

# Vereinheitlichung internationaler Bibliothekskataloge

Christian Scheel, Claudia Schmitz, and  
Ernesto William De Luca

Georg-Eckert-Institut — Leibniz-Institut für internationale Schulbuchforschung,  
Celler Straße 3, 38114 Braunschweig, Deutschland  
{scheel,schmitz,deluca}@gei.de  
<http://www.gei.de>

**Zusammenfassung** Aus dem Bereich der Schulbuchforschung beschreiben wir exemplarisch, welche Herausforderungen, Chancen und Stolpersteine es gab, ein Rechercheinstrument einer deutschen Schulbuchsammlung zu einem internationalem Rechercheinstrument auszubauen. Hierbei standen vor allem das Erstellen und Erweitern einer gemeinsamen Repräsentationsbasis, sowie das Wiederverwenden existierender Lösungen im Mittelpunkt. Diese Ausarbeitung hilft bei der zielorientierten Planung ähnlich ausgerichteter Vorhaben.

**Keywords:** Cross-Lingual IR, Cross-Cultural IR, Digital Humanities

## 1 Einleitung

Der Gegenstand der Schulbuchforschung ist die Untersuchung von gesellschaftlich (politisch), pädagogisch und fachwissenschaftlich sanktioniertem Wissen für die Bildung der jungen Generation. Zum Beispiel gewinnen Schulbücher auf der Suche nach „populärem Wissen“ für die historische Forschung zunehmend an Bedeutung, da sich in ihnen Weltanschauungen, Denkströmungen und erwünschte Wissensbestände widerspiegeln. Am Georg-Eckert-Institut<sup>1</sup> wird die Forschung deshalb vor allem von den Erziehungswissenschaften, der Geschichtswissenschaft, der Geographie, der Politikwissenschaft und der Religionswissenschaft bestimmt [5,6,4]. Die Vorteile für einen internationalen Blickwinkel liegen klar auf der Hand, weshalb die Idee entstand ein institutionelles Rechercheinstrument für Schulbücher zu einem internationalen Rechercheinstrument auszubauen.

Ein Rechercheinstrument speziell für Schulbücher ist deshalb notwendig, da Bibliotheken Schulbücher im Allgemeinen nicht forschungsadäquat als Schulbücher nachweisen, sondern schlicht als Bücher. Dem Schulbuchforscher ist es deshalb nicht möglich, explizit nach Schulbüchern zu recherchieren. Hinzu kommt, dass Schulbücher spezielle Eigenschaften wie „Geltungsland“, „Unterrichtsfach“, „Bildungslevel“, „Schulform“, etc. aufweisen, die ein Rechercheinstrument berücksichtigen muss.

<sup>1</sup> <http://www.gei.de/das-institut.html>

Erweitert man ein monolinguales Rechercheinstrument mit Katalogen aus internationalen Beständen, hat man es zuallererst mit zwei Problemen zu tun. Zum einen möchte der internationale Wissenschaftler das Rechercheinstrument in seiner Sprache benutzen, zum anderen nutzen die Bibliothekare natürlich ihre Muttersprache und gegebenenfalls eigene Codes um die Eigenschaften der Bücher zu beschreiben. Man muss also nicht nur jedes Attribut und jede Ausprägung dieser Attribute in jeder Sprache beschreiben können, sondern auch die Beschreibungen in den Katalogen auf eben diese Attribute und deren Ausprägungen abbilden (mappen) können.

Die Überführung vom institutionellem Rechercheinstrument TextbookCat<sup>2</sup> zum internationalen Rechercheinstrument International TextbookCat dient neben dem Aufbau einer erweiterbaren Architektur unter anderem dazu, Erfahrungen mit Workflows und Arbeitsschritten zu sammeln sowie etwaige Herausforderungen und notwendige Ressourcen besser einschätzen zu können. Um die Chancen und Fallstricke eines solchen Vorhabens offen zu legen, konzentrierte sich der International TextbookCat auf die Zusammenführung verschiedensprachiger Schulbuchdatenbanken aus drei Institutionen: dem Georg-Eckert-Institut, der Università degli Studi di Torino und der Universidad Nacional de Educación a Distancia. Die Zusammenarbeit verfolgte die Festlegung einheitlicher Standards zur Erfassung von Schulbuchdaten und die Anfertigung von Mappings auf ein einheitliches Datenformat.

Das entstandene Rechercheinstrument kann stellvertretend für jedes spezialisierte Rechercheinstrument stehen, bei dem geplant ist, die Datenbasis durch internationale Kataloge zu einem Metakatalog zu erweitern.

## 2 Die Ausgangslage

Um die Ausgangslage vor der Internationalisierung erfassen zu können, werden im Folgenden die institutionelle Schulbuchsammlung und internationale Schulbuchsammlungen beschrieben. Es folgt eine Beschreibung des Rechercheinstruments und eine Betrachtung der Vorteile eines internationalen Rechercheinstruments.

### 2.1 Die institutionelle Schulbuchsammlung

Die Schulbuchsammlung der Forschungsbibliothek<sup>3</sup> umfasst rund 175.000 Schulbücher, welche auch vor Ort hinterlegt sind. Um diese Schulbücher besser beschreiben zu können erarbeiteten die Bibliothekare ein Klassifikationssystem mit welchem jedes Buch der Sammlung detailliert beschrieben werden konnte (siehe Tabelle 1). Eine Besonderheit beim Geltungsland ist der teilweise

---

<sup>2</sup> <http://vufind.gei.de/vufind2/>

<sup>3</sup> <http://www.gei.de/bibliothek>

Gebrauch von Jahreszahlen zur genauen Spezifikation <sup>4</sup>. Die Klasse „Bundesland“ beinhaltet dabei nur Codes für die deutschen Bundesländer. Aufgrund der schulwissenschaftlichen Ausrichtung des Georg-Eckert-Instituts finden sich im Bereich „Unterrichtsfach“ nur die Fächer Geschichte, Sozialkunde, Geographie und Religion, sowie muttersprachlicher Unterricht. Mathematikschulbücher sind zum Beispiel nicht Teil der Sammlung, da man mit ihnen die Fragestellungen der Schulbuchwissenschaft nicht beleuchten kann. Die Klasse „Bildungsgang“ entspricht fast ausschließlich der Klassifikation der International Standard Classification of Education (ISCED) der UNESCO. Die „Schulform“ orientiert sich am deutschen Schulsystem, auch wenn bei circa 100.000 Schulbüchern des Bestandes das „Geltungsland“ nicht Deutschland ist. Die Klasse „Zeitraum“ ist redundant zum Publikationsjahr ermöglicht jedoch eine genaue Klassifikation zu bestimmten historisch relevanten Epochen. Die schulbuchspezifische Klasse „Publikationsform“ unterscheidet zum Beispiel zwischen Schulbuch, Lehrplan, Lehrmittel, Lehrerhandbuch, Aufgabensammlung, etc. Abbildung 1 zeigt die Abdeckung der annotierten Attribute des Klassifikationssystems in der institutionellen Schulbuchsammlung. Da ein großer Teil der Sammlung aus internationalen Schulbüchern besteht, haben die international gültigen Attribute erwartungsgemäß eine höhere Abdeckung.

**Tabelle 1.** Lokale Notation, wobei \_ den Platzhalter für Ziffern oder Buchstaben darstellt.

Code(s)	Klasse	Ausprägungen
l_ _ _	(Geltungs-) Land	181
b_ _ / b_bz	Bundesland / Besatzungszone Deutschland	16 / 4
u_ _ _	Unterrichtsfach/Lernbereich	15
k_ _	Klassenstufe/Bildungsgang	7
s_ _	Schulform Deutschland	11
z_ _ _	Zeitraum	15
d_ _	Publikationsform – Inhaltsform (Dokumenttyp)	12

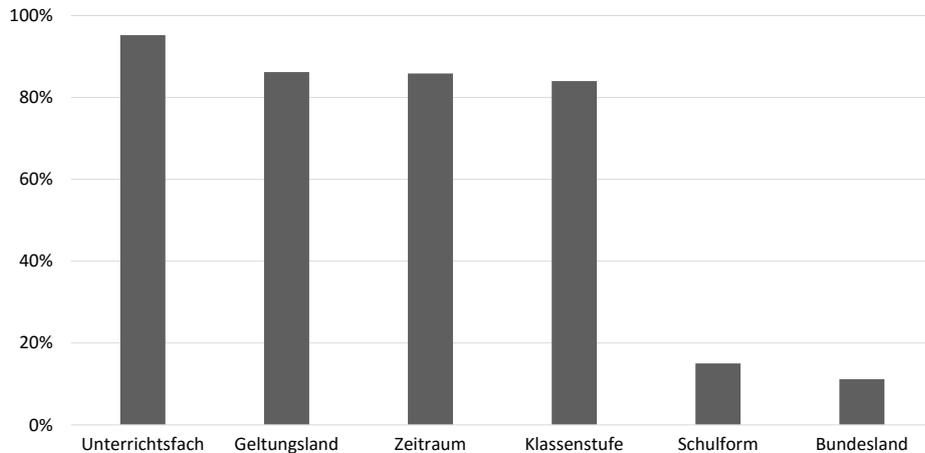
## 2.2 Internationale Schulbuchsammlungen

Internationale Sammlungen sind meist aus der Notwendigkeit heraus entstanden, bekannte Schulbücher geordnet zu erfassen, weil es keine explizite Behandlung von Schulbüchern in Bibliotheken gab. Anders als am Georg-Eckert-Institut, bei dem alle Schulbücher der Sammlung Teil der Forschungsbibliothek sind, stellen die Datenbanken internationaler Schulbuchforscher meist Bibliografien dar. Dies ist kein Nachteil, denn andererseits wären Informationen über Schulbücher, die

<sup>4</sup> Zum Beispiel:

l025: Jugoslawien (-1992)

l125: Jugoslawien, Föderative Republik (1992-2003)



**Abbildung 1.** Abdeckung der annotierten Attribute des Klassifikationssystems in der institutionellen Schulbuchsammlung.

in Klosterbibliotheken hinterlegt sind schlicht unbekannt. Zusätzlich ist die Qualität der Informationen gerade bei schulbuchspezifischen Eigenschaften als hoch einzuschätzen und ähnlich genug, dass sie sich in ein einheitliches Klassifikationssystem abbilden lassen. Beispielhaft werden im Folgendem drei prominente Sammlungen beschrieben. Weitere Sammlungen können [10] entnommen werden.

**EDISCO** Das Research Center for digitization and creation of digital libraries for the humanities der Universität Turin hostet die Datenbank EDISCO, die mit etwa 30.000 Datensätzen italienische Schulbücher aus dem 19. und 20. Jahrhundert nachweist.

**MANES** Das Department of History of Education and Comparative Education hostet die Datenbank MANES, die mit etwa 35.000 Datensätzen spanische, portugiesische und lateinamerikanische Schulbücher aus dem 19. und 20. Jahrhundert nachweist. Die Metadaten in MANES werden nach abgestimmten und festgeschriebenen Katalogisierungsrichtlinien erfasst [9].

**Emmanuelle Textbook Project** Alain Choppin war ein Pionier der Schulbuchforschung. Unter dem Titel „Emmanuelle Textbook Project“ erschuf er 1979 eine Sammlung französischer Schulbücher ab dem Jahr 1789 [2,3].

### 2.3 Der TextbookCat

Der TextbookCat ist das Rechercheinstrument des Georg-Eckert-Institut — Leibniz-Institut für internationale Schulbuchforschung. Neben der Recherche

nach Schulbüchern, die bestimmte Attribute aufweisen, wird der *TextbookCat* von Schulbuchforschern genutzt, um Forschungsfragen zu finden, zu beantworten und zu beleuchten.

Im *TextbookCat* spiegelt sich das Klassifikationssystem der Forschungsbibliothek als Facetten wider. Durch die Auswahl von Ausprägungen gewünschter Attribute über die Facetten können Benutzer bei der Recherche die Ergebnismenge manipulieren, sodass sie ihrer Forschung dienlich ist [1].

Eine Analyse der Log-Dateien des Rechercheinstruments zeigte, dass der Großteil der Benutzer die Facetten für Recherchen nutzte. Drei Viertel aller Rechercheanfragen basierten ausschließlich auf der Nutzung der Facetten. Dementsprechend wurde bei einem Viertel mit der Hinzunahme von Suchbegriffen, also der klassischen Textsuche, recherchiert.

Bei der Migration des *TextbookCat* sollten nicht nur die Datenbestände internationalisiert werden, sondern auch die Qualitätseigenschaften des Rechercheinstruments, wie die Abdeckung der schulbuchspezifischen Attribute (vgl. Abbildung 1), bestehen bleiben. So sollte verhindert werden, dass die Schulbuchsammlung des Georg-Eckert-Instituts in den Ergebnismengen überrepräsentiert wird.

## 2.4 Chancen des International TextbookCat

Die Erweiterung des institutionellen Rechercheinstruments *TextbookCat* zum internationalen Rechercheinstrument *International TextbookCat* erfolgt vor allem im Backend der Architektur. Bevor Sammlungen in sprachunabhängige Suchindexe aufgenommen werden können, müssen sie in eine einheitliche Datenstruktur überführt werden (vgl. [8]).

Die bekannten Schulbuchsammlungen nach und nach in einem Rechercheinstrument zu vereinen, bringt mehrere Vorteile.

1. Man nähert sich immer mehr einer generellen Repräsentationsstruktur von Schulbüchern und somit einem Standard. Wenn die Materie am Anfang zu groß erscheint, benötigt es einen iterativen Ansatz, um sich mit jedem weiteren Datensatz einem Standard zu nähern.
2. An einem Ort in verschiedenen Sammlungen zu recherchieren erspart den Aufwand für das Finden der individuellen Rechercheoberflächen der Sammlungen<sup>5</sup>.
3. Ein einheitliches Klassifikationssystem, welches Eigenschaften und Ausprägungen mit Codes beschreibt, unterstützt Multilingualität ideal, da die Codes in die jeweiligen Sprachen abgebildet werden können.
4. Auch wenn man nie weiß, wie viele Schulbücher man nicht im System hat, liefert ein vereinigttes Rechercheinstrument mit der Internationalität eine neue Dimension in der Schulbuchforschung.

---

<sup>5</sup> Beispielfhaft kann der Leser versuchen das Rechercheinstrument vom Emmanuelle Textbook Project zu finden.

### **3 Beobachtungen**

Im Idealfall hätten die Sammlungen der internationalen Partner eins zu eins auf die Datenstruktur des Georg-Eckert-Instituts gemappt werden können. Dass dies nicht der Fall sein konnte, war schon vor dem Vorhaben klar, da andere Länder andere Schulsysteme haben, die im erarbeiteten Klassifikationssystem keine Abbildung finden. Weitere Gründe wurden bei der Betrachtung der internationalen Sammlungen offenbart.

#### **3.1 Individuelle Felder in den Sammlungen**

Da die Erzeuger der Sammlungen keine Bibliothekare waren, wurden zum Teil nur Eigenschaften erfasst, die für Schulbücher im Speziellen, aber nicht Bücher im Allgemeinen wichtig sind. Beispielhaft trennt ein Bibliothekar Titel, Untertitel und Titelzusatz, wobei ein „Laie“ diese Informationen als ein Ganzes aufnehmen würde, weil die Trennung im Schulbuchkontext unwichtig erscheint.

#### **3.2 Interpretation individueller Felder**

Eine weitere Herausforderung stellten Felder dar, deren Bezeichnung nicht eindeutig waren, sodass sie von verschiedenen Personen unterschiedlich interpretiert wurden. Beispielhaft wurde das Feld „Verwendung“ bei MANES als „Verwendung für“ und als „Verwendung als“ interpretiert, was dazu führte, dass dort zum einen Personen und Gruppen, sowie Dokumenttypen vorzufinden waren.

#### **3.3 Unterschiede im Detaillierungsgrad**

Bei der Betrachtung der Sammlungen zeigte sich der Nachteil von Feldern, die mit nicht vordefinierten Einträgen beschrieben werden konnten. Das äußerte sich zum Beispiel in 155 Schulfächern bei EDISCO und 85 Schulfächern bei MANES die den 15 Schulfächern des Klassifikationssystems gegenüber standen. Auch wenn ein großer Teil der Varianten Schreibfehlern geschuldet war, blieben nach der Bereinigung 86 neue Kategorien übrig, die berücksichtigt werden mussten. Dies führte zu grundsätzlichen Entscheidungen, ob man Eigenschaften generalisieren (weniger Optionen), spezialisieren (alle Optionen) oder nicht abbildbare Eigenschaften einfach ignorieren sollte. Das Generalisieren oder Ignorieren würde zum Verlust an Informationen führen, während das Spezialisieren den Nachteil birgt, dass man jede Ausprägung (auch zukünftige) in jede unterstützte Sprache (eindeutig) übersetzen muss.

#### **3.4 Einträge lassen sich kaum mappen**

Aufgrund unterschiedlicher Bildungssysteme sind Felder wie die „Schulform“ nicht aufeinander abbildbar. Im Rechercheinstrument würde ein Filtern auf ein solches Feld meist nur zu Ergebnissen aus der speziellen Sammlung führen, was nicht gewollt war. Zudem wurden internationale Schulbücher in der Sammlung des Georg-Eckert-Instituts nicht hinsichtlich der „Schulform“ katalogisiert.

### 3.5 Unterschiedliches Verständnis der Materie

Der eigentliche Vorteil, dass man reine Schulbuchsammlungen zusammenführt, offenbart die Frage, ab wann ein Buch ein Schulbuch ist. Dass Mathematikbücher Schulbücher sind, auch wenn sie kein Bestandteil traditioneller Schulbuchforschung sind, ist klar ersichtlich. Anders verhält es sich bei Curricula oder wissenschaftlichen Abhandlungen zur Schulbuchforschung, bei denen die Nutzer des Rechercheinstruments entscheiden müssten, ob diese Bücher recherchierbar sein sollten. Sind Kochbücher, die in EDISCO zu finden sind, Schulbücher?

### 3.6 Generalisierung

Zusammenfassend lassen sich die Beobachtungen wie folgt generalisieren:

1. Datensatzfelder tragen oft Informationen für mehrere Felder.
2. Freitextfelder begünstigen Schreibfehler und mehrere Schreibvarianten für eine Ausprägung.
3. Datensatzfelder sind oft in der Muttersprache formuliert.
4. Ein bestehendes Klassifikationssystem kann zu speziell, aber auch zu generell sein.

Der Idealfall sähe wie folgt aus:

1. Datensatzfelder finden ihre genauen Entsprechungen in bibliographischen Datenformaten wie PICA<sup>6</sup> oder MARC21<sup>7</sup>.
2. Belegungen der Datensatzfelder werden über ein festes Vokabular bestimmt.
3. Belegungen der Datensatzfelder werden durch Codes beschrieben, die sich nach internationalen Standards richten.
4. Existierende Klassifikationssysteme lassen sich auf Codes internationaler Standards abbilden.

## 4 Vereinheitlichtes Klassifikationssystem

Das Klassifikationssystem beschreibt Klassen und deren Ausprägungen als Codes. Es soll garantieren, dass die beschriebenen Ausprägungen der relevanten Eigenschaften der Sammlungen im vereinigten Rechercheinstrument nutzbar sind. Auch wenn die Belegungen der Datensatzfelder nicht als Codes vorliegen, so lassen sie sich dennoch in solche überführen [7].

---

<sup>6</sup> <https://www.gbv.de/de/katricht/inhalt.shtml>

<sup>7</sup> <https://www.loc.gov/marc/>

## 4.1 Erstellen des Klassifikationssystems

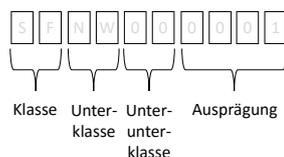
Der erste Schritt im Erstellen des Klassifikationssystems kann nur, wie in Abschnitt 3 beschrieben, das Untersuchen der Sammlungen sein. Idealerweise kann man den zu betrachtenden Datensatz in eine csv-Datei exportieren und ein Tabellenkalkulationsprogramm oder OpenRefine<sup>8</sup> nutzen, um sich einen Überblick zu verschaffen. Erfahrungsgemäß folgt als zweiter Schritt das Aufräumen der Daten, das zwar der Sammlung dient, aber für das Abbilden ins Klassifikationssystem keine Rolle spielt, da auch Schreibfehler und Schreibvarianten auf Codes abgebildet werden können.

Nachdem jede relevante Klasse bestimmt, beziehungsweise im Datensatz identifiziert wurde, müssen alle Ausprägungen notiert werden. Jede Ausprägung, die nicht schon Teil der Klasse ist, wird mit einem neuen Code ins Klassifikationssystem aufgenommen. Dieser Schritt ist nicht trivial, da entsprechende Sprach- und Fachkenntnisse vorausgesetzt werden, um zu erkennen, ob zwei Ausprägungen in zwei Sammlungen die gleiche Eigenschaft darstellen.

In der Praxis sollte man diesen Schritt aufgrund des Aufwands infrage stellen und schauen, ob es nicht ratsamer ist, eine Klasse durch existierende Standards zu beschreiben (siehe Abschnitt 4.2).

Ein intensives Beschäftigen mit einer Klasse führt in einigen Fällen zur Notwendigkeit, Klassen mit Unterklassen erweitern zu müssen, sodass das Klassifikationssystem im Prinzip baumartig repräsentiert werden müsste. Stattdessen sollten Ober- und Unterklassen durch die Codestructur repräsentiert werden (siehe Abbildung 2), bei der Teile des Codes auf die Klasse und Unterklasse(n) einer Ausprägung schließen lassen.

Ein einmal festgelegtes Klassifikationssystem sollte möglichst nicht um neue Klassen erweitert werden, da die bestehenden Sammlungen höchstwahrscheinlich nicht in diese Klasse abgebildet werden können. Anders verhält es sich bei den Ausprägungen bestehender Klassen, denn vorher unbekannte Eigenschaftsausprägungen sollten mit neuen Codes in das Klassifikationssystem aufgenommen werden. Ein Neustrukturieren mit Unterklassen ist möglich, wenn sich die bisher vergebenen Codes diesen Unterklassen zuweisen lassen, aber man muss aufpassen, da sich die Codes aufgrund der Unterklasse gegebenenfalls ändern.



**Abbildung 2.** Klassen, Unterklassen und potentielle weitere Unterklassen lassen sich direkt im Code abbilden, was eine lineare Repräsentation im Klassifikationssystem ermöglicht. Beispielhaft könnte der Code `SFNW000001` für „Schulfach > Naturwissenschaften > Mathematik“ stehen.

<sup>8</sup> <http://openrefine.org>

## 4.2 Klassifikation durch Standards

Es wird dringend dazu geraten Standards zu nutzen oder nach existierenden Klassifikationen Ausschau zu halten, da in deren Erstellung Expertisen und vor allem Zeit geflossen sind. Standards ermöglichen es, die Erstellung des eigenen Klassifikationssystems entscheidend zu beschleunigen.

Der Standard für Länder ist die ISO-3166-1, für subnationale Einheiten (wie Bundesländer) die ISO-3166-2, für Sprachen die ISO-639-2. Für anwendungsspezifische Klassen sollte man eine handvoll Ausprägungen als Suchbegriffe verwenden, um mithilfe einer Suchmaschine festzustellen, ob jemand bereits eine Klassifikation erstellt hat.

Ob im endgültigen System dann alle durch einen Standard definierten Ausprägungen genutzt werden ist zweitrangig, da das Rechercheinstrument nur genutzte Ausprägungen in Facetten anzeigen wird. Zeitgleich sollten keine nicht vergebenen Ausprägungen weggelassen werden, weil unklar ist, welche Datensätze mit welchen Eigenschaften zu einem späteren Zeitpunkt repräsentiert werden müssen.

Es ist ratsam, einen gefundenen Standard nicht mit dessen existierenden Codes zu beschreiben, sondern eigene Codes zu vergeben, die auf diesen Standard abbilden (vgl. Abbildung 2). Der Grund hierfür ist die Erweiterbarkeit, falls ein Standard nicht ganz genau auf die gewünschte Materie abbildet.

Klassifikationssystem	Übersetzungen		
	<u>Sprache X</u>	<u>Sprache Y</u>	<u>Sprache Z</u>
<Klasse A>	<Übersetzung XA>	<Übersetzung YA>	...
• <Ausprägung A1> → <Code A1>	• <Übersetzung XA1>	• <Übersetzung YA1>	• ...
• <Ausprägung A2> → <Code A2>	• <Übersetzung XA2>	• <Übersetzung YA2>	• ...
• <Ausprägung A3> → <Code A3>	• <Übersetzung XA3>	• <Übersetzung YA3>	• ...
...	...	...	...
• <Ausprägung An> → <Code An>	• <Übersetzung XAn>	• <Übersetzung YAn>	• ...
<Klasse B>	<Übersetzung XB>	<Übersetzung YB>	...
<Unterklasse BA> → <Code BA>	<Übersetzung XBA>	<Übersetzung YBA>	...
• <Ausprägung B1> → <Code B1>	• <Übersetzung XB1>	• <Übersetzung YB1>	• ...
• <Ausprägung B2> → <Code B2>	• <Übersetzung XB2>	• <Übersetzung YB2>	• ...
• <Ausprägung B3> → <Code B3>	• <Übersetzung XB3>	• <Übersetzung YB3>	• ...
...	...	...	...
<Unterklasse BZ> → <Code BZ>	<Übersetzung XBZ>	<Übersetzung YBZ>	• ...
• <Ausprägung Bm> → <Code Bm>	• <Übersetzung XBm>	• <Übersetzung YBm>	...
<Klasse C>	<Übersetzung XC>	<Übersetzung YC>	• ...
• <Ausprägung C1> → <Code C1>	• <Übersetzung XC1>	• <Übersetzung YC1>	...
...	...	...	...

**Abbildung 3.** Struktur des Klassifikationssystem im Zusammenspiel mit Übersetzungen

### 4.3 Übersetzungen

Das fertige Klassifikationssystem besteht aus Klassen und Codes. Diese müssen für jede Sprache in entsprechende Begriffe abbildbar sein, was eindeutige Übersetzungen benötigt. Hat man für die Klassifikation auf ISO-Normen zurück gegriffen, sollten entsprechende Übersetzungen verfügbar sein.

Für die Sprachen der Sammlungen, für die die Codes vergeben wurden stehen diese Begriffe ebenfalls zur Verfügung. Für andere Sprachen müssen die Übersetzungen mit hohem Aufwand zusammengetragen werden.

Da das Rechercheinstrument intern nur mit den Codes arbeitet und Übersetzungen nur in der Weboberfläche dargestellt werden, ist eine (vorläufige) Übersetzung durch automatisierte Systeme<sup>9</sup> besser als keine Übersetzung. Jede Übersetzung kann jederzeit überarbeitet und so Mehrdeutigkeit eliminiert werden.

## 5 Mapping

Nach dem Festlegen des Klassifikationssystems müssen alle relevanten Ausprägungen der Sammlungen auf deren Entsprechungen im Klassifikationssystem abgebildet werden, bevor sie in einem Suchindex zusammengefasst werden können.

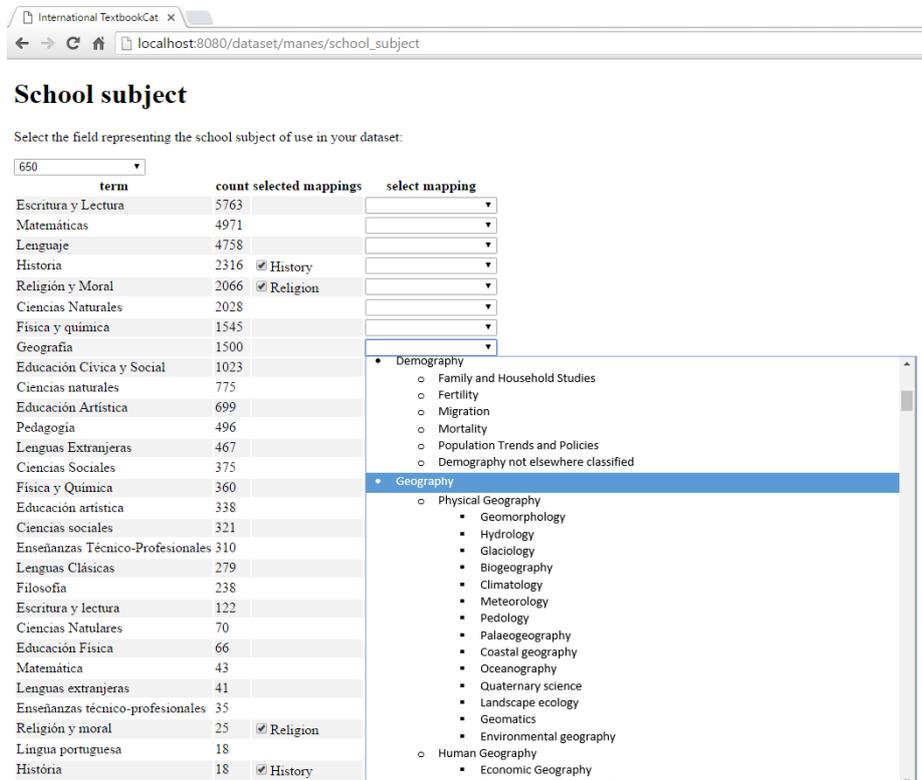
### 5.1 Dialoggestütztes Mapping

Da die Mapping-Regeln für jede neue Sammlung individuell erstellt und für den Indexierer in digitaler Form vorliegen müssen, haben wir ein System entwickelt, das den Prozess des Integrierens, über das Mapping, Indexieren und Testen unterstützt. Das Erzeugen der Mapping-Regeln wird dabei durch eine Weboberfläche unterstützt und kann jederzeit überarbeitet werden.

Um diesen Prozess zu unterstützen verlangen wir, dass jede Partnerdatenbank ihre Daten über eine standardisierte Schnittstelle (Z39.50, OAI-PMH) anbieten muss. Dies wird üblicherweise durch die Verwendung eines aktuellen Bibliothekssystems gewährleistet. Während der Indexierer über diese Schnittstelle die aktuellen Dokumentinformationen gewinnt und auf ihnen die Mapping-Regeln anwendet, werden für das Erstellen der Mapping-Regeln erst einmal nur alle Attribute und deren Ausprägungen gesammelt und nach Häufigkeit sortiert.

Der Dialog führt durch die domain-spezifischen Klassen (die im Rechercheinstrument als Facetten repräsentiert werden) und fragt, welches Feld einer bestimmten Klasse entspricht. Wie in Abbildung 4 zu sehen ist, werden die ermittelten Ausprägungen dieses Feldes nach der Häufigkeit sortiert präsentiert, um das Mapping festlegen zu können. Ziel ist, dass jedes Attribut eine Entsprechung im Klassifikationssystem findet. Mehrere Ausprägungen dürfen dabei auf das selbe Feld abgebildet werden. Eine Ausprägung kann aber auch auf mehrere Felder im Klassifikationssystem abgebildet werden.

<sup>9</sup> wie Google Translate <https://translate.google.com>



**Abbildung 4.** Dialoggestütztes Erstellen der Mapping-Regeln von einer Sammlung in das Klassifikationssystem.

## 5.2 Vom Mapping zum Suchindex

Mit den gegebenen Mapping-Regeln kann ein Indexierer ohne Umwege über eine weitere Repräsentation der Sammlungen einen homogenen und in Teilen sprachunabhängigen Suchindex aufbauen.

## 6 Zusammenfassung und Ausblick

Bei der Überführung vom institutionellen Rechercheinstrument *TextbookCat* zum internationalen Rechercheinstrument *International TextbookCat* musste eine allgemeingültige Repräsentationsstruktur gefunden werden, die internationale Schulbuchsammlungen nahezu vollständig abbilden kann. Der Schulbuchforschung wurde so ein Instrument zur Hand gegeben, welches die internationale Dimension in Schulbüchern besser beleuchten kann.

Wir haben gezeigt, wie ein Klassifikationssystem aussehen und welche Schritte man bei der Erstellung dieser Repräsentationsstruktur befolgen sollte. Länger-

fristig kann ein solches Klassifikationssystem als Standard verwendet werden, wenn es sich bewähren sollte.

Des Weiteren haben wir gezeigt, wie ein Dialogsystem eingesetzt werden kann, um komplette Sammlungen im erstellten Klassifikationssystem abzubilden.

Durch die Nutzung von Klassifizierungs-Codes als Beschreibung der Eigenschaften der Bücher ist das Rechercheinstrument sprachunabhängig, da diese Codes erst im Web-Browser in Begriffe der jeweiligen Sprache umgewandelt werden. Zur kompletten Multilingualität fehlt jedoch eine entsprechende Behandlung der textuellen Suchanfragen und der Volltexte.

Sobald Daten gänzlich durch Codes beschrieben werden können, ist das Verknüpfen mit Ontologien trivial. In diesem Zustand können dann semantische Verfahren aufsetzen, um einen noch größeren Mehrwert für die Schulbuchforschung zu generieren.

## Literatur

1. Calhoun, K., Cellentani, D., et al.: Online Catalogs: What Users and Librarians Want: an OCLC report (2009), <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>
2. Choppin, A.: EMMANUELLE: a data base for textbooks' history in Europe. *Historical Social Research* 14(4), 52–58 (1989), <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-51666>
3. Choppin, A.: The Emmanuelle Textbook Project. *Journal of Curriculum Studies* 24(4), 345–356 (1992), <http://dx.doi.org/10.1080/0022027920240404>
4. Fiedler, M., Scheel, C., Weiß, A., De Luca, E.: Welt der Kinder. Semantisches Information Retrieval als Zugang zu Wissensbeständen des 19. Jahrhunderts. In: FVWG und GCDH Workshop für Wissenschaftsgeschichte und Digital Humanities in Forschung und Lehre (2016)
5. Fuchs, E., Kahlert, J., Sandfuchs, U.: Schulbuch konkret: Kontexte - Produktion - Unterricht. Klinkhardt (2010)
6. Fuchs, E., Niehaus, I., Stoletzki, A.: Das Schulbuch in der Forschung: Analysen und Empfehlungen für die Bildungspraxis. Eckert. Expertise, V&R Unipress (2014), <http://www.gei.de/publikationen/eckert-expertise/>
7. de Groat, G.: Future Directions in Metadata Remediation for Metadata Aggregators. Tech. rep., Digital Library Federation (2009)
8. Mayr, P., Petras, V.: Cross-concordances: terminology mapping and its effectiveness for information retrieval. CoRR abs/0806.3765 (2008), [http://archive.ifla.org/IV/ifla74/papers/129-Mayr\\_Petras-en.pdf](http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf)
9. Ossenbach, G.: Research about school handbooks in Latin America: The project MANES contribution. *Historia de la Educación* 19(0) (2013), <http://revistas.usal.es/index.php/0212-0267/article/view/10797>
10. Ossenbach, G.: Textbook databases and their contribution to international research on the history of school culture. *History of Education & Children's Literature* IX(1), 163–174 (2014)