

Change Request zu SIP Schema arelda_v3.13.2

Einleitung	1
1 Change Request <i>bentō</i>	2
2 Zusätzliche Schema-Änderung	6
3 Beilagen	7

Einleitung

Die Aufgabestellung in *bentō* "3.2.3 Metadaten" Punkt 6 lautet:

"Zusätzlich muss es im SIP eine Möglichkeit geben, weitere Metadaten jeglicher Art aus dem Quellsystem als Key-Value-Paare festzuhalten. Damit besteht zwar die Gefahr, im SIP (und folglich im AIP) einer unkontrollierten Entwicklung der Metadaten den Weg zu ebnen. Eine Möglichkeit, im SIP und AIP zusätzlich zu den minimal spezifizierten weitere Metadaten festzuhalten, ist jedoch ein breites Desiderat und somit unabdingbar."

Eine Lösung für dieses Problem, die keine Einschränkungen und Anpassungen am BAR PackageHandler wie auch an anderen, bereits entwickelten Tools bedingt (Abwärtskompatibilität), und nur wenige Änderungen am heutigen BAR-SIP *arelda_v3.13.2* verlangt, ist nicht einfach. Der Vorschlag in der Folge erfüllt diese Forderung bedingt.

Die vorgeschlagene Lösung bettet strukturierte Informationen in einem neuen optionalen Knoten **<zusatzDaten>** ein, der unter **<ordnungssystemposition>**, **<dossier>** und **<dokument>** im SIP eingehängt werden kann. **<zusatzDaten>** erlaubt strukturierte Darstellung von Daten als eine Liste von **Key-Value-Paaren**. Unstrukturierter Text soll grundsätzlich weiterhin im Element **<bemerkung>** untergebracht werden.

Der Knoten **<zusatzDaten>** ist bei der Entität Ordnungssystemposition in der Reihenfolge (*sequence*) hinter dem Knoten **<sonstigeBestimmungen>**, bei den Entitäten Dossier und Dokument aber hinter dem Knoten **<bemerkung>** angelegt. Die Positionierung in der Reihenfolge ist eigentlich ohne Bedeutung, wird aber durch das Schema festgelegt.

Weil ein neuer Knoten **<zusatzDaten>** in der XML Datei auftaucht, ist es erst nach Anpassung bzw. Erweiterung bestehender Ingest-Tools möglich, den Inhalt dieser strukturierten Daten zu sehen und zu validieren.

1 Change Request *bentō*

Es wurde bewusst versucht, die Schemaerweiterung sprachlich an die (deutschsprachigen) areldaTag-Namen anzupassen. Folgende Übersetzungen für Begriffe aus den METS und EAD Schemata wurden gewählt:

otherData	zusatzDaten
property	merkmal
key	schlüssel
value	wert

1.1 Schema-Erweiterungen

An der Master Schemadatei *arelda_v3.13.2.xsd* müssen vier Änderungen vorgenommen werden. Zusätzlich kommt in den Ordner metadata/xsd eine neue Schemadatei *zusatzDaten.xsd*.

1.1.1 metadata.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<paket xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://bar.admin.ch/arelda/v3.13.2"
xsi:schemaLocation="http://bar.admin.ch/arelda/v3.13.2 xsd/arelda_v3.13.2_bento_CR.xsd" xsi:type="paketSIP">
```

1.1.2 Schema: *arelda_v3.13.2_bento_CR.xsd*

Vier Änderungen an *arelda_v3.13.2.xsd*:

```
<xs:include schemaLocation="zusatzDaten.xsd"/>

<xs:complexType name="ordnungssystemposition" abstract="true">
  <xs:sequence>
    <xs:element name="federfuehrendeOrganisationseinheit" type="federfuehrendeOrganisationseinheitOrdnungssystemposition" minOccurs="0"/>
    <xs:element name="klassifizierungskategorie" type="klassifizierungskategorieOrdnungssystemposition" minOccurs="0"/>
    <xs:element name="datenschutz" type="datenschutzOrdnungssystemposition" minOccurs="0"/>
    <xs:element name="oeffentlichkeitsstatus" type="oeffentlichkeitsstatusOrdnungssystemposition" minOccurs="0"/>
    <xs:element name="oeffentlichkeitsstatusBegrueundung" type="oeffentlichkeitsstatusBegrueundungOrdnungssystemposition" minOccurs="0"/>
    <xs:element name="sonstigeBestimmungen" type="sonstigeBestimmungenOrdnungssystemposition" minOccurs="0"/>
    <xs:element name="zusatzDaten" type="zusatzDaten" minOccurs="0"/>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="dossier" abstract="true">
  <xs:sequence>
    <xs:element name="zusatzmerkmal" type="zusatzmerkmal" minOccurs="0"/>
  </xs:sequence>
</xs:complexType>
```

```

<xs:element name="titel" type="titelDossier"/>
<xs:element name="inhalt" type="inhalt" minOccurs="0"/>
<xs:element name="formInhalt" type="formInhalt" minOccurs="0"/>
<xs:element name="erscheinungsform" type="erscheinungsformDossier" minOccurs="0"/>
<xs:element name="federfuehrendeOrganisationseinheit" type="federfuehrendeOrganisationseinheitDossier" minOccurs="0"/>
<xs:element name="entstehungszeitraum" type="historischerZeitraum"></xs:element>
<xs:element name="entstehungszeitraumAnmerkung" type="entstehungszeitraumAnmerkung" minOccurs="0"/>
<xs:element name="klassifizierungskategorie" type="klassifizierungskategorieDossier" minOccurs="0"/>
<xs:element name="datenschutz" type="datenschutzDossier" minOccurs="0"/>
<xs:element name="oeffentlichkeitsstatus" type="oeffentlichkeitsstatusDossier" minOccurs="0"/>
<xs:element name="oeffentlichkeitsstatusBegrueundung" type="oeffentlichkeitsstatusBegrueundungDossier" minOccurs="0"/>
<xs:element name="sonstigeBestimmungen" type="sonstigeBestimmungenDossier" minOccurs="0"/>
<xs:element name="bemerkung" type="bemerkungDossier" minOccurs="0"/>
<xs:element name="zusatzDaten" type="zusatzDaten" minOccurs="0"/>
</xs:sequence>
<xs:attribute name="id" type="idDossier" use="required"/>
</xs:complexType>

<xs:complexType name="dokument" abstract="true">
  <xs:sequence>
    <xs:element name="titel" type="titelDokument"/>
    <xs:element name="autor" type="autor" minOccurs="0" maxOccurs="unbounded" ecore:name="autoren"/>
    <xs:element name="erscheinungsform" type="erscheinungsformDokument"/>
    <xs:element name="dokumenttyp" type="dokumenttyp" minOccurs="0"/>
    <xs:element name="registrierdatum" type="historischerZeitpunkt" minOccurs="0"></xs:element>
    <xs:element name="entstehungszeitraum" type="historischerZeitraum" minOccurs="0">
    <xs:element name="klassifizierungskategorie" type="klassifizierungskategorieDokument" minOccurs="0"/>
    <xs:element name="datenschutz" type="datenschutzDokument" minOccurs="0"/>
    <xs:element name="oeffentlichkeitsstatus" type="oeffentlichkeitsstatusDokument" minOccurs="0"/>
    <xs:element name="oeffentlichkeitsstatusBegrueundung" type="oeffentlichkeitsstatusBegrueundungDokument" minOccurs="0"/>
    <xs:element name="sonstigeBestimmungen" type="sonstigeBestimmungenDokument" minOccurs="0"/>
    <xs:element name="bemerkung" type="bemerkungDokument" minOccurs="0"/>
    <xs:element name="zusatzDaten" type="zusatzDaten" minOccurs="0"/>
    <xs:element name="dateiRef" type="xs:IDREF" minOccurs="0" maxOccurs="unbounded" ecore:name="dateien"
      ecore:unsettable="true" ecore:reference="datei"/>
  </xs:sequence>
  <xs:attribute name="id" type="idDokument" use="required"/>
</xs:complexType>

```

1.1.3 Schema: zusatzDaten.xsd

Zusätzliche Schemadatei:

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified" attributeFormDefault="unqualified">
  <xs:complexType name="zusatzDaten" mixed="false">
    <xs:annotation>
      <xs:documentation>Merkmal - Werte Gruppen erlaubt das festhalten weiterer Metadaten in tabellarisch strukturierter Form</xs:documentation>
    </xs:annotation>
    <xs:sequence>
      <xs:element name="merkmal" maxOccurs="unbounded">
        <xs:complexType>
          <xs:simpleContent>
            <xs:extension base="xs:string">
              <xs:attribute name="name" type="xs:string" use="required">
                <xs:annotation>
                  <xs:documentation>Name des Merkmal - Werte Paares</xs:documentation>
                </xs:annotation>
              </xs:attribute>
              <xs:attribute name="order" type="xs:integer" use="optional">
                <xs:annotation>
                  <xs:documentation>Das Attribut erlaubt das Ordnen einer Reihe gleicher Merkmale</xs:documentation>
                </xs:annotation>
              </xs:attribute>
            </xs:extension>
          </xs:simpleContent>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```

1.1.4 Beispielausschnitt aus metadata.xml

Folgende Tabelle wird in *metadata.xml* als Zusatzdaten zum Dossier 22.06.12 festgehalten. Man beachte, dass Titel im Original ein Zweizeiler ist:

Titel	X. Nachtrag zum Volksschulgesetz
Titel	Zusammenfassung
Erscheinungsjahr	2007

```

<dossier id="DS-0">
  <titel>X. Nachtrag zum Volksschulgesetz 123</titel>
  <erscheinungsform>nicht digital</erscheinungsform>
  <klassifizierungskategorie>nicht klassifiziert</klassifizierungskategorie>
  <datenschutz>>false</datenschutz>
  <zusatzDaten>
    <merkmal name="Titel" order="1">X. Nachtrag zum Volksschulgesetz</merkmal>
    <merkmal name="Titel" order="2">Zusammenfassung</merkmal>
    <merkmal name="Erscheinungsjahr">2007</merkmal>
  </zusatzDaten>
  <aktenzeichen>22.06.12</aktenzeichen>

```

1.2 Validieren und bearbeiten mit bestehenden Tools

Für die Validierung mit bereits bestehenden und noch nicht angepassten Tools reicht leider eine Transformation der strukturierten Daten in CDATA¹ Daten nicht. Der Grund dafür ist, dass an der Stelle im arelda Schema keine Daten, sondern nur Knoten erlaubt sind und nicht geparste Daten dennoch den Status von Text annehmen. Die einzige Lösung ist die Transformation zu XML Kommentar (Achtung XML-Kommentar² und nicht XML-Annotation³).

In transformierter Form:

```

<dossier id="DS-0">
  <titel>X. Nachtrag zum Volksschulgesetz 123</titel>
  <erscheinungsform>nicht digital</erscheinungsform>
  <klassifizierungskategorie>nicht klassifiziert</klassifizierungskategorie>
  <datenschutz>>false</datenschutz>
  <!-- <zusatzDaten>
    <merkmal name="Titel" order="1">X. Nachtrag zum Volksschulgesetz</merkmal>
    <merkmal name="Titel" order="2">Zusammenfassung</merkmal>
    <merkmal name="Erscheinungsjahr">2007</merkmal>
  </zusatzDaten> -->
  <aktenzeichen>22.06.12</aktenzeichen>

```

¹ Als CDATA markierte Daten werden vom XML Parser nicht geparst <http://www.w3.org/TR/REC-xml/#sec-cdata-sect>

² Siehe <http://www.w3.org/TR/REC-xml/#sec-comments>

³ XML Schema Annotation beschreiben ein Schema <http://www.w3.org/TR/xmlschema-0/#CommVers>

2 Zusätzliche Schema-Änderung

Folgende Schema-Änderung ist aus Kompatibilitätsgründen mit *xmllint* gewünscht:

2.1 "the base type is expected to be a complex type"

Um eine Erweiterung von "eigenschaft" mit Attributen zu realisieren, muss "eigenschaftDatei" als *complexType* definiert werden und von einem *base type* abstammen, der ebenfalls als *complexType* definiert ist. Ein aus einem *simpleType* abgeleiteter *complexType* wiederum ist immer eine *extension* zum *simpleType*, keine *restriction*. Die gewünschte Änderung ist rein formal und hat keinen Einfluss auf die gewünschte Funktionalität.

arelda_v3.13.2.xsd

```
951 <xs:complexType name="eigenschaftDatei">
952   <xs:complexContent>
953     <xs:extension base="eigenschaft">
954       <xs:attribute name="name" type="text4" use="required"/>
955     </xs:extension>
956   </xs:complexContent>
957 </xs:complexType>
```

datei.xsd (original)

```
xmllint.exe -noout -schema xsd\arelda_v3.13.2.xsd metadata.xml
```

```
xsd/arelda_v3.13.2.xsd:951: element complexType: Schemas parser error: complex type 'eigenschaftDatei':
```

```
If using <complexContent>, the base type is expected to be a complex type.
```

```
The base type '{http://bar.admin.ch/arelda/v3.13.2}eigenschaft' is a simple type.
```

```
<xs:simpleType name="eigenschaft">
  <xs:annotation>
    <xs:documentation>Generisches Attribut, welches im Zusammenhang mit Wert verwendet wird. Wird bspw. eine CSV-Datei abgeliefert, dann können
    mittels Attributpaar "eigenschaft" und "wert" Feld- und Zeilendelemter, Texterkennungszeichen oder Spaltennamen angegeben werden.
    </xs:documentation>
  </xs:annotation>
  <xs:restriction base="text4"/>
</xs:simpleType>
```

datei.xsd (mit gewünschter Änderung)

```
xmllint.exe -noout -schema xsd\arelda_v3.13.2_bento_CR.xsd metadata.xml  
metadata.xml validates
```

```
<xs:complexType name="eigenschaft">  
  <xs:annotation>  
    <xs:documentation>Generisches Attribut, welches im Zusammenhang mit Wert verwendet wird. Wird bspw. eine CSV-Datei abgeliefert, dann können  
    mittels Attributpaar "eigenschaft" und "wert" Feld- und Zeilendelemter, Texterkennungszeichen oder Spaltennamen angegeben werden.  
  </xs:documentation>  
  </xs:annotation>  
  <xs:simpleContent>  
    <xs:extension base="text4"/>  
  </xs:simpleContent>  
</xs:complexType>
```

3 Beilagen

In der Beilage `arelda_v3.13.2_bento_CR.zip` befindet sich ein Beispielablieferung mit den entsprechend angepassten Schemadateien und die geänderten Schemadateien in `arelda_v3.13.2_bento_CR.schema.zip`.