

Beschreibung der XML-Dateien der Retrodigitalisierung von NDB und ADB

Version	Datum	Änderung	Verantwortlich
0.1	2018-06-06	Erstfassung	Matthias Reinert

Die unter data.deutsche-biographie.de ausgelieferte XML-Dateistruktur orientiert sich an der TEI-P5 Empfehlung: es werden *vorwiegend* Elemente aus dem core Modul genutzt.

Abschnitte

Die Hauptabschnitte im XML korrespondieren mit dem Inhalt

Inhaltstyp	Abschnitt	Inhalte
Allgemeines	<teiHeader>	Rechte, Änderungen, Teilnehmer des Projekts
Registereintrag	<particDesc> <profileDesc>	Personen, Relationen Orte, Relationen
wenn vorhanden:	<body>	
NDB-Artikel	<div subtype="ndb">	Artikelinhalte ergänzt um Angaben zum Vorgänger und Nachfolger im NDB Band
ADB-Artikel	<div subtype="adb">	Artikelinhalte ergänzt um Angaben zum Vorgänger und Nachfolger im ADB Band

Der Abschnitt <teiHeader>

Statischer Inhalt, absichtlich in allen XML-Dateien redundant aufgeführt, da diese einzeln ausgeliefert werden

Der Abschnitt <fileDesc>

<respStmt> listet die Teilnehmenden in verschiedenen Projektabschnitten auf, sofern sie identifizierbare Änderungen an den XML-Tags vorgenommen haben, ist ihr Kürzel verzeichnet, das – in einer internen XML-Fassung – unter @resp mitgeführt wird

<publicationStmt> listet unter <availability> die drei Lizenzmöglichkeiten auf

- CC-BY-NC-SA für die ADB Artikel
- CC-BY-NC-ND für die NDB Artikel
- CC-0 für die Metadaten / alle anderen Daten

Der Abschnitt <profileDesc>

für <particDesc> und <settingsDesc>

die genannten Personen – bis auf biographierte Person/Familie – werden nicht redundant aufgeführt, da sie eigene Instanzen in XML haben, dagegen werden die Orte redundant aufgeführt, sie haben noch keine eigenen Instanzen

Der Abschnitt <particDesc>

Unter <listPerson> wird bisher nur die biographierte Person aufgeführt, da die anderen Personen/Familien in ihren eigenen XML-Dateien beschrieben werden.

< person@role =entry	Eigenschaften der biographierten Person/Familie
@sex	1 = Mann, 2 = Frau, 9 = Familie, 0 = unbekannt
@n	Schlüsselfeldzahl „sfz[0-9][0-9_-]+“
@xml:id	GND Nummer „ppn[0-9-X]+“
< persName@default >	Namensansetzung
< persName@search >	Suchvarianten des Namens
< persName@sort >	Sortiervariante des Namens
<floruit>	Angabe zum Wirkungszeitraum
<occupation>	Semikolon separierte Liste; Angabe einer Berufsklassifikation im Attribut @
<faith>	Semikolon separierte Liste
<note type="adb ndb"> <rs> @type @ref <name	Registerangaben zu einer Person – Referenz auf die @type=adb ndb mit dem Ziel @ref=sfz...

<pre> subtype="author" type="adb ndb"> <bibl type="adb ndb"> <list@type <item@n <list> <ref target="URL">Seitenanga be ggf. Ergänzungen</ref> </pre>	<ul style="list-style-type: none"> - Name des Autors @type=adb ndb - Zitierangabe mit Bandangaben je nach @type=adb ndb - - geschachtelte Liste um die Struktur NDB > Band > Seitereferenz abzubilden als list@type=(a n)db > item@n=BANDNR > list > item > ref+
<pre> <birth> <death> <date> <placeName> </pre>	
<pre> <listRelation type="spatial"> <relation name="" active="Biographierte r" passive="OrtsID"> </pre>	
<pre> <listRelation type="genealogical"> <relation name="" active="Biographierte r" passive="SFZ"> </pre>	

Der Abschnitt <stettingsDesc>

<pre> <listPlace> <place@xml:id> </pre>	<p>Die Liste der genannten Orte</p> <p>- die ID ist intern und bezieht sich auf im Aufbau befindliche Ortsdatenbank</p>
<pre> <location><geo> </pre>	<p>Koordinaten</p>
<pre> <placeName> </pre>	<p>Ortsname, ggf. mit normalisierter Fassung unter @n wenn bekannt</p>
<pre> <idno type="gnd osm odb"> </pre>	<p>ID-Nummer die auf Normdaten/Datenbanken referenziert</p> <p>GND</p> <p>OpenStreetmap (kein persitenter Identifier)</p>

Der Abschnitt `<div type="entry" subtype="adb|ndb">`

... hat weitere Attribute @n und @xml:id

<div type="prev"> <head>	Vorgänger im Band, dessen Name und @sfz
<div type="prev"> <head>	Nachfolger im Band, dessen Name und @sfz
<div type="citation">	Zitiernamen dieses Biographierten
<div type="kopf"> <p> <seg type="notname"> <birth@n> <date@type,@n> <placeName@type> <death@n> <date@type,@n> <placeName@type> <placeName type="tomb"> <faith> </seg>	<ul style="list-style-type: none"> – Kopfzeile der Biographie, in einem <p> der Namen enthält und ggf. ein <seg> für den Rest (Beruf, Geburt, Tod, Konfession) = <i>alle Daten sind bereits in den Metadaten kodiert</i> – Geburt und Tod mit @n als normalisierte Datumsangabe, tw. @calendar=Julian (nicht interpretiert) – Ortsnamen wenn identifiziert @key = interne OrtsID, @resp = Zuordnungskürzel, @corresp = lat,lon Angabe, @osmname = wenn bekannt –
<div type="geneal">	
<div type="leben">	
<div type="A W L P Qu">	
	– Verweise auf Familienmitglieder – nur in sog. Familienartikeln möglich

Elemente unterhalb von <p>

<seg>	bisher nur im <div@kopf>, sonstungenutzt
<choice><abbr type="acronym"><expan> <choice><abbr type="lemma"><expan>	Abkürzung, die als Akronym nicht durch <expan> ersetzt werden sollte Das Lemma ist abgekürzt und i.d.R. vereinfacht die Kurzform das Lesen.
<choice><abbr><expan>	Abkürzung, die durch <expan> erklärt werden sollte, die Ersetzung ist nur in den Bänden NDB 1-12 realisiert, in den übrigen Bänden sind die <expan> automatisch eingesetzt und tw. nicht eindeutig [Psychologe, Psychologie]
<choice><orig><reg>	Trennungen von Worten an <pb> und <milestone> werden durch <reg> zusammengesetzt angegeben
<name type="(person place org job)">	
<ref target="">	Link auf <ul style="list-style-type: none"> – Personenartikel – Literaturtitel – Projekte wie GW, dMGH
<hi rend="(italics antiqua sup sub)">	kann überall unterhalb von <p> kann vorkommen:
<lb/> <milestone /> <pb />	Zeilenumbruch Spaltenwecheln Seitenwechsel