

Das Klassifikationen-Datenmodell

2015.05 2016.06 2017.06
2016-03-18

by Robert Stephan, Jens Kupferschmidt

*In diesem Abschnitt wird die Struktur einer Klassifikation
und ihre Repräsentation im XML-Format beschrieben.*

Table of contents

1 Das Klassifikationen-Datenmodell.....	2
1.1 Elemente einer Klassifikation.....	2
1.2 Besonderheiten.....	3
1.3 XML Format.....	4
1.4 Für Entwickler.....	5

1 Das Klassifikationen-Datenmodell

Für das Importieren und Exportieren von Klassifikationen wird in MyCoRe ein **XML-Format** verwendet.

1.1 Elemente einer Klassifikation

Eine Klassifikation besteht aus:

1.1.1 Rotelement

Das Rotelement (`<mycoreclass>`) enthält die ID der Klassifikation, ein oder mehrere Bezeichner `label` in verschiedenen Sprachen und einen Container `categories`, der die Einträge der Klassifikation speichert.

Die ID kann maximal 32 Zeichen lang sein und darf die Zeichen `:`, `$`, `{`, `}` nicht enthalten.

1.1.2 Kategorie

Als Kategorien (`<category>`) bezeichnen man die Einträge der Klassifikation. Kategorien können hierarchisch geschachtelt werden. Sie enthalten eine ID, einen oder mehrere Bezeichner `label` für verschiedene Sprachen und optional eine URL `url`. Diese kann verwendet werden, um auf eine Webseite mit weiteren Informationen oder der Beschreibung des Klassifikationseintrages zu verlinken.

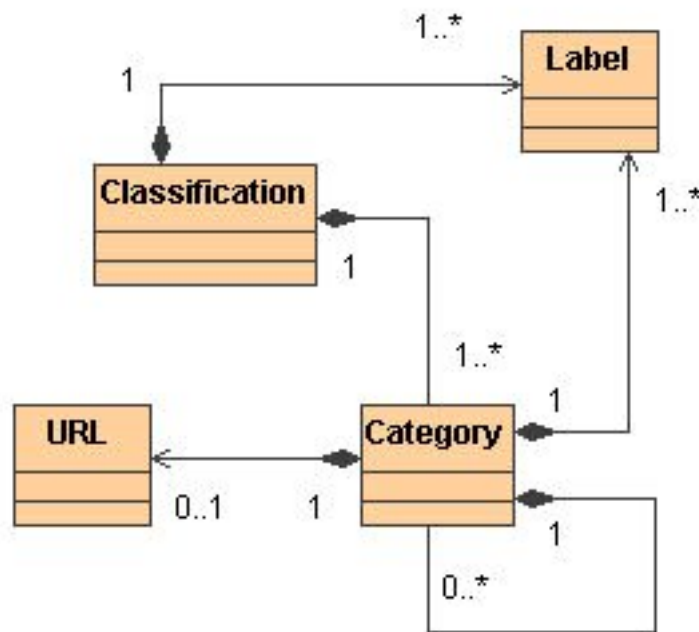
Die ID kann maximal 128 Zeichen lang sein und darf die Zeichen `:`, `$`, `{`, `}` nicht enthalten.

1.1.3 Bezeichner

Bezeichner (`<label>`) werden verwendet, um die Klassifikation und ihre Einträge zu benennen. Label sind wiederholbar, um auch Übersetzungen des Begriffes speichern zu können. Sie bestehen aus einem Sprachattribut `xml:lang` und einem Textfeld `text`. Optional können zusätzliche Angaben zur Verwendung des Eintrages in einem Beschreibungsfeld `description` abgelegt werden.

Die Länge von Text und Beschreibung kann jeweils maximal 4096 Zeichen betragen. Das Sprachattribut kann maximal 255 Zeichen lang sein und muss gemäß den Regeln des [RFC 4646](https://www.ietf.org/rfc/rfc4646.txt) (<https://www.ietf.org/rfc/rfc4646.txt>) aufgebaut sein.

Folgende Grafik soll die Beziehung zwischen den Klassifikationselementen verdeutlichen:



1.2 Besonderheiten

1.2.1 x- Sprachcodes

In MyCoRe wird das Label zur Speicherung zusätzlicher Informationen für die Weiterverarbeitung der Klassifikationen verwendet. Dazu werden zusätzliche Label mit privaten Sprachcodes (gemäß den Regeln des RFC4646) definiert. Die Sprachcodes bestehen aus dem Prefix `x-` und weiteren Blöcken aus jeweils max. 8 Zeichen (Zahlen oder Buchstaben), die durch `-` getrennt werden.

Folgende Sprachcodes werden verwendet:

- **x-uri** für Bezeichner in Form einer URI.
URIs sind mit URLs (HTTP-Adressen) verwandt und werden vor allem im Context von Linked Data eingesetzt.
Im [MODS](http://www.loc.gov/standards/mods/) (<http://www.loc.gov/standards/mods/>) -Metadatenmodell werden sie zur Speicherung von Klassifikationen in Form von `authorityURI` und `valueURI` Attributen verwendet. Anhand des `x-uri`-Labels kann MyCoRe diese Klassifikationen erkennen und verarbeiten.
- **x-auth** als Marker für Klassifikationen, die im [MODS](http://www.loc.gov/standards/mods/) (<http://www.loc.gov/standards/mods/>) -Metadatenmodell mit dem Attribut `authority` gekennzeichnet wurden.
- **x-mapping** für das [Klassifikationsmapping](http://www.loc.gov/standards/mods/..../documentation/basics/classifications/classification_mapping.html) ([../documentation/basics/classifications/classification_mapping.html](http://www.loc.gov/standards/mods/..../documentation/basics/classifications/classification_mapping.html)) .

Wenn Metadaten z.B. über die OAI-Schnittstelle an externe Services ausgeliefert werden, werden häufige andere Klassifikationseinträge benötigt. Um Mehrfacheingaben zu vermeiden, bietet MyCoRe mit dem Klassifikationsmapping die Möglichkeit, dass nur eine primäre Klassifikation erfasst wird. MyCoRe trägt dann die äquivalenten Einträge der anderen Klassifikation automatisch dazu. Die Werte werden als Leerzeichen-getrennte Liste von Klassifikationseinträgen der Form classid:catid gespeichert.

1.2.2 Format von IDs

In hierarchischen Klassifikationen kann es sinnvoll sein, die Hierarchiestufen beispielsweise durch einen Punkt (.) oder Numerierung in den IDs abzubilden:

- Institutionen: Unis # Unis.Rostock # Unis.Rostock.UB
- DDC: 5 # 56 # 567

Für das Funktionieren des Klassifikationssystems ist das jedoch keine Voraussetzung.

1.3 XML Format

Das folgende XML-Grundgerüst kann für das Schreiben eigener Klassifikationen verwendet werden:

```
<?xml version="1.0" encoding="UTF-8" ?>
<mycoreclass
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="MCRClassification.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  ID="..." >
  <label xml:lang="de" text="..." description="..."/>
  <label xml:lang="en" text="..." description="..."/>
  <label xml:lang="x-uri" text="..." description="..."/>
  ...
  <categories>
    <category ID="...">
      <label xml:lang="..." text="..." description="..."/>
      ...
      <url xlink:href="..." />
      <category ID="...">
        <label xml:lang="..." text="..." description="..."/>
        ...
      </category>
    </category ID="...">
      <label xml:lang="..." text="..." description="..."/>
      ...
    </category>
  </categories>
</mycoreclass>
```

Für die Validierung kann die XML-Schema Datei [MCRClassification.xsd](http://server.mycore.de/svn/mycore/trunk/mycore-base/src/main/resources/MCRClassification.xsd) (<http://server.mycore.de/svn/mycore/trunk/mycore-base/src/main/resources/MCRClassification.xsd>) aus dem SVN verwendet werden.

1.4 Für Entwickler

Intern erfolgt die Speicherung der Klassifikationen in mehreren verknüpften *Datenbanktabellen*:

- In der Tabelle *mrcategory* werden Klassifikationseinträge und ihre hierarchische Struktur gespeichert. Die Hierarchie wird als *Nested Set Model* (s. [Wikipedia](https://en.wikipedia.org/wiki/Nested_set_model) (https://en.wikipedia.org/wiki/Nested_set_model)) abgebildet.
- Die Tabelle *mrcategorylabels* enthält die mehrsprachigen Bezeichner und Beschreibungen der Klassifikationseinträge
- Das Mapping von MyCoRe-Objekten zu Klassifikationseinträgen wird in der Tabelle *mrcategorylink* gespeichert