

Der Nutzen der Speicherung von AE Daten in einem modernen Datenbankformat

Hartmut VALLEN *, Gabriel CORNEANU *, Thomas THENIKL *
* Vallen Systeme GmbH, 82057 Icking, www.vallen.de

Kurzfassung. Etwa 1985 entschieden wir uns, AE Daten in einem binären Format zu speichern. Das binäre Format bot die schnellste und kompakteste Form der Datenspeicherung und nutzte damit optimal die damals doch recht begrenzte Rechenleistung und Speichergröße. Dass dieses Format noch heute verwendet wird, spricht für seine Vorteile, z.B. die unübertroffene Echtzeitanalyse rate von über 100.000 Hits pro Sekunde.

Leider führt die Komplexität des binären Formats zu einer gewissen Unflexibilität, da so unterschiedliche Datenarten wie kompakte Hit- und Zeitgetriebene Merkmalsdaten, großvolumige Wellenformdaten, Anwendertexte (Labels) und andere, verwaltet werden müssen. Dies erschwert das Hinzufügen neuer Merkmale und Funktionen.

Nach jahrelanger Analyse moderner Alternativen entschieden wir uns für das Format SQLite3 und die Datenbankbibliothek SQLite. Nach [1] ist SQLite die meist angewendete "Structured Query Language" (SQL) Datenbank-Bibliothek der Welt. SQLite ist "transaktional", das bedeutet u.A., dass eine Änderung der Daten nur erfolgt, wenn eine Transaktion, einschließlich Schreibvorgang auf Datei, durch keine der folgenden Störungen unterbrochen wurde:

- Program Crash,
- Betriebssystem Crash, oder
- Stromausfall.

Somit ist SQLite mindestens so sicher wie das bisherige Format, wo die "Preallocation" bei o.g. Störungen die Datenrettung in den meisten Fällen ebenfalls ermöglicht.

SQLite ist nun in unserer Software (Acquisition32, VisualAE...) stabil und anwendungsreif implementiert. Eine Echtzeitanalyse rate von über 100.000 Hits pro Sekunde wird weiterhin unterstützt, erfordert aber eine Dual Core CPU.

In Zukunft können wir das Format so flexibel an neue Anforderungen anpassen wie nie zuvor. Endlich können wir vielen Wünschen nach Funktionserweiterungen näher treten.

Wir planen die Freigabe einer Softwareversion, die sowohl das neue als auch das bisherige Dateiformat unterstützt, für 2013. Anwender, die sofort testen wollen was auf sie zukommt, (Rückmeldungen erwünscht), erhalten auf Anforderung eine Beta-Version.

Es wird Sie überraschen, wie gut und fehlerfrei die neue Software - auch mit existierenden Daten - funktioniert und wie einfach und Anwender-freundlich der Umstieg erfolgt.

Wir hoffen, mit der Implementierung von SQLite zu einem wesentlichen Fortschritt der AE Technologie beizutragen.

1. Rolle und Bedeutung des AE Dateiformats

Die Datenerfassungseinheit eines AE-System misst die AE-Signale von den AE-Sensoren, konvertiert sie in digitale Datensätze und speichert diese auf einen Permanentenspeicher, meist eine Magnetplatte. Gleichzeitig liest die Analyseeinheit des Systems die Daten von der Platte (in Wirklichkeit meist aus dem Cache im Arbeitsspeicher), führt eine Reihe von Analyseschritten durch, zB logische Filterung, Ortung, Clustering, Statistik, und visualisiert die Ergebnisse in grafischer und/oder numerischer Form.

Die Datenspeicherung ist somit der zentrale Punkt, wo sich Datenerfassung und -analyse treffen. Das Datenformat entscheidet über die Geschwindigkeit, Genauigkeit, Ergiebigkeit, Volumen und die Anpassungsfähigkeit des Systems an die Anwendung

2. Bisheriges binäres Datenformat

Mit dem bisher verwendeten binären Dateiformat kann man mit einem modernen PC die Merkmalsdaten von weit über 100.000 Hits pro Sekunde erfassen und gleichzeitig analysieren und die Ergebnisse in Echtzeit anzeigen. Das binäre Format ist außerdem sehr kompakt und gestattet einfach den direkten Zugriff auf beliebige Datensätze innerhalb der Datei. Im Gegensatz dazu, muss eine Textdatei in der Regel vom Anfang ab gelesen werden. Ein schneller direkter Zugriff auf beliebige Datensätze würde das Führen einer zusätzlichen Indextabelle erfordern.

2.1 Anforderungen an das Dateiformat

Das Dateiformat muss unterschiedliche Datenarten speichern und bereitstellen können, z.B.

- Hit-gesteuerte AE-Merkmalsdaten,
- Zeit-gesteuerte AE-Merkmalsdaten,
- externe Parameter, Zeit- und Hit-gesteuert
- automatisch erzeugte Labels,
- vom Bediener eingegebene Labels
- AE-Wellenformen,
- aus Wellenformen abgeleitete Merkmalsdaten.

Das Dateiformat muss außerdem während der Erfassung den Zugriff auf die Daten von mehreren Analysetasks unterstützen. All dies führt zu einer erheblichen Komplexität, die es schwierig und aufwändig macht, den Funktionsumfang zu erweitern und neue Merkmale oder Datenarten, wie zB Wetterdaten, zu implementieren.

2.2 Begrenzungen des bisherigen binären Dateiformat

Die Begrenzungen des bisherigen Dateiformats waren:

- Entweder war "True Energy" (vom quadrierten AE-signal) oder "Signal Strength" verfügbar, und die Auswahl musste vor der Erfassung entschieden werden
- Es wurden immer 8 externe Parameter (PA0 to PA7) geschrieben
- Maximal 26 Zeichen Text pro Label eingebbar
- Zeitstempel ist begrenzt auf ca 280 Tage bei 100 ns, bzw. 70 Tage bei 25 ns Auflösung.
- Maximal 254 Merkmale können an VisualAE übergeben werden, was hinsichtlich der vom TR-Feature Extractor optional gewonnenen Merkmale ein Engpass darstellt.
- Gleiche Dauer der Wellenformspeicherung für kurze und lange Bursts

2.3 Konsequenzen

Ein neues, modernes Datenformat würde den Aufwand für die Implementierung neuer Funktionen und neuer Merkmale drastisch reduzieren.

Ein neues, flexibleres Datenformat kann den Fortschritt der AE Technologie beschleunigen!

Für Vallen Systeme GmbH war die Implementierung eines neuen Formats seit Jahren ein wesentliches Ziel. Noch vor dem Kolloquium wird eine Softwareversion erscheinen, die auf einem neuen, modernen Datenformat basiert.

3. Online Datenverarbeitungsprozesse und Dateitypen

Abb. 1 zeigt den Zusammenhang zwischen den in Echtzeit laufenden Prozessen und den beteiligten Dateitypen der bisherigen Softwareversionen (bis R20120509).

Es sind folgende **Prozesse** beteiligt:

- Y5acq32.exe Erfassungsprogramm für Merkmale und Wellenformen
- VisualAE.exe Datenanalyseprogramm
- TRFeatex.exe Wellenform-Merkmalgewinnung

Es sind folgende **Dateitypen** (Dateierweiterung) beteiligt:

- VAC Speichert die Erfassungseinstellungen
- PRI Speichert AE-Merkmale, externe Parameter, Labels u.a.
- TRA Speichert Wellenformdaten
- TRF Speichert Merkmale des TR-Feature-Extractors (zB Spectral Ratios)

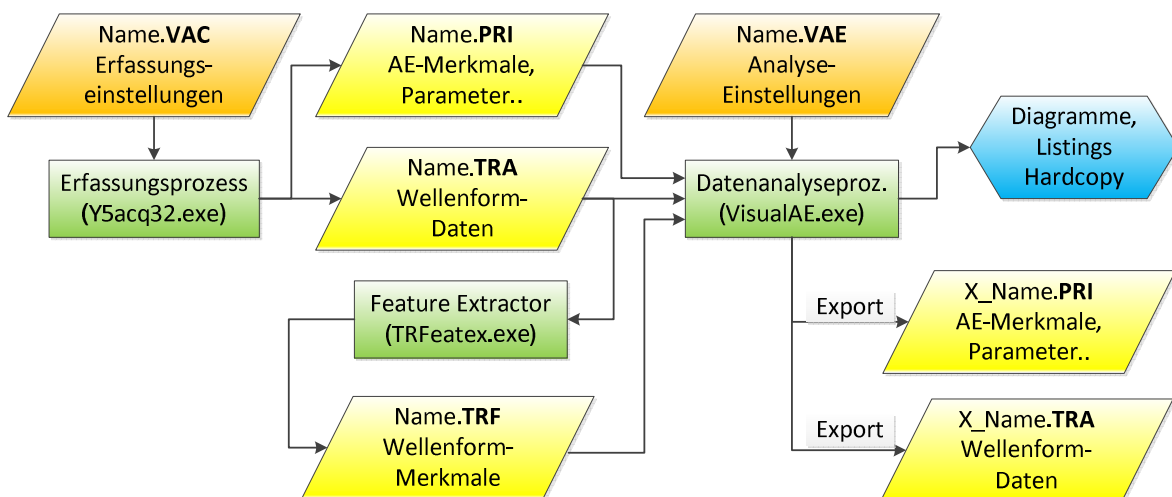


Abb. 1 Datenverarbeitungsprozesse und Dateitypen (bis R2012.0509)

Mit der Umstellung auf das neue Datenbankformat erhalten die Dateien eine andere Dateierweiterung, die "DB" enthält. Dies ist für den Anwender die größte Neuerung. Dahinter steht, dass die neuen Dateien eine komplette Datenbankstruktur enthalten, die flexibel und selbsterklärend variiert und an neue Herausforderungen angepasst werden können.

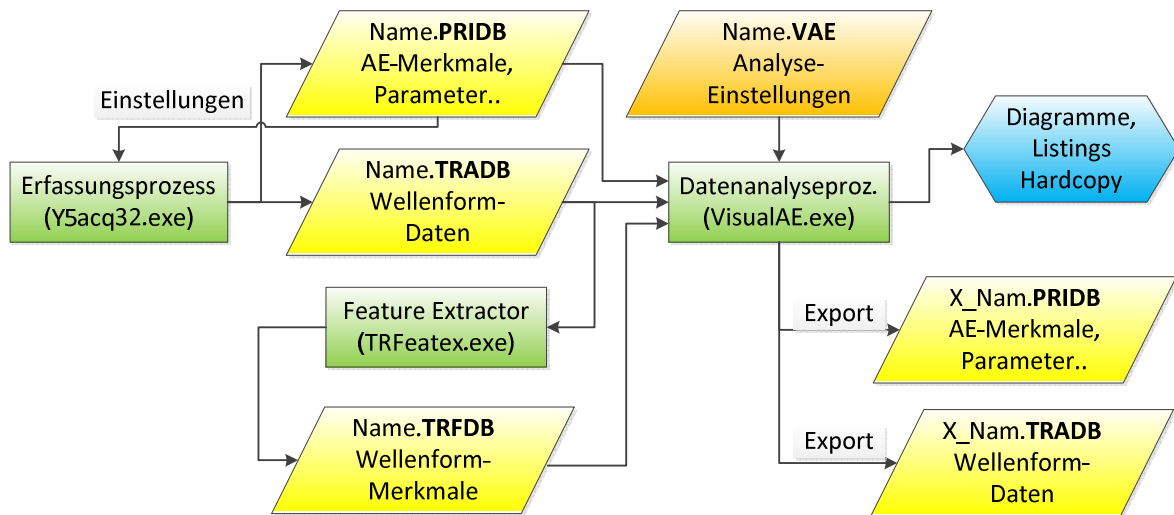


Abb. 2 Dateitypen (PRIDB, TRADB, TRFDB) nach Umstellung auf Datenbankformat

Wesentlich ist, dass VisualAE, VisualTR und VisualClass auch nach der Umstellung noch die bisherigen Dateitypen PRI, TRA und TRF lesen und verarbeiten können. Dies wird durch Abb. 3 verdeutlicht. Die Datenexportfunktion von VisualAE generiert PRIDB und TRADB Formate.

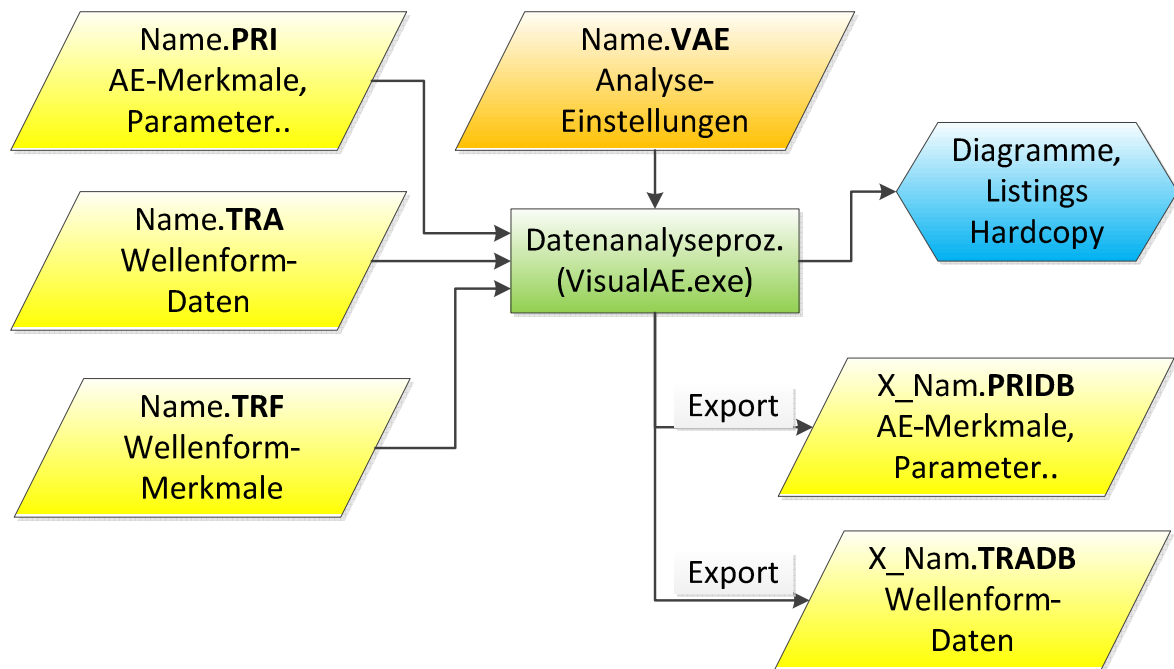


Abb. 3 VisualAE kann weiterhin PRI, TRA und TRF Dateien verarbeiten.

4. Alternative Dateiformate

Im Verlauf der Analyse der alternativen Dateiformate kamen zwei in die engere Wahl: Sqlite3 und HDF5.

4.1 Bibliothek und Dateiformat HDF5 (Hierarchical Data Format 5)

HDF5 ist ein Vielzweck-Dateiformat und eine Software-Bibliothek für die Speicherung aller Arten von wissenschaftlichen Daten. Siehe [2].

Allerdings haben wir festgestellt, dass HDF5 in damals verfügbarer Version auch zwei wesentliche Einschränkungen besitzt:

1. Die interne Struktur arbeitet mit konstanten Satzlängen, was bedeutet, dass es nicht möglich sein wird, kurze Bursts in kurzer Satzlänge und lange Bursts mit langer Satzlänge zu speichern.
2. HDF5 unterstützt nicht, dass mehrere Analysetasks - simultan zur Daten schreibenden Erfassung – die Daten lesen können.

4.2 Bibliothek SQLite und Dateiformat SQLite3

Nach www.sqlite.org ist SQLite die meist angewendete "Structured Query Language" (SQL) Datenbank-Bibliothek der Welt. Sie bietet folgende Vorteile:

- Sehr leistungsfähige Datenbanksprache und -bibliothek, leicht erlernbar, ermöglicht eine sich selbstbeschreibende Datenstruktur
- Einfacherer Zugriff (im Vergleich zu HDF5)
- Es sind viele kostenlose Werkzeuge verfügbar die gespeicherte Daten und ihre Struktur anzeigen können
- Wird in tausenden Anwendungen genutzt, d.h. es ist sehr stabil und ausgereift.

Nachteile im Vergleich zum binären Format:

- Erfordert höhere Rechenleistung, aber mit einer Dual Core CPU werden weiterhin über 100,000 Hits/s in Echtzeit erreicht
- Weniger kompakt, d.h. die Dateien werden etwas größer.

5. Der Nutzen von SQLite für die AE Technologie

- Datentransparenz: Die Daten können mit bewährten Werkzeugen wie Matlab ohne Konversion gelesen und verarbeitet werden. Kostenlose Werkzeuge gestatten Einsicht in Struktur und Daten, zB SQLite Expert. Siehe Abb. 4. Dort wird im rechten Teil eine Tabelle angezeigt, welche die unterstützten Variablen und Ihre Umrechnung vom gespeicherten Ganzzahlformat in die physikalische Einheit definiert. Alle gemessenen Daten stehen im Rohformat und in Einheiten umgerechnet in diversen Tabellen zur Verfügung.
- Flexibilität: Die Implementierung neuer Funktionen und Merkmale wird erleichtert: Fortschritt in AE wird beschleunigt
- Die Speicherlänge von Wellenformen kann an die Dauer eines Bursts angepasst werden. Kurze Bursts belegen weniger Speicher als lange. (Die komplette Realisierung erfolgt voraussichtlich in 2014)
- In VisualAE werden nun beschreibende Namen für 38 Wellenform-Merkmale angezeigt, z.B. "T2_SR5" für "Spectral Ratio of time segment 2 and frequency segment 5". Bisher war es nur möglich im TR Feature Extractor neun von 38 unterstützten Wellenform-Merkmalen einen von 9 abstrakten Namen ("FEx08" bis "FEx16") zuzuweisen und in VisualAE anzuzeigen.
- Beliebig viele externe Parameter können unterstützt werden. Erste Version unterstützt statt 8 nun 16 externe parametrische Eingänge (PA0-PA15)

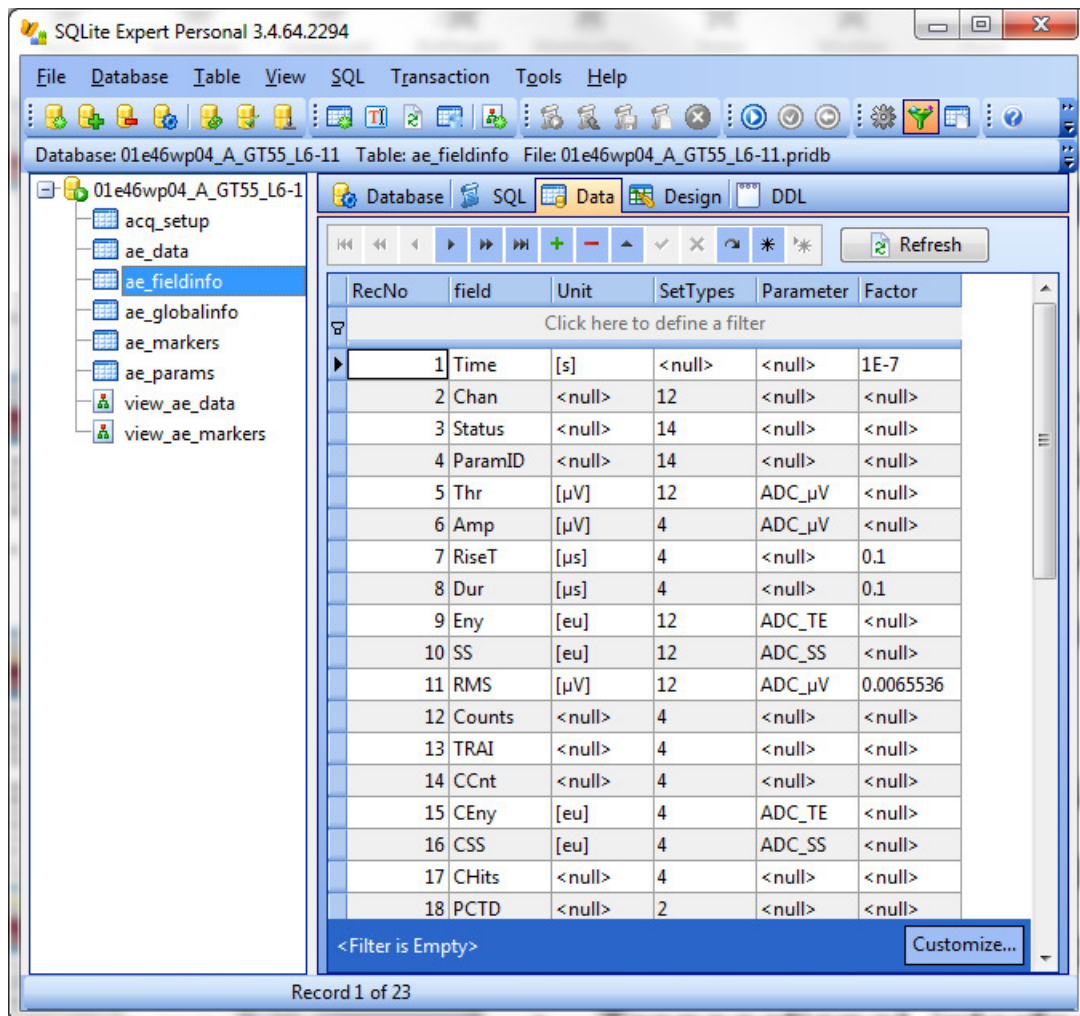


Abb. 4 Struktur und Daten können mit kostenlosen Werkzeugen angezeigt werden

- Energie und Signalstärke sind nun gleichzeitig verfügbar. Bisher musste sich der Anwender vor der Messung für eine der Größen entscheiden. Siehe Abb. 5.

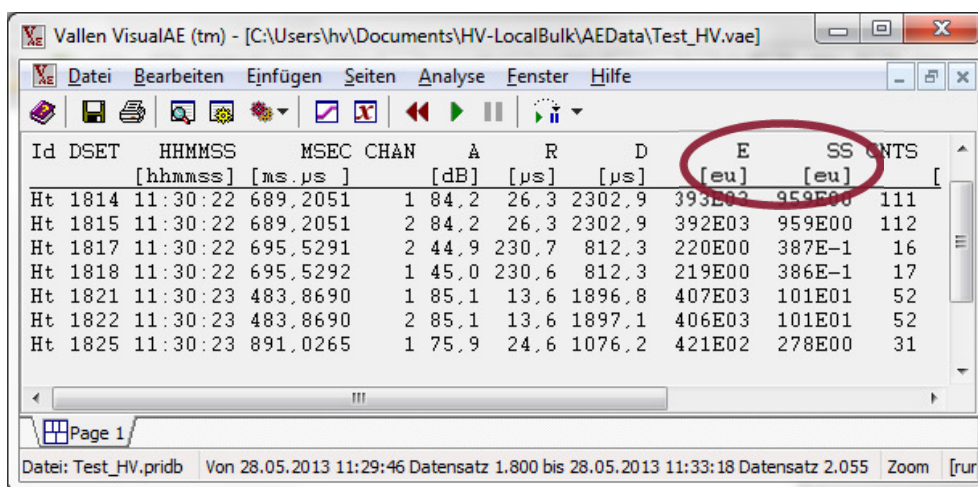


Abb. 5 Energie (E) und Signal Strength (SS) sind gleichzeitig verfügbar

- Der Zeitstempel umfasst nun bei 25 ns Auflösung eine Dynamik von Jahrhunderten; bisher ca 70 Tage bei gleicher Auflösung.
- Neue Datenarten, zB aus Wetterstationen, können unterstützt werden

- Nicht erforderliche Merkmale können abgeschaltet werden, dies ermöglicht höchstmögliche Datenraten und spart Speicherplatz (siehe Abb. 6).

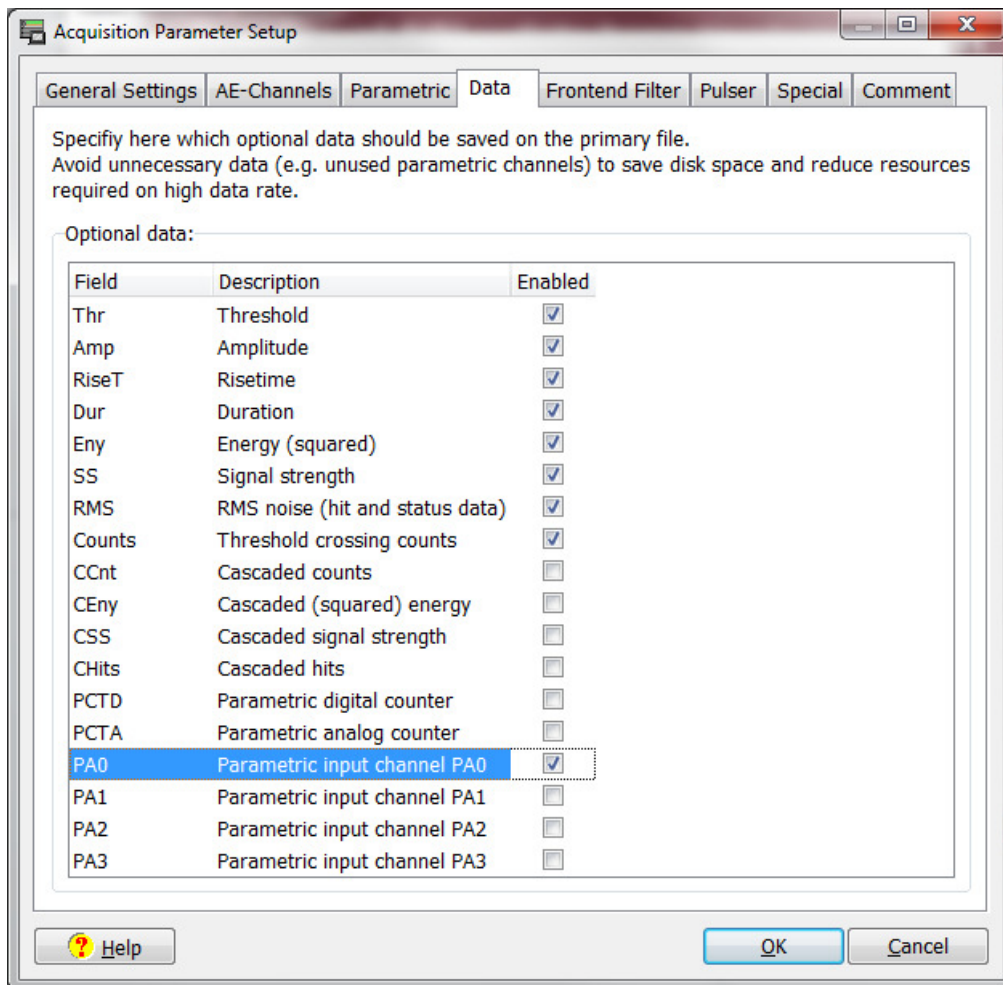


Abb. 6 Die Erfassung nicht erforderlicher Merkmale kann abgeschaltet werden.

6. Kompatibilität mit bisheriger Software

Wir erleben oft, dass zu viele Neuerungen die Anwender erschrecken, weil sie befürchten, sich nicht mehr auszukennen. Man kennt das ja von Microsoft-Produkten.

Dem Anwender der Vallen AE Suite Software wird es leicht gemacht, auf das neue Datenformat umzusteigen. Wie die Sequenz von Abb. 1 bis 3 zeigt, merkt der Anwender den Umstieg eigentlich nur an den Bezeichnungen von drei Dateitypen (Endungen), die nun nicht mehr PRI, TRA, und TRF heißen sondern PRIDB, TRADB und TRFDB. Mit dem "DB" in der Endung wird zum Ausdruck gebracht, dass jede Datei eine komplette und transparente Datenbank beinhaltet.

VisualAE, VisualTR und VisualClass können beide Formate lesen, der Anwender muss die Software nicht um-installieren, um alte Daten neu zu analysieren. Daten die von Softwareversionen ab 2013 erzeugt werden, können von vorherigen Softwareversionen nicht verarbeitet werden.

Im Vallen Control Panel (Abb. 7) sehen Sie das neue Werkzeug "PRI2DB". Es dient zur bequemen Konvertierung bisheriger PRI und TRA Dateien in PRIDB und TRADB Dateien. Nach einer Auswahl beliebig vieler Dateisätze können diese alle mit einem Klick konvertiert werden.

Die Menüs (das Graphical User Interface, GUI) sind nahezu unverändert geblieben, so kennt sich der Anwender gleich wieder aus.

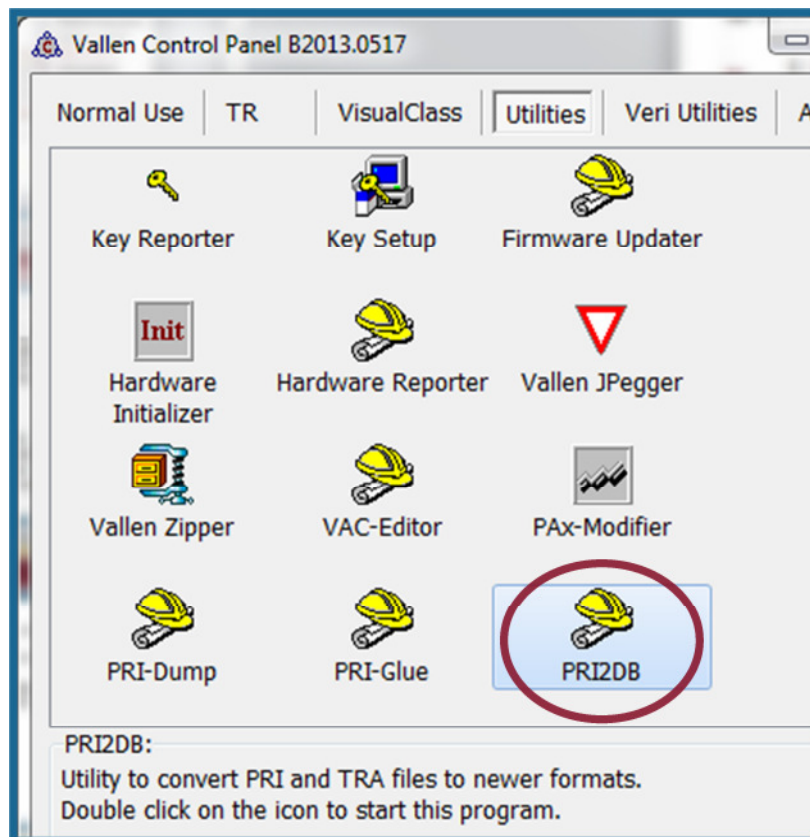


Abb 7 Menüs sind weitestgehend unverändert, PRI2DB ist ein neuer Daten-Konverter

7. Ausblick

Für Vallen Systeme GmbH war die Implementierung eines neuen Dateiformats über Jahre hinweg ein wesentliches Anliegen und eine große Aufgabe. Manche Kundenwünsche mussten warten, da sie mit dem bisherigen Format nicht effizient implementierbar waren.

Das neue Format gestattet nun auch "fremder" Software vollen Lesezugriff auf alle Datenarten, sogar in Echtzeit. Kunden können nun, unabhängig von uns als Hersteller, eigene Analyseprogramme entwickeln und neue Anwendungen erschließen.

Wir wünschen uns, dass das neue Format zum Erfolg unserer Kunden und zum weiteren Fortschritt der AE Technologie einen wichtigen Beitrag leisten wird.

Referenzen

[1] <http://www.sqlite.org>

[2] <http://www.lrz.de/services/software/mathematik/hdf5/>