAE, VisualTR and VisualClass can read the previous as well as the new file format. There is no need to reinstall previous software just for being able to read data acquired in the past. But the client should be aware that data generated by the new software can't be read by previous software releases.

In Vallen Control Panel (Fig. 7) you find the new tool "PRI2DB". This tool converts any number of PRI and TRA files in one step to PRIDB and TRADB files.

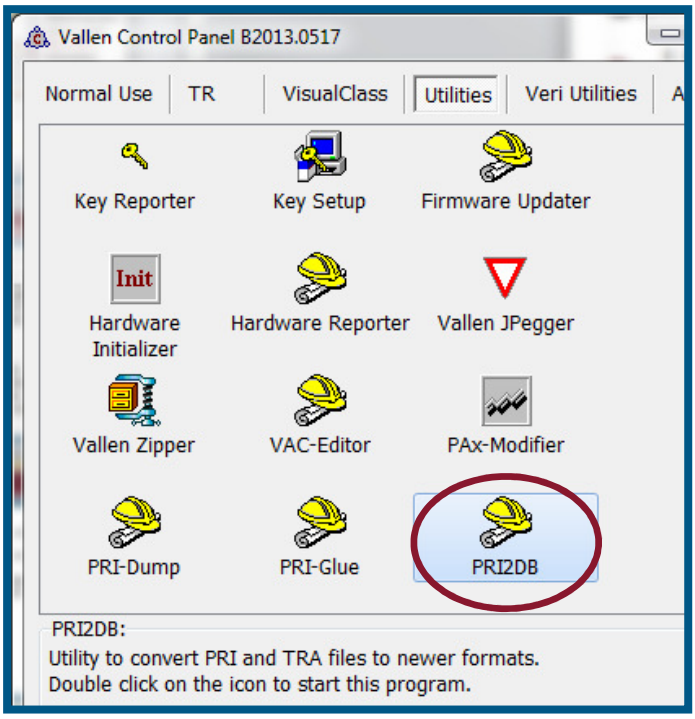

Fig. 7 New Tool "PRI2DB" added, other program names and usage have not been changed remarkably

7. Prospect

The Implementation of the new data format has been a serious concern for us over the last years and a challenging task. The new data file format gives us the chance to realize a number of new features addressed by our customers that could not be done in an efficient manner with the previous data file format.

The new format allows read-access of all kind of stored data by third party software, even during data acquisition. Customers can now, independently of us, develop their own software and perhaps develop new markets for the AE testing method.

Our sincere intention was that the new data file format will contribute to the success of our customers and to further progress of the AE testing technology.

References

[1] http://www.sqlite.org
[2] http://www.lrz.de/services/software/mathematik/hdf5/