

Reihenfolge von Dokumenten und Subdossier in Dossier (RFC 2021-18)

1 Ausgangslage

Werden eCH-0160 SIPs verwendet um gescannte Objekte, z.B. Dossiers und Dokumente zu *ingestieren*, wird schnell klar, dass die gescannten Objekte, d.i. Dokumente und Subdossier in einem Dossier eine Reihenfolge im dem Scanner vorgelegten Dossier haben. Die Reihenfolge ist durch die Reihenfolge der Unterlagen im Mäppchen vorgegeben und soll erhalten bleiben.

Bis anhin wurde in Digitalen Dossiers in eCH-0160 dieses Problem ignoriert, bzw. auf das Entstehungsdatum von Dossier und Dokument oder auf das Aktenzeichen als ordnendes Element abgestellt. Beides, Entstehungsdatum oder Aktenzeichen ist beim Scannen nicht direkt verfügbar und bildet auch die Reihenfolge des analogen Dossiers nicht in jedem Fall adäquat ab.

2 Problemstellung

Die Reihenfolge in welcher Dokumente und Subdossiers in metadata.xml, d.i. XML Textdatei vorkommen ist nicht ausschlaggebend für die logische Reihenfolge dieser Objekte im XML DOM (Document Object Model)¹ und kann von jedem Programm, z.B. Package Handler bei der Darstellung geändert werden. In metadata.xml hingegen müssen gemäss Vorgabe XSD auf derselben Hierarchie-Ebene immer zuerst alle Subdossiers und dann alle Dokumente eingetragen werden, ansonsten ist das metadata.xml nicht valide. Auf diese Weise können Reihenfolgen, bei denen sich z. B. Subdossiers und Dokumente abwechseln, nicht abgebildet werden.

3 Lösungsvorschlag

Eine einfache Lösung ist jedem Objekt (Dossier oder Dokument) ein spezifisches Zusatzdatum mit zu geben, dass dann von der Darstellungssoftware (Package Handler, DIP Creator etc.) für die Darstellung in korrekte Reihenfolge ausgewertet wird. Das ist leider nicht wirklich nachhaltig, weil Zusatzdaten nur für menschenlesbare Daten gedacht sind und darum für die Zusatzdaten auch kein spezifisches kontrolliertes Vokabular definiert ist. Sinnvoller wäre es hier eine Lösung zu implementieren, die sich an den Vorschlag für RFC 2021-16: Abbildung von Versionen und Repräsentationen anlehnt, weil wir es hier mit einem ähnlichen Problem, aber auf höherer Stufe zu tun haben.

Erweiterter Lösungsvorschlag:

Jedes Element in der Ordnungssystem Navigation kann ein Attribut *order* erhalten, so wie das schon bei den Elementen *Vorgang* und *Aktivität* der Fall ist:

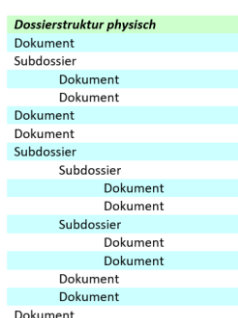
```
<titel>Geschäft eröffnen</titel>  
  <aktivitaet order="1">  
    <vorschreibung>KRG 1 KR-Geschäft anmelden</vorschreibung>
```

¹ https://de.wikipedia.org/wiki/Document_Object_Model

Konkret sind das die Elemente *Ordnungssystemposition*, *Dossier*, *Dokument*, *Datei*, *Vorgang* und *Aktivität*.

Bei Datei ersetzt *order* das ursprünglich im RFC 2021-16 vorgesehene Attribut *sequence/reihung*, was ja logisch das gleiche ist. Grund dafür ist, dass das Attribut *order* für die Elemente *Aktivität* und *Vorgang* bereits vorhanden ist.² Damit wird die Konsistenz bei der Benennung des Attributs bei allen Elementen gewahrt. Damit ist es möglich die physische Strukturierung eines Papierdossiers eins zu eins abzubilden. Das *order* Attribut kann entweder eine Zahl oder ein Datum sein, die Sortierfolge ist die natürliche Sortierordnung. Das Attribut ist optional.

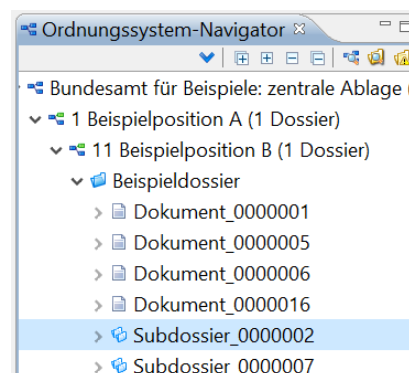
Beispiel:
Ein Dossier ist wie folgt strukturiert



Die Vergabe der Nummerierung, bzw. des Sortierwertes kann fortlaufend auf einer Ebene erfolgen und auf jeder darunterliegenden Ebene neu beginnen.

Im Beispiel sähe das also so aus:
(Die Abbildung zeigt jeweils die heutige Anordnung im Package Handler)

Dokument_00000001	order="1"
Subdossier_00000002	order="2"
Dokument_00000005	order="3"
Dokument_00000006	order="4"
Subdossier_00000007	order="5"
Dokument_00000016	order="6"



Es ist aber auch denkbar, dass bedingt durch den Scanprozess die Nummerierung über Dossier und Subdossier fortlaufend stattfindet.

Dokument_0000001	order="1"
Subdossier_0000002	order="2"
Dokument_0000003	order="3"
Dokument_0000004	order="4"
Dokument_0000005	order="5"
Dokument_0000006	order="6"
Subdossier_0000007	order="7"
...	
Dokument_0000016	order="16"

² Vgl. dazu [STAN d DEF 2015-06-24 eCH-0160 V1.1 ArchivischeAblieferungsschnittstelle DataDictionary.xlsx](#), eCH-0160 Archivische Ablieferungsschnittstelle (SIP), v1.1

In beiden Fällen ist in der jeweiligen hierarchischen Ebene eine klare Reihenfolge gegeben. Ergibt sich der Fall, dass zwei Elemente (Dossier, Dokumente) in einer Ebene den gleichen Wert für *order* haben, bedeutet das, dass zwischen diesen Elementen keine Reihenfolge festgelegt ist.

Beliebige Informationen können aber auch weiterhin in den Zusatzdaten von Dossier und Dokument gespeichert werden. Für eine Interpretation oder Darstellung der Anordnung von Dossiers und Dokumenten in den Tools ist das aber ohne Belang.

4 Umsetzung in XML

Wie oben dargelegt kann neu ein Attribut *order* den Elemente *Ordnungssystemposition*, *Dossier*, *Dokument*, *Datei*, *Vorgang* und *Aktivität* zugeordnet werden.

```
<dokument id="_s04MZwdEeugYO7zUkufmw" order="1">
  <titel>Dokument_0000010</titel>

<dossier id="_HaqngJwdEeugYO7zUkufmw" order="2">
  <titel>Subdossier_0000002</titel>
```

Das Attribut ist vom Typ xs:integer oder xs:date.

Das Attribut ist optional.

Die Schema Einschränkung, dass in einem Dossier erst Subdossiers und erst dann Dokumente aufgelistet werden können, soll zugunsten einer freien Mischbarkeit von Dossier und Dokumenten aufgegeben werden. Damit ist es dann auch möglich, dass der Aufbau der XML Datei *metadata.xml* parallel zum Scan-Prozess stattfinden kann. Die damit im XML abgebildete Reihenfolge der Subdossiers / Dokumente erleichtert lediglich die Erstellung des *metadata.xml*. Für die Dokumentation der Reihenfolge der Elemente ist das Attribut *order* zu verwenden, da die Reihenfolge im *metadata.xml* nachwievor nicht ausschlaggebend für die logische Reihenfolge dieser Objekte ist.

↑ ordnungssystemposition	
= id	_jcFb4JwcEeugYO7zUkufmw
⟨⟩ nummer	11
⟨⟩ titel	Beispielposition B
↑ dossier	
= id	_nNShEJwcEeugYO7zUkufmw
⟨⟩ titel	Beispieldossier
⟨⟩ erscheinungsform	digital
▼ entstehungszeitraum	
▼ dossier (2)	
▼ dokument (4)	