

Unisens 2.0

How to create your own unisens.xml file

Malte Kirst

1. September 2010

Einführung

Dieses Dokument soll die Erstellung einer Unisens-Headerdatei anhand eines Beispiels erklären. Für das XML-Format spielt die Reihenfolge der Attribute keine Rolle. Optionale Attribute sind grau.

Das Modell für einen Unisens-Datensatz innerhalb des **unisens**-Elements ist:

```
(customAttributes?, context?,  
  (signalEntry | valuesEntry | eventEntry | customEntry)* , group* )
```

Für jedes Entry innerhalb eines Datensatzes gilt:

```
(customAttributes?,  
  (binFileFormat | csvFileFormat | xmlFileFormat | customFileFormat) )
```

XML-Elemente und -Attribute

XML Version

```
<?xml version="1.0" encoding="utf-8"?>
```

Dieser Teil ist immer zu Beginn einer validen XML-Datei notwendig, hier wird der Zeichensatz und die XML-Version definiert.

Root-Element

```
<unisens measurementId="smpleData" timestampStart="2009-10-01  
  T12:23:11" version="2.0" comment="Example data set, derived from  
  Holtervergleich_0001." xmlns="http://www.unisens.org/unisens2.0"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://www.unisens.org/unisens2.0 http://www.  
  unisens.org/unisens2.0/unisens.xsd">
```

Die **measurementId** ist eine Eindeutige ID für diesen Datensatz. In **timestampStart** wird der Aufnahmebeginn dieses Datensatzes festgehalten alle Daten-Einträge beginnen zu diesem Zeitpunkt. Das Format entspricht der ISO 8601. Die Angabe der Millisekunden ist optional. Ebenso optional ist die Angabe des Attributes **comment**, der einen Kommentar zu diesem Datensatz speichert. Für manche Datensätze kann es sinnvoll sein, die

Gesamtlänge anzugeben. Hierfür ist das optionale Attribut `duration` gedacht. Die Gesamtlänge wird immer in Sekunden angegeben. Das Attribut `version` bezeichnet die Version der Unisens-Spezifikation, mit der dieser Datensatz geschrieben wurde. Die Attribute `xmlns`, `xmlns:xsi` und `xsi:schemaLocation` sind in dieser Form notwendig, damit Unisens-Anwendungen das XML-Schema richtig interpretieren können.

Eigene Attribute

```
<customAttributes>
  <customAttribute key="evaluated" value="yes" />
  <customAttribute key="originalDataSet" value="
    Holtervergleich_0001" />
</customAttributes>
```

Dieser Teil ist optional. Mit Hilfe des Elements `customAttributes` können Schlüssel-Wert-Paare zu einem Datensatz gespeichert werden. Dabei ist die Zeichenkette in `key` der Schlüssel und die in `value` der zugehörige Wert. Die Anzahl der `customAttribute`-Elemente ist nicht limitiert.

`customAttributes` können sowohl global für den gesamten Datensatz angelegt werden als auch innerhalb von einem `*Entry`.

Kontext-Informationen

```
<context schemaUrl="patient.xsd" />
```

Dieser Teil ist optional. In einer separaten XML-Datei mit dem Namen `context.xml` und dem in `schemaUrl` definierten Schema können weitere Daten zu diesem Datensatz abgelegt werden. Pro Datensatz kann es nur eine `context.xml` geben.

Signale

```
<signalEntry adcResolution="16" baseline="0" adcZero="0" comment="
  FhG-Elektroden" contentClass="ECG" dataType="int16" id="
  ecg_m500_250.bin" lsbValue="1" sampleRate="250" source="FhG-
  Elektroden, ADS8345" unit="uV">
  <binFileFormat endianness="LITTLE" />
  <channel name="CH1" />
  <channel name="CH2" />
</signalEntry>
```

Das Element `signalEntry` ist für alle quasi-kontinuierlich abgetasteten Messdaten gedacht. Als Dateiformat bietet sich hier ein binärformat an, da dies sehr kompakt ist. Es kann aber auch XML oder CSV verwendet werden, was jedoch bei großen Datensätzen zu Geschwindigkeitseinbußen und sehr großen Dateien führen kann.

Die `contentClass` gibt die Inhaltsklasse der gespeicherten Daten an, zum Beispiel ECG für EKG-Werte oder BLOODPRESSURE für Blutdruckwerte. In Tabelle 1 ist eine Liste mit Empfehlungen für verschiedene Inhaltsklassen. Die `id` enthält den vollständigen Dateinamen der zugehörigen Datendatei im selben Ordner. Die Abtastrate des Signals wird in `sampleRate` abgelegt, die Abtastrate kann auch eine Fließkommazahl sein. Der Datentyp, in dem die Daten vorliegen, wird in `dataType` gespeichert. Dies ist auch der Datentyp, in dem die Daten von der Unisens-Library zurückgegeben werden (s. Tab. 2).

Ein Kommentar kann optional in **comment** angegeben werden, ein Name für diesen Eintrag kann optional in **name** vergeben werden.

Der verwendete AD-Konverter (ADC) wird mit den folgenden Attributen beschrieben: **adcResolution** gibt die Auflösung in Bit an und **lsbValue** ist das Least Significant Bit (LSB) des ADC in der unter **unit** angegebenen Einheit. Ist das LSB unbekannt, wird hier 1 angegeben. Eine Liste der empfohlenen Einheiten ist in Tabelle 3 angegeben. Die **baseline** ist der Ausgangswert des ADC, wenn am Eingang der physikalische Wert 0 der angegebenen Einheit anliegt (also z. B. 0 V bei einem Spannungsmesser oder 0 g bei einem Beschleunigungsmesser). Hiermit kann also ein Offset des ADC eingestellt werden. Das Attribut **adcZero** beinhaltet den Ausgangswert des ADC, wenn am Eingang genau in der Mitte steht. Bei bipolaren ADC mit Zweierkomplement-Ausgang ist dieser Wert normalerweise 0.

Das optionale Attribut **source** enthält die Beschreibung der Datenquelle, üblicherweise ist dies der Sensorname. Dies kann im optionalen Attribut **sourceId** weiter spezifiziert werden, wenn der Sensor eine eigene ID hat (dies kann z. B. eine MAC-Adresse sein).

Um Dateien im Binärformat einzubinden, muss das **binFileFormat** definiert werden. Neben einem optionalen Kommentar in **comment** wird in **endianess** die Byte-Reihenfolge angegeben, entweder LITTLE für little endian oder BIG für big endian. Bei Windows-Systemen und Intel-Prozessoren hat man es normalerweise mit little endian zu tun, während Motorola-Prozessoren oft mit big endian arbeiten.

Für jeden Kanal des Signals wird ein **channel**-Element angelegt. Im Attribut **name** ist der Name des Kanals angegeben.

Einzelmesswerte

```
<valuesEntry contentClass="BLOODPRESSURE" id="bloodpressure.csv"
  comment="Blutdruck" sampleRate="1" lsbValue="1" unit="mmHg"
  dataType="int16">
  <csvFileFormat comment="csv, 2 channel" separator="," />
  <channel name="systolisch" />
  <channel name="diastolisch" />
</valuesEntry>
```

Das Element **valuesEntry** ist für alle nicht-kontinuierlich abgetasteten Messdaten gedacht. Als Dateiformat bietet sich hier CSV an, da dies menschenlesbar, übersichtlich und kompakt ist. Es kann aber auch XML oder ein Binärformat verwendet werden. Die Attribute entsprechen alle dem **signalEntry**. Die Abtastfrequenz in **sampleRate** dient hier als Zeitbasis für einen samplegenauen Datenzugriff.

Werden die Daten als CSV-Datei gespeichert, muss das **csvFileFormat** angegeben werden. Hierfür kann das Trennzeichen in der CSV-Datei im Attribut **separator** definiert werden, Standard ist ein Semikolon. Der Kommentar **comment** ist optional, ebenso die Angabe eines Dezimaltrennzeichens **decimalSeparator** innerhalb der Werte (standardmäßig ein Punkt).

Einzelmesswerte können in mehreren Kanälen vorliegen, diese werden entsprechend den Signalen angegeben.

Ereignisse

```
<eventEntry comment="Reference trigger list" contentClass="TRIGGER
  " id="trigger_reference.csv" sampleRate="1000" source="PADSY /
  M. KIRST" typeLength="1">
```

```

    <csvFileFormat separator=";" />
  </eventEntry>

```

Das Element `eventEntry` soll bei nicht messtechnisch erfassten Ereignissen verwendet werden. In diesem Element können keine ADC-spezifischen Attribute verwendet werden, ansonsten entsprechen die Attribute denen von Einzelmesswerten oder Signalen. Die Abtastfrequenz in `sampleRate` dient hier ebenfalls als Zeitbasis für einen samplegenauen Datenzugriff.

Werden Ereignisse im Binärformat gespeichert, sind die Attribute `typeLength` und `commentLength` anzugeben. Diese geben die Länge der Zeichenkette des Ereignisses und dessen optionalen Kommentars an. Alle Ereignisse und Kommentare müssen in diesem Fall die gleiche Länge haben und dürfen sie nicht überschreiten.

Um dieses zu umgehen, bietet es sich für Ereignisse an, nicht das Binärformat zu verwenden, sondern CSV- oder XML-Dateien. Dafür muss ein entsprechendes Element angelegt werden. Im Falle einer XML-Datei ist dies das `xmlFileFormat`, das außer einem optionalen `comment` keine weiteren Attribute hat.

Proprietäre Daten

```

<customEntry id="P1010061.JPG" contentClass="IMAGE">
  <customAttributes>
    <customAttribute key="mime" value="img/jpeg" />
  </customAttributes>
  <customFileFormat fileName="jpg" comment="picture from
    digital camera" />
</customEntry>

```

Proprietäre Daten, die nicht von Unisens unterstützt werden, können als `customEntry` angelegt werden. Dieses können zum Beispiel weitere Datensätze in eigenen Formaten sein (z. B. EDF) oder auch Bilder oder Videos von den Messungen. In diesem Beispiel ist das `customEntry` noch mit `customAttributes` versehen, was aber nicht notwendig ist. Ein `customEntry` hat in den meisten Fällen ein freies Dateiformat, das mit `customFileFormat` angegeben wird. Hier kann das Format mit `fileName` näher spezifiziert werden, so dass Anwendungen darauf reagieren können.

Gruppierungen

```

<group id="default_ecg" comment="ECG and trigger list">
  <groupEntry ref="ecg_m500_250.bin" />
  <groupEntry ref="trigger_reference.csv" />
</group>

```

Dieser Teil ist optional. Hiermit können mehrere Einträge zu einer Gruppe zusammengefasst werden. Die Anzahl der Gruppen und die Anzahl der Elemente einer Gruppe ist nicht limitiert. Eine Gruppe `group` hat eine eindeutige ID `id` und kann mit einem Kommentar `comment` versehen werden. Die einzelnen Einträge `groupEntry` verweisen mit `ref` jeweils auf die ID eines `SignalEntry`, `ValuesEntry`, `EventEntry` oder `CustomEntry`.

Root-Element schließen

```

</unisens>

```

Am Ende der Datei muss selbstverständlich das Root-Element geschlossen werden.

Konstanten und Empfehlungen

Tabelle 1: Empfehlung für die Anwendung des Attributes `contentClass`

Klasse	Beschreibung
ECG	EKG-Signal
TRIGGER	Trigger-Liste (EKG-Annotation)
ACC	Beschleunigungssignal
IMP	Impedanzsignal
RAW	Rohdaten
BLOODPRESSURE	Blutdruck
PLETH	Pleth-Signal
RESP	Atmung
MARKER	Patienten-Marker
ARTIFACT	Liste mit Artefakt-Markern

Tabelle 2: Von Unisens spezifizierte Datentypen

Datentyp	Größe (Byte)	Wertebereich
double	8	$4.9 \cdot 10^{-324} \dots 1.7976931348623157 \cdot 10^{308}$
float	4	$1.4 \cdot 10^{-45} \dots 3.4028235 \cdot 10^{38}$
int32	4	$-2147483648 \dots 2147483647$
int16	2	$-32768 \dots 32767$
int8	1	$-128 \dots 127$
uint32	4	$0 \dots 4294967295$
uint16	2	$0 \dots 65535$
uint8	1	$0 \dots 255$

Tabelle 3: Empfehlung für physikalische Einheiten

Einheit	Beschreibung
uV, μ V, mV, V, kV	Spannung
g	Beschleunigung
°C	Temperatur

Beispiel-Datei

```
<?xml version="1.0" encoding="utf-8"?>
<unisens measurementId="sampleData" timestampStart="2009-10-01
  T12:23:11" version="2.0" comment="Example data set, derived from
  Holtervergleich_0001." xmlns="http://www.unisens.org/unisens2.0"
```

```

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.unisens.org/unisens2.0 http://www.
unisens.org/unisens2.0/unisens.xsd">
<customAttributes>
  <customAttribute key="evaluated" value="yes" />
  <customAttribute key="originalDataSet" value="
    Holtervergleich_0001" />
</customAttributes>
<context schemaUrl="patient.xsd" />
<signalEntry adcResolution="16" baseline="0" adcZero="0" comment="
  FhG-Elektroden" contentClass="ECG" dataType="int16" id="
  ecg_m500_250.bin" lsbValue="1" sampleRate="250" source="FhG-
  Elektroden, ADS8345" unit="uV">
  <binFileFormat endianness="LITTLE" />
  <channel name="CH1" />
  <channel name="CH2" />
</signalEntry>
<signalEntry adcResolution="16" baseline="0" adcZero="0" comment="
  Holter-EKG mit Ambu Bluesensor Klebeelektroden, Downsampling"
  contentClass="ECG" dataType="int16" id="ecg_padsy_250.bin"
  lsbValue="3.417" sampleRate="250" source="TELESMART H/P" unit="
  uV">
  <customAttributes>
    <customAttribute key="numberOfVes" value="22" />
    <customAttribute key="schnipp" value="schnapp" />
  </customAttributes>
  <binFileFormat endianness="LITTLE" />
  <channel name="A" />
  <channel name="B" />
</signalEntry>
<signalEntry adcResolution="12" baseline="2048" adcZero="0"
  comment="ACC sensor connection board (textile)" contentClass="
  ACC" dataType="int16" id="acc_textile_50.bin" lsbValue="0.00294
  " sampleRate="50" source="ST Microelectronics MEMS Inertial
  Sensor" sourceId="LIS3LV02DQ" unit="g">
  <binFileFormat comment="binary, 3 channels" endianness="LITTLE" /
  >
  <channel name="x" />
  <channel name="y" />
  <channel name="z" />
</signalEntry>
<eventEntry comment="Reference trigger list" contentClass="TRIGGER
  " id="trigger_reference.csv" sampleRate="1000" source="PADSY /
  M. Kirst" typeLength="1">
  <csvFileFormat separator=";" />
</eventEntry>
<customEntry id="P1010061.JPG" contentClass="IMAGE">
  <customAttributes>
    <customAttribute key="mime" value="img/jpeg" />
  </customAttributes>
  <customFileFormat fileName="jpg" comment="picture from
  digital camera" />
</customEntry>
<valuesEntry contentClass="BLOODPRESSURE" id="bloodpressure.csv"

```

```

        comment="Blutdruck" sampleRate="1" lsbValue="1" unit="mmHg"
        dataType="int16">
    <csvFileFormat comment="csv, 2 channel" separator="," />
    <channel name="systolisch" />
    <channel name="diastolisch" />
</valuesEntry>
<group id="default_ecg" comment="ECG and trigger list">
    <groupEntry ref="ecg_m500_250.bin" />
    <groupEntry ref="trigger_reference.csv" />
</group>
</unisens>

```