

Rudolf Ungváry und Péter Király

# Bemerkungen zu der Qualitätsbewertung von MARC-21-Datensätzen

## 1 Einführung

Király (2019a) beschreibt eine neuartige Methodologie der systematischen Analyse der Bewertung der Qualität von MARC-21-Datensätzen. In seiner Forschungsarbeit untersucht er die Gesamtkataloge von 16 wichtigen National- und Forschungsbibliotheken. Die Ergebnisse beleuchten einige typische Probleme und liefern hilfreiche Messgrößen für das Verständnis von qualitätsentscheidenden Strukturmerkmalen eines Katalogs. In dieser weiterführenden Studie wollen wir eine tiefergehende Analyse einer kleineren Menge von Katalogen durchführen.

Dies sind unsere Forschungsfragen:

- (a) Was sind jeweils die Hauptmerkmale der Katalogisierungspraxis in jeder der Bibliotheken?
- In welchem Umfang werden die MARC-21-Felder und -Unterfelder genutzt (z. B.: Wird das Feld 648 *Zeitschlagwort*<sup>1</sup> verwendet?)?
  - Wie vollständig werden Unterfelder verwendet (z. B.: Wie oft kommen die Unterfelder \$0 *Identifikations- oder Standardnummer des Datensatzes* oder \$1 *Persistenter Link zum Datensatz*<sup>2</sup> in einer Titelaufnahme vor?)?
  - Besteht Stimmigkeit zwischen verwandten Unterfeldern (z. B. Wie verhalten sich die Unterfelder \$a *Benennung* und \$2 *Quelle* zueinander?)?
  - In welchem Umfang werden die für den lokalen Gebrauch reservierten 9XX-Felder und die \$9-Unterfelder verwendet?
  - Wie oft kommen im MARC-21-Standard nicht enthaltene Felder und Unterfelder vor?

---

<sup>1</sup> Die Übersetzung der Feldbezeichnungen richtet sich nach der Feldbeschreibung der Titeldaten der Deutschen Nationalbibliothek und der Zeitschriftendatenbank im Format MARC 21, Version 3.3 mit Stand 9. November 2020, <https://d-nb.info/1220864196/34> (22.12.2020), sowie nach dem MARC21 – Handbuch in deutscher Sprache der Schweizerischen Nationalbibliothek, <https://ead.nb.admin.ch/web/marc21/dmarcein12.htm> (22.12.2020). Dort, wo es keine deutsche Entsprechung gab, wurde eigens übersetzt.

<sup>2</sup> Im Engl. *real world object URI*.

(b) Welche Schlussfolgerungen können im Hinblick auf das verwendete Katalogisierungssystem (z. B. inwieweit die Eingabe fehlerhafter Werte toleriert wird) sowie auf die bei der Sacherschließung verwendeten kontrollierten Vokabulare (Schlagwörter, Thesauri, Klassifikationen, Taxonomien) gezogen werden?

(c) Last but not least: Können die Kataloge der einzelnen Bibliotheken im Hinblick auf die jeweilige Katalogisierungspraxis miteinander verglichen werden (z. B. im Hinblick darauf, in welchen Fällen geografische Koordinaten in Feld 034 *Kodierte kartografische mathematische Angaben*; bzw. in Feld 653 *Indexierungsterme – nicht normiert* für Deskriptoren, die nicht einem kontrollierten Vokabular entnommen sind, eingegeben oder nicht eingegeben werden)?

Da die Katalogdatensätze keine direkte Information über alle Details des Katalogisierungsprozesses enthalten (eingesetzte Software, lokale Katalogisierungsregeln und -gewohnheiten sowie Änderungen derselben), können wir nur begrenzte Schlussfolgerungen ziehen. Tatsächlich können nur die Spezialist:innen der jeweiligen Bibliothek stimmige Schlussfolgerungen ziehen, weil es Fälle geben kann, in denen nur sie die Hintergründe für ein bestimmtes Merkmal oder einen Fehler kennen. Die daraus gewonnenen Schlüsse sind allerdings innerhalb der Bibliothekswelt geläufig.

Gegenstand dieser Studie sind nur die inhaltsbezogenen Datenelemente (043 *Code für geografische Gebiete*, 044 *Ländercode der veröffentlichenden/herstellenden Stelle*, 045 *Zeitabschnitt der Eintragung*, 052 *Geografischer Klassifikationscode*, 072 *Schlagwortkategoriecode*, 080 *Notation nach der Universal Decimal Classification*, 082 *Notation nach der Dewey Decimal Classification*, 084 *Andere Notation*, 085, 6XX-Feldergruppe für *Schlagworteintragungen*) sowie einige andere damit verwandte Datenelemente (008 *Datenelemente mit fester Länge zur physischen Beschreibung*, 034 *Kodierte kartografische mathematische Angaben*, 041 *Sprachcode*). Verglichen wurden die Katalogdaten der Deutschen Nationalbibliothek (DNB), der Universitätsbibliothek Gent (GENT), der Bibliothek der Ungarischen Akademie der Wissenschaften (MTAK), der Universitätsbibliothek Szeged (SZTE) und des Ungarischen Verbunds (MOKKA). Die Daten sind entweder offen zugänglich,<sup>3</sup> oder sie wurden uns für Forschungszwecke zur Verfügung gestellt. Gelegentlich beziehen wir auch die Széchényi-National-

---

<sup>3</sup> Wir pflegen eine Liste von Open-Access-Katalogen auf Github: <https://github.com/pkiraly/metadata-qa-marc> (22.12.2020).

bibliothek (die Ungarische Nationalbibliothek, NSZL) und die Finnische Nationalbibliothek in die Betrachtung mit ein.

Bei der Nennung von MARC-21-Datenelementen wenden wir die folgende Konvention an: 010\$a *Kontrollnummer der Library of Congress*, wobei 010 die Feldnummer (*tag*) ist, a der Code für das Unterfeld und *Kontrollnummer der Library of Congress* die Benennung des Datenelements. Bei Kontrollfeldern schreiben wir 008/18–21 *Illustrationen*, um ein Datenelement in Feld 008 zwischen der 18. und 21. Zeichenposition zu bezeichnen.

## 1.1 Methodologie

Die Analyse wurde mit dem Forschungssoftwarepaket QA Catalogue durchgeführt. Seine Code-Funktionen wurden in Király (2019a) beschrieben, die Programmoberfläche in Király (2020). Es handelt sich um Open-Source-Software, die über Github<sup>4</sup> zugänglich ist. Das Programm deckt alle Kontrollfelder und Datenfelder sowie deren Indikatoren und Unterfelder ab (also alle Datentypen und Werte). Die Anzahl der in den einzelnen Feldern und Unterfeldern vorhandenen Datenelemente wird berechnet und ihr Verhältnis mit dem aller Titelaufnahmen des Katalogs verglichen. Die Nutzer:innen des Programms können Listen der in den Datenelementen vorhandenen Werte mitsamt der Angabe, in welchen Katalogdatensätzen sie vorkommen, ausgeben.

650 — Subject Added Entry - Topical Term			
\$1		≡	9 0.00%
\$2	Source of heading or term	≡	733,857 57.82%
\$3	Materials specified	≡	1,922 0.15%
\$a	Topical term or geographic name entry element	≡	733,809 57.81%
\$b	Topical term following geographic name entry element	≡	26 0.00%

**Abb. 1:** Einige Unterfelder von Feld 650 *Schlagworteintragung – Sachschlagwort*: Wenn die blauen Codes für die Unterfelder in der ersten Spalte *location* angeklickt werden, wird eine Liste aller Titelaufnahmen angezeigt, in denen das jeweilige Unterfeld vorkommt. In der zweiten Spalte *label* stehen die Benennungen der Felder und Unterfelder. In der Spalte *terms* ste-

<sup>4</sup> Github: <https://github.com/pkiraly/metadata-qa-marc> (22.12.2020).

hen Links zu den Listen der Werte (Zahlen, Code-Ausdrücke,<sup>5</sup> Ausdrücke natürlicher Sprache und Texte), die in den Unterfeldern erscheinen – wobei jeder Wert mit den Aufnahmen verlinkt ist, in denen er vorkommt. Die Datensätze können sowohl in lesefreundlichem Format als auch als MARC-21-Aufnahme angezeigt werden. Die Spalte *count* zeigt an, in wie vielen Titelaufnahmen das Unterfeld vorkommt, während die letzte Spalte den prozentualen Anteil der Titelaufnahmen an der Gesamtzahl der Titelaufnahmen im jeweiligen Katalog anzeigt.

Für die Kontrollfelder (00X-Felder) werden auch die Werte der einzelnen Positionen – bei denen es sich theoretisch nur um feststehende (einem kontrollierten Vokabular entnommene) Ausdrücke handeln kann – aufgeführt (Abb. 2).

path	message url	instances	records
<b>control subfield level issues</b>		3,656,937	363,434
control subfield: invalid code (158 variants)		4,807	4,556
[+]			
008/18-21 (tag008book18)	'r' in 'ar' <a href="#">i</a>	3,392	3,392
008/18-21 (tag008book18)	'r' in 'arb' <a href="#">i</a>	292	292
008/18-21 (tag008book18)	'r' in 'r' <a href="#">i</a>	45	45
008/18-21 (tag008book18)	'r' in 'arf' <a href="#">i</a>	28	28

**Abb. 2:** Fehleranzeigedetails für die 00X-Felder. Wird das blaue [+] angeklickt, so werden alle Positionen mit Problemen angezeigt. In der Spalte message stehen die eigentlichen Probleme. Das blaue i ist ein Link zu der entsprechenden Seite des MARC-21-Standards. Die letzten beiden Spalten zeigen die Häufigkeiten und die Anzahl der betroffenen Aufnahmen an.

## 2 Vergleich von Fehlern in Kontrollfeldern und Unterfeldern im vorgeschriebenen Format

Die Analyse hebt auch Fehler in Kontrollfeldern (00X) hervor. Es gibt nur wenige Zeichenwerte in diesen Feldern, die einem kontrollierten Vokabular entnommen sind und sie müssen in bestimmten Zeichenpositionen des 00X-Felds stehen. Fehler in vorgeschriebenen Unterfeldern mit numerischem Format (z. B. ISBN) und in Unterfeldern zur Verknüpfung (*linkage*) mit vorgeschriebenem Format (z. B. \$6) werden ebenfalls angezeigt.

Das Feld 008 *Datenelemente mit fester Länge zur physischen Beschreibung* enthält eine Folge von 40 Zeichen, die in positiongebundene Datenelemente un-

<sup>5</sup> Code-Ausdrücke (in der bibliothekarischen Praxis in der Regel Notationen) sind die in symbolischer Sprache geschriebenen Elemente kontrollierter Vokabulare, während der Code selbst das Sprachsystem des Vokabulars ist.

terteilt ist. Fehler in diesem Feld sind besonders schwerwiegend. Die Werte in diesem Feld sind mit der inhaltlichen Suche eng verbunden.<sup>6</sup> Die positionengebundenen Werte mit fester Länge (wenige Zeichen lang) bezeichnen Typ und Genre des Dokuments, was in Verbindung mit den Deskriptoren der 6XX-Feldergruppe (aber im Prinzip auch in Verbindung mit den Werten jedes anderen Felds) zur Suchverfeinerung nützlich sein kann. In anderen Worten: Der Dokumententyp ist bei der Einengung einer Suche ein sehr wichtiger Aspekt. Im Vergleich zur Bedeutsamkeit seiner Datenelemente bekommt das Feld 008 nicht die Aufmerksamkeit, die es eigentlich verdient, obwohl die Codes für die Formate, die in jeder Position eingetragen werden können, früh festgelegt wurden (in den 1970ern) und aus der Warte der Erschließungstheorie nicht als wirklich kohärentes System angesehen werden können.

Obwohl die positionsgebundenen Formatcodierungen manchmal lokal modifiziert wurden,<sup>7</sup> ist die Typologie in MARC 21 selbst dieselbe geblieben und sie wurde nie grundsätzlich reorganisiert, um die Codierungen theoretisch stimmiger zu machen. Das Hauptproblem generell ist aber, dass Bibliothekssysteme den Nutzer:innen keine Möglichkeit anbieten, sie für die Suche zu nutzen.

In den 00X-Feldern sind der numerische Wert (z. B. Datum) sowie die Form und Zeichen für die Datenelemente streng geregelt. Die positionsgebundenen Zeichen werden kontrollierten Vokabularen entnommen, die Teil von MARC 21 sind. Im Verhältnis zu den betreffenden Regeln sind die Anzahl und die Art der Fehler auffällig hoch. Auch ist das nicht nur bei einzelnen Bibliotheken so, sondern fällt generell auf. Wir nehmen an, dass das nicht in erster Linie mit Sorgfalt und Katalogisierungsqualität, sondern eher mit einer unterwickelten Fehlererkennung der jeweiligen Katalogisierungssoftware zu tun hat. Selbst eine sehr sorgfältige intellektuelle Eingabe dieser nicht natürlichsprachigen Werte kann Fehler nicht vermeiden.

---

<sup>6</sup> Dies scheint wenig bekannt zu sein. In der Spezifikation von MOKKA, dem gemeinsamen ungarischen Verbundkatalog, werden beispielsweise nur 6XX-Felder als inhaltsbezogene Felder angesehen. Aber die Frage des Dokumententyps lässt sich von der Suche nach Themen und Notationen nicht trennen, da sie selten nur für einen Dokumententyp allein gesucht werden. Die Suche mit Hilfe eines Dokumententyp ist jedoch überwiegend mit Schlagwörtern anderen Typs verbunden. Man sucht sehr selten nur mit einem Form- bzw. Genreschlagwort.

<sup>7</sup> Die Wichtigkeit der Modifikationen wird in einer einzigen ungarischen Quelle, bei Bilicsi (2018), erwähnt – in Verbindung damit, dass die Typologie für den internen Gebrauch verbessert wurde: „This can be observed in more structured fields (especially 007, 008) and in the expanded possibilities of physical description of new media; media can be better put into the categories that are most relevant of them, their main characteristics can be specified more precisely.“ Diese Quelle macht auch Angaben zu anderen Änderungen bei MARC-21-Feldern, -Unterkategorien, und Indikatoren, die hauptsächlich zum Zweck der Anpassung an RDA vorgenommen wurden.

Tab. 1: Positions- und Unterfeldfehler

MARC-21-Daten- element		DNB	GENT	MTAK	SZTEK	MOKKA
Probleme in Kontrollfeldern						
ungültige Werte <sup>I</sup>	Anzahl	11 238	864 505	281 454	3 656 937	20 317 130
	Sätze	3 889	333 635	110 011	363 434	5 240 827
Probleme in anderen Feldern						
mehrfache Verwen- dung von Feldern, die nur einmal er- laubt sind	Anzahl	740	67	8	240	4 705
	Sätze	740	65	8	240	4 698
undefinierte Felder <sup>II</sup>	Anzahl	–	–	308 347	2 884	1 115 714
	Sätze	–	–	268 696	2 626	479 623
Indikatorprobleme	Anzahl	53 161	5 066	192 690	5 015	3 541 444
	Sätze	8 066	1 303	178 461	4 271	2 298 715
Unterfeldprobleme <sup>III</sup>	Anzahl	2 368 900	529 103	87	31 197	34 759 334
	Sätze	2 483 725	430 833	57	26 900	8 864 772
– undefinierte Unterfelder <sup>IV</sup>	Anzahl	2 478 770	520 863	520 863	11 417	33 465 038
	Sätze	2 365 446	429 314	429 314	8 048	8 861 086
– 65X	Anzahl	728	7	13	75 <sup>V</sup>	682
	Sätze	768	7	10	60	624
– mehrfache Ver- wendung von Unter- feldern, die nur ein- mal erlaubt sind	Anzahl	–	–	6 452	1 900	881 787
	Sätze	–	–	3 867	1 857	776 565

I Meistens im Feld 008 im Datenelement *Art des Inhalts* in Position 24–27.

II Meistens lokal definierte Felder, deren Benennung 9 enthält.

III Falsche oder beschädigte ISBN-/ISSN-Werte, mehrfache Verwendungen von Unterfeldern, die nur einmal erlaubt sind, undefinierte Unterfelder, ungültige Werte.

IV Inklusive lokal definierter \$9-Unterfelder von unterschiedlichen Feldern.

V Alle im Feld 653 *Indexierungsterm* – nicht normiert!

In Zeile 5, *mehrfache Verwendung von Feldern, die nur einmal erlaubt sind*, fällt besonders auf, dass es sich formal um eine sehr einfache Art von Fehler handelt.

Zu Zeile 6: Ein eher kleiner Teil der undefinierten Felder und Unterfelder mag für lokale Zwecke eingesetzt worden sein, aber in den meisten Fällen zeigen die Werte eine so hohe Inkonsistenz (Rauschen), dass wir meinen, dass das

Vorhandensein dieser Unterfelder im Katalog wohl eher mit Tippfehlern als mit speziell entwickelten lokalen Definitionen zu tun hat.

Im Fall der beiden ungarischen wissenschaftlichen Bibliotheken können HUNMARC-Felder und -Unterfelder eine Rolle spielen, die von QA Catalogue nicht erkannt werden. Es ist auch möglich, dass solche Fehler beim Import von Titelaufnahmen externer Bibliotheken entstanden sind.

Die Werte dieser Datenelemente sowie die Felder, Unterfelder und die vorgeschriebenen Ein-Zeichen-Indikatoren sind so strukturiert, dass ihre formale Überprüfung relativ einfach ist. Insofern wäre es leicht, ihre Kontrolle in das Katalogisierungssystem zu integrieren. Es ist wahrlich keine schwierige Programmieraufgabe, die Eingabe ungültiger Datenelementwerte, die mehrfache Verwendung von Feldern, die nur einmal erlaubt sind, die Verwendung undefinierter Felder, Unterfelder, Indikatoren usw. zu verhindern oder solche Fehler nachträglich zu erkennen. Selbst bei Bibliotheken mit einem sehr guten Ruf kommen solche Fehler manchmal millionenfach vor. (Beispielsweise wird in der DNB mit PICA katalogisiert, die MARC-21-Aufnahmen werden durch einen Konversionsprozess und für Datenaustauschzwecke erzeugt, so dass die fast zweieinhalb Millionen undefinierten Unterfelder auch auf Konversionsfehler zurückgehen können).

Die kompletten bibliografischen Daten wurden mit einer Filterung in den Verbundkatalog MOKKA migriert. Es ist auffällig, wie viele Fehler trotzdem transferiert wurden. Wer das Problem kennt, weiß, dass das Administrieren von Filterprogrammen und lokalen Feldern sowie deren Dokumentation normalerweise nicht perfekt läuft und die Dokumentation schwer zugänglich ist. Das mag auch bei MOKKA der Fall sein: Nur die Filterspezifikation der Széchényi-Nationalbibliothek ist verfügbar.<sup>8</sup> Die Spezifikationsdokumente werden auf den MOKKA-Webseiten bereitgestellt, aber keine davon enthält die Konversionsspezifikation für die bibliografischen Daten der einzelnen Lieferbibliotheken.<sup>9</sup>

Es wäre hilfreich, wenn es in Katalogisierungssystemen bei der Definition lokaler Felder und Unterfelder die Möglichkeit gäbe, diese nicht nur zu benennen, weil die Benennungen wegen ihrer Kürze nicht informativ genug sind. Es sollten auch längere Erklärungen und Definitionen angefügt werden können, die sich im System speichern und als Online-Hilfe nutzen ließen.

---

<sup>8</sup> Szabó (2003) hat die Fehler der NSZL analysiert. Diese Analyse ist auf den MOKKA-Seiten nicht zu finden (<http://www.mokka.hu/web/guest/katalogizalasi-szabalyok>, 23.12.2020), sondern nur die Katalogisierungsregeln und die von Szabó bereitgestellte Anleitung.

<sup>9</sup> Natürlich heißt das nicht, dass es keine Dokumente, die die Kriterien und Anforderungen für jedes Filterprogramm enthalten, gibt. Eher muss vermutet werden, dass es sich um fehlende oder unveröffentlichte Manuskripte handelt.

Es folgen einige beispielhafte Fehler aus dem SZTE-Katalog. Ähnliche kommen in anderen Katalogen vor.

**Tab. 2:** Einige Fehlertypen aus dem Bestand der SZTE

Fehlertyp	Fehlerbeispiel	Sätze	Anzahl
Fehler in Kontrollfeld 008 (720 Varianten)		3 656 937	363 434
Code-Probleme (158 Varianten)		4 807	4 556
008/18–21	Im Datenelement kommt undefiniertes r vor.	3 392	3 392
Wertprobleme (562 Varianten)		3 652 130	359 259
008/00–05	Wert 2002 widerspricht dem Datenformat YYMMDD.	13	13
008/29	Wert [köz] ist in dieser Position nicht erlaubt.	225 388	225 388
008/06	Wert x ist in dieser Position nicht erlaubt.	1	1
Indikatorprobleme		333	332
65\$ind1	Der erste Indikator sollte leer sein, enthält aber 1.	33	33
Unterfeldprobleme		31 197	26 900
080\$6	. , und ) sind in diesem Unterfeld ungültige Zeichen. Die \$6-Verknüpfung ist ein besonderes Unterfeld, um Repräsentationen verschiedener Schriftsysteme für dieselbe Information zu verbinden. Hier sollte ein bestimmtes Format eingehalten werden, um Schriftsysteme entsprechend einem kontrollierten Vokabular laut MARC 21 Appendix A* zu bezeichnen (z. B. Na = kyrillische Schrift). Punkt, Komma und schließende Klammer sind hier nicht erlaubt.	8	8
Subfield: ISBN-Probleme (10 348 Varianten)		13 115	13 012
020\$a	Wert 0238-3349(fúzött) ist keine ISBN.	8	8

\* <https://www.loc.gov/marc/bibliographic/ecbcdntf.html> (23.12.2020).

Kurz gesagt: Fehler in Datenelementen, die numerische Werte enthalten, welche vorwiegend in vorgeschriebenem Format oder als einem einfachen, kontrol-



lierten Vokabular entnommene Zeichenwerte anzugeben sind, könnten vermieden werden, wenn die Katalogisierungssysteme selbst es erlaubten, Eingaben systematisch zu prüfen. Das setzt natürlich voraus, dass Katalogisierer:innen auch in der Lage sind, ihren Fehlererkennungsbedarf akkurat zu kommunizieren, wenn solche Systeme ausgewählt werden. Weiterhin wäre es gut, wenn man sie in der Ausbildung darauf auch vorbereiten würde.

### 3 01X–09X Feldergruppe Nummern und Codes

Das Verhältnis der Anzahl der Werte in Feld 040 *Katalogisierungsquelle* zur Anzahl anderer Identifier und Feldbelegungen in Titelaufnahmen gibt einen Eindruck von der Vollständigkeit der Katalogisierung. Darüber hinaus enthält die Feldergruppe Sprach- und Länderkennungen (*codes*), die bei der inhaltlichen Suche eine Rolle spielen. Die Geokoordinaten in Feld 034 *Kodierte kartografische mathematische Angaben* sind für geografische Quellen wichtig, weil sie es ermöglichen, bei Suchen nach geografischen Namen sie auf Landkarten anzuzeigen.<sup>10</sup> Schließlich enthält die Feldergruppe sieben Felder für die inhaltliche Suche, sie sind für Notationen ausgewählter Klassifikationssysteme gedacht. Die meisten Felder dieser Gruppe können also bei der vergleichenden Analyse berücksichtigt werden.

#### 3.1 Identifier im bibliografischen Datensatz

Man kann sich denken, dass eine Nationalbibliothek (wie die DNB) kaum einmal einen Datensatz von einer anderen großen Bibliothek erhält, der das Feld 010\$a *Kontrollnummer der Library of Congress* (LoC) enthält. Daher sind die 0,03% DNB-Sätze mit LoC-Kontrollnummer eher Ausnahmen. Mehr davon gibt es offenbar in GENT (14,56%). Also muss vielleicht erklärt werden, warum es wissenschaftliche Bibliotheken mit einer vernachlässigbaren Menge an LoC-Sätzen (oder damit identischen Sätzen) gibt, obwohl man vermuten kann, dass eine größere Menge davon übernommen wurde.

---

<sup>10</sup> Darüber hinaus sind die Geokoordinaten einzelner Siedlungen, bewohnter Orte, geomorphologischer Formationen (Berge, Hügel, Gewässer usw.) in Ortslexika und -verzeichnissen (*gazetteers*) erfasst (z. B. in *Geonames* oder in Publikationen nationaler Statistikbehörden). Das Herunterladen der Geokoordinaten von in Katalogen erfassten Orten sollte also ohne besondere Schwierigkeiten möglich sein.

Tab. 3: Datensatz-IDs in Feldergruppe 01X–09X

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
010\$a Kontrollnummer der Library of Congress	5 894	0,03	256 940	14,56	1	0,00	161	0,01	56 668	0,64
015 Nummer der Nationalbibliografie										
\$2 Quelle	11 345 731	59,63	26 219	1,18	–	–	4	0,00	598	0,01
\$a Nummer der Nationalbibliografie	11 345 731	59,63	71 517	1,49	–	–	7	0,00	6 291	0,07
\$z Gelöschte / Ungültige Nummer der Nationalbibliografie	1 995 499	10,49	637	0,04	–	–	–	–	12	0,00
016 Kontrollnummer der nationalbibliografischen Agentur										
\$2 Quelle	19 Mio.	100	31 802	1,80	–	–	5	0,00	843	0,01
\$a Datensatzkontrollnummer	19 Mio.	100	32 155	1,82	–	–	6	0,00	847	0,01
019\$a lokal definiertes Datenelement	–	–	2	0,00	–	–	–	–	–	–

Das Feld 016 *Kontrollnummer der nationalbibliografischen Agentur* ist für die Deutsche Nationalbibliothek natürlich vollständig. Das lässt sich auch daran ablesen, dass seine Häufigkeit (19 Millionen) mit der Häufigkeit in Feld 040 *Katalogisierungsquelle* übereinstimmt (wie in Tab. 5 gezeigt) – was übrigens im Vergleich zu anderen Bibliotheken ein Beleg für einen sehr sorgfältigen Katalogisierungsprozess ist. Ein weiteres Indiz dafür ist die hohe Anzahl gelöschter Nummern.

In anderen Bibliotheken kommt das Feld 016 mit nationalbibliografischen Identifiern kaum und das Feld 019 mit lokalen Identifiern fast gar nicht vor. Die Situation ähnelt der beim Feld 015 *Nummer der Nationalbibliografie*. Im Fall der DNB sind 59,63 % der Katalogdatensätze in der Nationalbibliografie enthalten. Im Fall der Ungarischen Nationalbibliothek (die in dieser Untersuchung nicht berücksichtigt wird), beträgt die Erfassungsrates der Bücher ca. 65 %, was von Jahr zu Jahr aber schwanken kann. Auch bei anderen Bibliotheken ist diese Kennung praktisch nicht in den Datensätzen enthalten. Die Erklärung für das halbe Dutzend Ausnahmen ist vielleicht der reine Zufall, wahrscheinlich handelt es sich aber um importierte Aufnahmen. Es ist erwähnenswert, dass der ungarische Verbundkatalog MOKKA keine bibliografischen Identifier der Nationalbibliothek enthält. Die Erklärungen und Interpretation von den Expert:innen der einzelnen Bibliotheken wären nicht nur interessant, sondern könnten auch fachlich sehr hilfreich sein.

### 3.2 Internationale Standardnummern (ISBN, ISSN)

Die undefinierten Unterfelder der beiden Standardnummern und ihre Werte stehen stellvertretend für etliche Probleme. Obwohl die absolute Zahl der Aufnahmen, die diese Art von Fehler enthalten, vernachlässigbar ist, so kann keine allgemeine Schlussfolgerung im Hinblick auf die Qualität der betreffenden Kataloge gezogen werden, viel mehr ist dies ein deprimierender Beweis für die Unfähigkeit von Katalogisierungssoftware, Fehler zu prüfen.

Solch ein Rauschen kam fast ausschließlich in ungarischen Katalogen vor und war sonst eher rar. Wahrscheinlich liegt die Ursache dafür hauptsächlich in der Qualität der Katalogisierungssoftware. Leider hat es meistens finanzielle Gründe, wenn keine bessere Software eingesetzt wird. Infolgedessen sind nicht nur die Möglichkeiten der Software begrenzt und es stehen keine Mittel für Verbesserungen zur Verfügung, sondern die finanzielle Knappheit führt auch dazu, dass ausgebildete Fachkräfte die Fachrichtung wechseln.

Tab. 4: ISBN und ISSN

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
020 Internationale Standardbuchnummer (ISBN)										
\$a ISBN	5 721 373	30,07	6 21 661	35,24	2 17 607	20,68	6 45 528	50,86	2 814 100	31,66
\$c Bezugsbedingungen*	7 044 386	37,02	9 950	0,56	45	0,00	25 153	1,98	31 914	0,36
\$q qualifizierende Zusatzinformation	-	-	87 028	4,93	8 095	0,77	6	0,00	4 529	0,05
\$z Gelöschte / ungültige ISBN	85 751	0,45	8 869	0,50	1 187	0,11	4 283	0,34	49 487	0,56
\$6 Verlinkung	-	-	68	0,00	2	0,00	12	0,00	794	0,01
\$8 Feldverknüpfung und Reihenfolge	-	-	3	0,00	-	-	6	0,00	38	0,00
\$9 lokal definiertes Unterfeld	5 759 980	30,27	-	-	-	-	91	0,01	22	0,00
Unterfelder, die nur in HUNMARC verfügbar sind										
\$d Preisabweichung	-	-	-	-	-	-	28	0,00	973	0,01
\$h zusätzliche Information	-	-	-	-	-	-	2	0,00	1 331	0,01
\$i Notiz	-	-	-	-	-	-	25	0,00	5 443	0,06
\$j Einband	-	-	-	-	-	-	283	0,00	1 155	0,01
unbekannte Unterfelder										
\$-	-	-	-	-	-	-	-	-	2	0,00
\$0	-	-	-	-	-	-	25	0,00	45	0,00
\$1	-	-	-	-	-	-	9	0,00	146	0,00
\$2	-	-	-	-	2	0,00	18	0,00	80	0,00

Fortsetzung **Tab. 4:** ISBN und ISSN

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$3	-	-	-	-	1	0,00	21	0,00	94	0,00
\$4	-	-	-	-	2	-	3	0,00	31	0,00
\$5	-	-	-	-	-	-	3	0,00	30	0,00
\$7	-	-	-	-	1	0,00	1	0,00	16	0,00
\$C	-	-	-	-	-	-	-	-	3	0,00
\$b	-	-	-	-	2	0,00	23	0,00	415	0,00
\$e	-	-	-	-	-	-	1	0,00	12	0,00
\$f	-	-	-	-	-	-	-	-	2	0,00
\$g	-	-	-	-	-	-	55	0,00	58	0,00
\$k	-	-	-	-	-	-	-	-	1	0,00
\$l	-	-	-	-	-	-	2	0,00	2	0,00
\$n	-	-	-	-	-	-	1	0,00	-	-
\$r	-	-	-	-	-	-	-	-	2	0,00
\$s	-	-	-	-	-	-	3	0,00	5	0,00
\$t	-	-	-	-	-	-	2	0,00	1	0,00
022 Internationale Standardseriennummer (ISSN)										
\$a ISSN	1,65 Mio.	8,68	427	0,02	19357	1,84	69743	5,49	139375	1,57
\$c price (nicht mehr gültiges Unterfeld)	1 343	0,01	-	-	-	-	28	0,00	906	0,01
\$l ISSN-L	40 582	0,21	50	0,00	6	0,00	4	0,00	3	0,00
\$m Gelöschte ISSN-L	-	-	-	-	-	-	-	-	-	-
\$y Falsche ISSN	-	-	-	-	10	0,00	119	0,02	247	0,00
\$z Gelöschte ISSN	42	0,00	15	0,00	47	0,00	212	0,02	154	0,00
\$2 Quelle	-	-	42	0,00	2	0,00	-	0,00	3	0,00

Fortsetzung **Tab. 4:** ISBN und ISSN

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$6 Verlinkung	-	-	-	-	1	0,00	-	-	5	0,00
\$8 Feldverknüpfung und Reihenfolge	-	-	-	-	-1	0,00	-	-	-	-
Unterfelder, die nur in HUNMARC verfügbar sind										
\$i Notiz	-	-	-	-	30	0,00	-	-	67	0,00
\$j Einband	-	-	-	-	-	-	-	-	83	0,00
unbekannte Unterfelder										
\$-	-	-	-	-	1	0,00	-	-	-	-
\$0	-	-	-	-	3	0,00	2	0,00	21	0,00
\$1	-	-	-	-	2	0,00	1	0,00	7	0,00
\$3	-	-	-	-	1	0,00	-	-	-	-
\$4	-	-	-	-	1	0,00	-	-	1	0,00
\$5	-	-	-	-	-	-	1	0,00	-	-
\$7	-	-	-	-	1	0,00	-	-	-	-
\$S	-	-	-	-	1	0,00	-	-	-	-
\$b	-	-	-	-	-	-	-	-	1	0,00
\$f	-	-	-	-	-	-	3	0,00	-	-
\$g	-	-	-	-	-	-	1	0,00	5	0,00
\$s	-	-	-	-	-	-	-	-	7	0,00
\$t	-	-	-	-	-	-	-	-	1	0,00
\$w	-	-	-	-	-	-	-	-	1	0,00
\$x	-	-	-	-	-	-	-	-	59	0,00

\* In HUNMARC: Preis.

Die unbekanntenen Unterfelder, die in den beiden ungarischen Bibliothekskatalogen und in MOKKA vorkommen, enthalten entweder nicht nachvollziehbare Werte, besondere ISBN- oder ISSN-Werte oder den Preis, da es in HUNMARC ein gesondertes Unterfeld \$c *Preis* sowohl für das ISBN- als auch für das ISSN-Feld gibt. Nach MARC 21 werden allerdings im ISBN-Unterfeld \$c die *Bezugsbedingungen* angegeben und es existiert kein ISSN-Unterfeld \$. Als externe Betrachter wissen wir zu wenig über die Katalogisierungspraxis in den beiden hier behandelten ungarischen wissenschaftlichen Bibliotheken, aber wir sind sicher, dass sich viele unbekanntene Unterfelder erklären lassen. Bei den untersuchten nicht-ungarischen Bibliotheken gibt es jedoch – zumindest was die internationalen Standardnummern betrifft – keine solchen Zweifelsfälle.

An sich ist es merkwürdig, dass es überhaupt möglich ist, Unterfelder einzufügen, die im Standard nicht existieren, und ihnen einen Wert zuzuweisen, ohne dass das zumindest zu einer Fehlermeldung führt. Katalogisierungssysteme sollten sicherstellen, dass bei Veränderungen des Standards kompetente und autorisierte Anwender:innen neue Felder und Unterfelder definieren können, damit diese auch verifiziert werden können. Eines ist sicher: Bei geeigneten Fehlerprüfungsfunktionalitäten der Katalogisierungssysteme wären solche Fehler signifikant seltener, selbst wenn die finanziellen Mittel für eine Generalrevision nicht vorhanden sind.

Die Abbildungen 3 und 4 zeigen Beispiele von Unterfeldern, die nicht existieren, mit uneinheitlichen Inhalten.

## Terms

### 0201 Isbn 1

200 Ft (1)  
 250,- Ft (1)  
 480 Ft (1)  
 56662 347 2 (1)  
 600 Ft (1)  
 67,- Ft (1)  
 85278 116 5 (1)  
 890,-Ft (1)  
 963 00 2061 0 (1)

**Abb. 3:** Unterfeld \$1 in SZTE

## Terms

### 020b Isbn b

2.köt. (1)  
 963-05-2628-X (1)

**Abb. 4:** Unterfeld \$b in MTAK

### 3.3 Einige weitere spezielle Identifier und Kennung der Katalogisierungsquelle

Auch Unterfelder mit speziellen Identifikatoren könnten indirekt inhaltsbezogene Sortierungen erlauben. Leider wird dies in Katalogsystemen noch lange nicht nutzungsfreundlich angeboten. Es ist eigentlich nicht möglich, dem System die simple Suchanfrage zu stellen, ob Ergebnisse zu bestimmten Schlagwörtern z. B. eine Kennung eines Musikverlages haben oder nicht. Nicht nur bei der Entwicklung von zukünftigen Bibliothekssystemen sollten erfahrene Katalogisierer:innen einbezogen werden, sondern sie sollten auch befähigt sein zu erkennen, welche Qualitätsprüfungsmechanismen in der täglichen Katalogisierungsarbeit gebraucht werden.

Es ist bemerkenswert, dass nur der Katalog der DNB Eintragungen von Geokoordinatenwerten (034\$d) – in vernachlässigbarer Menge im Verhältnis zum Gesamtbestand – hat, obwohl diese wesentlich für eine Darstellung der Orte auf einer Karte im Netz außerhalb des Katalogs sind. Ihr Fehlen ist allerdings in diesem Fall überhaupt nicht durch die Katalogisierungssoftware bedingt.

Die Summe aller belegten Felder 040 entspricht der Anzahl aller verarbeiteten Titelaufnahmen. Im Prinzip sollten die anderen Unterfelder von Feld 040, \$a, \$b, \$c denselben Wert enthalten – außer \$d *Bearbeitungsstelle* und \$e *Beschreibungsfestlegungen*. Dass dies möglich ist, zeigt die numerische Übereinstimmung der DNB-Werte. Auch für die MTAK-Werte ist dies annähernd erreicht. Für GENT ist nur 040\$a *Original-Katalogisierungsstelle* belegt. Vielleicht meinte man, dass Interessierte sich die Unterfeldeinträge im Katalog der jeweiligen Ursprungsinstitution anschauen?

Die wenigen Eintragungen in Feld 040\$e *Beschreibungsfestlegungen* stehen in engem Zusammenhang mit dem Ein-Zeichen-Wert an der 18. Position der Satz Kennung (*leader*), der die Form der Formalerschließung bezeichnet. Dieser Wert ist von entscheidender Bedeutung dafür, ob bei der Programmierung von Konversionssoftware das Augenmerk z. B. auf das Vorkommen von Interpunktionszeichen in einem der Datenfelder wie etwa 245\$a *Titel* gelenkt wird.

Es stellt sich die Frage, warum in anderen Bibliothekskatalogen diese Einheitlichkeit fehlt? Kann es sein, dass das Unterfeld \$b *Katalogisierungssprache* nicht obligatorisch ist? Oder kann das Katalogisierungssystem etwa nicht prüfen, ob die Werte in 040\$b *Katalogisierungssprache* zu den Werten in Position 35–37 des Kontrollfelds 008 *Sprache* passen? (Weiterhin kann es sich natürlich auch immer um importierte Titelaufnahmen handeln.)



Tab. 5: Spezielle Identifier und Katalogisierungsquelle

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
024\$a Andere Standard-nummer oder Code	8 479 Mio.	44,56	103 619	5,87	1 644	0,16	867	0,07	115 987	1,30
028\$a Verlegernummer	1 408	7,40	7 391	0,42	-	-	-	-	256 322	2,88
030\$a CODEN	6 386	0,03	37	0,00	-	-	12 108	0,95	58	0,00
032\$a Postregistrierungsnummer	813	0,00	10	0,00	-	-	1	0,00	1	0,00
034\$d Koordinaten – westlichster Längengrad	55 880	0,29	46	0,00	-	-	-	-	12	0,00
035\$a System-Kontrollnummer	19 Mio.	100	283 252	16,06	774 923	73,64	864 342	68,10	8 888 923	99,99
040 Katalogisierungsquelle										
\$a Original-Katalogisierungsstelle	19 Mio.	100	1 764 209	100	1 052 265	100	1 268 276	100	8 889 434	100,00
\$b Katalogisierungssprache	19 Mio.	100	-	-	1 052 286	99,99	972 901	76,65	588	0,01
\$c Übertragungsstelle	19 Mio.	100	-	-	1 052 202	99,99	4 470	0,35	223	0,00
\$d Bearbeitungsstelle	19 Mio.	100	-	-	119 379	11,34	237 527	18,71	235 425	2,65
\$e Beschreibungsfestlegungen	606 123	3,19	-	-	22	0,00	10	0,00	4	0,00

### 3.4 Sprachcodes

Ähnlich wie bei den beiden *Internationalen Standardnummern* gibt es einige unbekannte Unterfelder zu Feld 041 *Sprachcode* in ungarischen Bibliotheken, die in MARC 21 und HUNMARC nicht vorgesehen sind und die als Werte zumeist Sprachcodes enthalten. Dabei mag es sich um lokal definierte Unterfelder handeln (die hoffentlich dokumentiert sind, auch wenn externe Forscher:innen das mangels Zugänglichkeit der Dokumentation nur schwer herausfinden können); es könnten aber auch Eingabefehler sein. Oder sind auch für dieses Problem importierte Titelaufnahmen der Grund? Falls dies der Fall sein sollte, stellt sich die Frage, warum das datenerhaltende Katalogisierungssystem die Fehler nicht herausfiltern konnte?

Keiner der Kataloge gibt die Sprache in allen Titelaufnahmen an und es ist auffällig, dass im GENT- und im SZTE-Katalog sehr wenige Aufnahmen Sprachcodes enthalten. Die DNB hat den höchsten Anteil an sprachlichen Kennungen, allerdings ist die DNB ja auch eine Nationalbibliothek. Nach Erfahrungswerten eines der Autoren ist dieser Anteil bei der Ungarischen Nationalbibliothek ähnlich dem der DNB.

Nach MARC 21 sind die ersten drei Zeichen sowohl in Feld 040\$b *Katalogisierungssprache* als auch in Feld 041\$a *Sprachcode* üblicherweise dieselben wie die Positionen 35–37 des Feldes 008 *Sprache*. Es fragt sich, wie das in der Praxis erreicht wird und, wenn nicht, warum nicht.

Tab. 6: Sprachcodes

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
041 Sprachcode										
\$a Sprachcode des Textes/der Tonspur oder des separaten Titels	15,7 Mio.	82,58	101 564	5,76	993 704	94,43	1 488 840	11,73	4 898 948	55,11
\$b Sprache der Zusammenfassung oder des Abstracts	-	-	4 791	0,27	14 261	1,36	15 930	1,26	103 089	1,16
\$d Sprachcode von gesungenem oder gesprochenem Wort	-	-	19	0,27	2	0,00	237	0,02	15 407	0,17
\$e Sprachcode von Operntexten	-	-	15	0,00	25	0,00	80	0,01	4 439	0,05
\$f Sprachcode des Inhaltsverzeichnis	-	-	100	0,01	1 095	0,10	1 680	0,13	12 193	0,14
\$g Sprachcode von Begleitmaterial mit Ausnahme von Operntexten	-	-	119	0,01	2 439	0,23	3 212	0,25	32 333	0,36
\$h Sprachcode der Original- und/oder Zwischenübersetzung des Textes	569 679	2,99	12 014	0,68	55 266	5,25	70 562	5,56	380 844	4,28
\$j Sprachcode der Untertitel oder Beschriftung	-	-	25	0,00	18	0,00	2	0,00	32	0,00
\$k Sprachcode der Zwischenübersetzung	-	-	9	0,00	41	0,00	1	0,00	23	0,00
\$m Sprachcode von Begleitmaterialien mit Ausnahme von Operntexten	-	-	9	0,00	11	0,00	1	0,00	7	0,00

Fortsetzung **Tab. 6:** Sprachcodes

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$n Sprachcode der Originalsprache von Operntexten	-	-	2	0,00	1	0,00	12	0,00	23	0,00
\$2 Quelle	-	-	-	-	-	-	1	0,00	2	0,00
\$6 Verlinkung	-	-	-	-	1	0,00	-	-	1	0,00
\$8 Feldverknüpfung und Reihenfolge	42 000	0,22	-	-	-	-	-	-	-	-
unbekannte Unterfelder										
\$0	-	-	-	-	1	0,00	-	-	1	0,00
\$1	-	-	-	-	-	-	-	-	1	0,00
\$3	-	-	-	-	1	0,00	-	-	-	-
\$4	-	-	-	-	-	-	-	-	3	0,00
\$7	-	-	-	-	1	0,00	-	-	-	-
\$c	-	-	-	-	-	-	6	0,00	31	0,00
\$i	-	-	-	-	2	0,00	6	0,00	13	0,00
\$l	-	-	-	-	9	0,00	29	0,00	60	0,00
\$p	-	-	-	-	1	0,00	4	0,00	6	0,00
\$q	-	-	-	-	-	-	-	-	1	0,00
\$r	-	-	-	-	1	0,00	8	0,00	41	0,00
\$s	-	-	-	-	2	0,00	3	0,00	21	0,00
\$t	-	-	-	-	2	0,00	2	0,00	10	0,00
\$u	-	-	-	-	3	0,00	1	0,00	2	0,00
\$v	-	-	-	-	-	-	-	-	1	0,00
\$x	-	-	-	-	-	-	16	0,00	11	0,00

### 3.5 Die inhaltsbezogenen Felder der Feldergruppe 01X–09X mit Ausnahme der 08X-Felder

Die Werte in den Feldern 043–052 stammen aus kontrollierten Vokabularen. Auch für diese gilt, dass sie die Suche und Auswahl nach Themen ermöglichen. Beispielsweise könnte man mit einer einfachen Abfrage in nutzungsfreundlichem Format den Wert eines beliebigen dieser Felder mit einem beliebigen Schlagwort aus 6XX kombinieren – etwa einen Deskriptor aus Feld 651 *Schlagworteintragung – Geografischer Name* mit einem *Ländercode der veröffentlichenden/herstellenden Stelle* aus 044. Sagen wir *651\$aPilisborosjenő AND 044\$bDE-BW* (wobei *Pilisborosjenő* ein ungarisches Dorf mit teilweise deutscher Bevölkerung ist und *DE-BW* der Code für Baden-Württemberg nach ISO 3166-2:DE<sup>11</sup>). Solche Kombinationen mögen aus heutiger Sicht nicht nach einer sehr realistischen Aufgabe klingen, aber in dem Maß, wie sich die Informationsmengen im Netz, inklusive in Online-Katalogen, mit der Zeit vervielfältigen, wird das Herausfiltern von Rauschen wichtiger werden denn je.

Bei den vier ausgewählten Bibliotheken und in MOKKA spielen diese Felder mit wenigen Ausnahmen eine untergeordnete Rolle. Vermutlich ist es kaum bekannt, dass sie auch für die Suche verwendet werden könnten.

Es ist bemerkenswert, dass *044\$a MARC-Ländercode* in MOKKA häufig vorkommt, obwohl seine Verwendung für Ungarn untypisch ist. Zugleich ist *\$c ISO-Ländercode* normalerweise leer, obwohl die Verwendung des Feldes in Ungarn obligatorisch ist. Wahrscheinlich lässt sich das damit erklären, dass das Unterfeld *\$c* für die ISO 3166-1- und ISO 3166-2-Ländercodes im Jahr 2001 als neues Unterfeld in MARC 21 eingeführt wurde und *\$a* seither für MARC-Ländercodes reserviert bleibt. Da es vermutlich zumindest für diesen Fall keine nachträgliche Übertragung der Inhalte in das neue Unterfeld gab, blieben die ISO-Ländercodes im Unterfeld *\$a* stehen und wurden so bei einer Datenübernahme nach MOKKA transferiert.

Im DNB-Katalog wurde offenbar eine Datenanpassung der ISO-Ländercodes vorgenommen, denn fast 100 % der Titelaufnahmen enthalten das Feld *044\$c ISO-Ländercode*. Leider haben wir keinen Zugang zu den NSZL-Daten, aber soweit wir wissen, wurde auch dort eine Datenanpassung vorgenommen. Eine andere Frage ist es, ob die Werte in Feld *044\$c ISO-Ländercode* zu denen in Position 15–17 von Feld 008 *Ort der Veröffentlichung, Herstellung oder Realisierung* passen. In anderen Worten: Wird das von Katalogisierungssystemen geprüft?

---

<sup>11</sup> Internationale Organisation für Normung: ISO 3166-2:2020. Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code. <https://www.iso.org/standard/63546.html> (25.12.2020).

Tab. 7: Inhaltliche Identifier der Feldergruppe 01X–09X mit Ausnahme der 08X-Felder

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
043 Code für geografische Gebiete										
\$a Code für geografische Gebiete	-	-	91 420	5,18	-	-	14	0,00	8 144	0,09
\$b lokaler Code für geografische Gebiete	-	-	28	0,00	-	-	-	-	1	0,00
\$c ISO-code	536 976	2,82	84	0,00	-	-	14	0,00	-	-
\$2 Quelle des lokalen Codes für geografische Gebiete	-	-	2	0,00	-	-	-	-	-	-
unbekannte Unterfelder										
\$d	-	-	-	-	-	-	-	-	13	0,00
\$x	-	-	-	-	-	-	3	0,00	-	-
044 Ländercode der verfügbaren/ herstellenden Stelle										
\$a MARC-Ländercode	-	-	-	-	3	0,00	18	0,00	963 690	10,84
\$b lokaler Code für eine Unterinheit	-	-	556	0,03	-	-	1	0,00	2	0,00
\$c ISO-Ländercode	18,8 Mio.*	98,76	1 328	0,08	-	-	-	-	30	0,00
\$9 lokal definiertes Unterfeld	-	-	-	-	3	0,00	18	0,00	-	-
unbekannte Unterfelder										
\$d	-	-	-	-	-	-	-	-	1	0,00
\$f	-	-	-	-	-	-	1	0,00	-	-

Fortsetzung **Tab. 7:** Inhaltliche Identifier der Feldergruppe 01X–09X mit Ausnahme 08X-Felder

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$h	-	-	-	-	-	-	1	0,00	2	0,00
052 Geografischer Klassifikationscode										
\$a Gebietscode nach geografischer Klassifikation	-	-	185	0,01	-	-	-	-	31	0,00
\$b Untergebietscode nach geografischer Klassifikation	-	-	42	0,00	17 625	1,67	-	-	25	0,00
072 Schlagwortkategoriecode										
\$a Schlagwortkategoriecode	-	-	5 949	0,34	17 625	1,67	3	0,00	3 581	0,04
\$x Schlagwortkategoriecode – Untereinheit	-	-	453	0,03	-	-	-	-	-	-
\$2 Quelle	-	-	5 581	0,32	17 423	0,66	-	-	28	0,00
090 [lokal definiertes Feld]										
\$6	-	-	-	-	-	-	1	0,00	-	-
\$7	-	-	-	-	-	-	1	0,00	-	-
\$a	-	-	-	-	-	-	11 370	16	-	-
\$b	-	-	-	-	-	-	16	0,00	-	-
\$j	-	-	-	-	-	-	4	0,00	-	-
\$a	-	-	-	-	-	-	1	0,00	-	-
\$b	-	-	-	-	-	-	1	0,00	-	-

\* Die hohe Zahl (nahezu 100 %) zeugt eventuell davon, dass die vollständige Nutzung der MARC-Felder etwas ernster genommen wurde.

### 3.6 Felder für Klassifikationen in 08X

Die Werte in den Feldern 080, 082, 083, 084 und 085 sind Notationen klassischer Klassifikationssysteme. Die ersten drei Felder werden für die Universelle Dezimalklassifikation (UDK) und die Dewey-Dezimalklassifikation (DDC) verwendet; 084 kann für jedes beliebige hierarchische oder aufzählende System und 085 für die Abbildung von Elementen synthetisierender Klassifikationssysteme verwendet werden.

In Ungarn nutzen die meisten Bibliotheken die UDK, einige die DDC; wahrscheinlich werden die Werte in den anderen beiden Feldern nicht aktiv während des eigenen Katalogisierungsprozesses erfasst, sondern entstehen wohl durch den Import von Titelaufnahmen.

Die DNB setzt nur die DDC ein; ein signifikanter Anteil der Titelaufnahmen enthält entsprechende Notationen. Die Praxis der DNB dürfte sich nur mit der NSZL vergleichen lassen. Die Verwendung anderer und synthetisierter Klassifikationen ist auch für die DNB (und die Finnische Nationalbibliothek) typisch. Alles in allem scheinen nicht-ungarische Bibliotheken etwas mehr Gebrauch von Klassifikationssystemen zu machen als die ungarischen Bibliotheken.

All das legt nahe, dass diese klassischen, mit Kunstsprache operierenden Klassifikationssysteme im Bibliothekswesen immer noch von eminenter Bedeutung sind. Wäre es nicht eine Idee, wenn bei einer Suchanfrage mit einem bestimmten Schlagwort Nutzer:innen die Option hätten, weitere Titel der Trefferliste hinzuzufügen, welche mit einer UDK- oder sonstigen Notation erschlossen sind und wobei Klassenbenennung und Suchwort übereinstimmen? Zweifellos ist das bisher noch ein Traum, obwohl eine technische Implementierung nicht zu schwer erscheint.<sup>12</sup>

---

**12** Es gab einmal – und gibt wahrscheinlich immer noch einen hochrangigen ungarischen Bibliotheksleiter, der die UDK nicht mehr einsetzen will. Noch schlimmer ist die Absicht, die „einfachere“ Version der UDK einzusetzen. Dies bedeutet ein grundlegendes Fehlverstehen der UDK. Hoffentlich werden mit kommenden Generationen von Bibliothekssoftware, in denen sich einfacher mit UDK suchen lässt, solche unprofessionellen Vorstellungen verschwinden.



Tab. 8: Inhaltliche Identifier in der Feldergruppe 01X–09X

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
080 Notation nach der Universellen Dezimalklassifikation										
\$a Notation	-	-	100 245	5,68	535 691	50,91	1 038 987	81,85	3 786 644	42,60
\$b individualisierender Notationszusatz	-	-	5	0,00	3	0,00	14	0,00	1 235	0,01
\$x Elemente einer Hilfstabelle	-	-	26	0,00	6	0,00	81	0,01	184	0,00
\$0 IDN des Normdatensatzes	-	-	-	-	6	0,00	58	0,00	124	0,00
\$1 Persistenter Link zum Datensatz	-	-	-	-	1	0,00	28	0,00	92	0,00
\$2 Ausgabennummer	-	-	-	-	3	0,00	127	0,01	613	0,01
\$6 Verlinkung	-	-	-	-	1	0,00	93	0,01	271	0,00
\$8 Feldverknüpfung und Reihenfolge	-	-	2	0,00	16	0,00	126	0,01	553	0,01
\$9 (lokal definiert) unbekannte Unterfelder	-	-	-	-	13	0,00	53	0,00	62	0,00
\$A	-	-	-	-	4	0,00	-	-	12	0,00
\$	-	-	-	-	2	0,00	-	-	-	-
\$#	-	-	-	-	-	-	-	-	160	0,00
\$!	-	-	-	-	1	0,00	-	-	-	-
\$c	-	-	-	-	-	-	8	0,00	466	0,01
\$d	-	-	-	-	-	-	40	0,00	116	0,00
\$e	-	-	-	-	-	-	1	0,00	3	0,00

Fortsetzung **Tab. 8:** Inhaltliche Identifier in der Feldergruppe 01X–09X

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$g	-	-	-	-	-	-	12	0,00	17	0,00
\$h	-	-	-	-	-	-	-	-	10	0,00
\$i	-	-	-	-	1	0,00	-	-	1	0,00
\$j	-	-	-	-	-	-	-	-	1	0,00
\$l	-	-	-	-	-	-	-	-	5	0,00
\$m	-	-	-	-	-	-	2	0,00	-	-
\$n	-	-	-	-	-	-	2	0,00	5	0,00
\$p	-	-	-	-	-	-	2	0,00	3	0,00
\$q	-	-	-	-	-	-	2	0,00	5	0,00
\$r	-	-	-	-	-	-	2	0,00	1	0,00
\$s	-	-	-	-	-	-	16	0,00	53	0,00
\$t	-	-	-	-	-	-	-	-	7	0,00
\$v	-	-	-	-	-	-	-	-	22	0,00
\$w	-	-	-	-	-	-	1	0,00	-	0,00
\$y	-	-	-	-	-	-	-	-	4	0,00
\$z	-	-	-	-	-	-	-	-	2	0,00
\$3	-	-	-	-	1	0,00	126	0,01	453	0,01
\$4	-	-	-	-	1	0,00	-	-	13	0,00
\$5	-	-	-	-	-	-	49	0,00	4 362	0,05
\$7	-	-	-	-	-	-	-	-	157	0,00
082 Notation nach der Dewey Decimal Classification										
\$a Notation	7 365 118	38,71	192 905	10,39	1	0,00	745	0,06	86 932	0,98

Fortsetzung **Tab. 8:** Inhaltliche Identifier in der Feldergruppe 01X–09X

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$b individualisierender Notationszusatz	-	-	4 082	0,23	-	-	1	0,00	75	0,00
\$q Vergabestelle	-	-	493	0,03	-	-	-	-	8	0,00
\$2 Ausgabennummer	7 365 077	38,70	112 572	6,38	-	-	50	0,00	39 051	0,44
\$8 Feldverknüpfung und Reihenfolge	-	-	264	0,01	-	-	-	-	4	0,00
unbekannte Unterfelder										
\$d	-	-	-	-	-	-	-	-	7	0,00
\$v	-	-	-	-	1	0,00	-	-	-	-
\$x	-	-	769	0,04	-	-	-	-	-	-
\$0	-	-	-	-	-	-	-	-	2	0,00
\$3	-	-	-	-	-	-	-	-	2	0,00
\$5	-	-	-	-	-	-	-	-	1	0,00
\$7	-	-	-	-	-	-	-	-	1	0,00
083 Zusätzliche DDC-Notation										
\$a Notation	-	-	582	0,03	-	-	7	0,00	21	0,00
084 Andere Notation										
\$a Notation	4 602 135	24,18	52 588	2,98	1	0,00	5 136	0,40	87	0,00
085 Synthetische Notation und ihre Bestandteile										
\$a Nummer, unter der sich die Anweisungen befinden – einzelne Nummer oder Anfangsnummer des Bereichs	1 348 402	7,09	-	-	-	-	3	0,00	9	0,00

Die Kultur der UDK-Erschließung in ungarischen Bibliotheken hat eine fast hundertjährige Tradition, was ihre signifikante Verwendung sowohl in der Nationalbibliothek als auch in wissenschaftlichen Bibliotheken erklärt. Bei den MTAK-Daten passen die 080-Werte in den Unterfeldern \$0 *IDN des Normdatensatzes*, \$1 *Persistenter Link zum Datensatz*, \$b und \$x *Elemente einer Hilfstafel* entweder überhaupt nicht zum erwarteten Inhalt oder sie sind häufig unzulässig. Ähnlich sieht es bei Feld 080 in den SZTE-Daten aus. Auch die unbekannteren Unterfelder enthalten überwiegend irgendeine Art an Notationen. Bei HUNMARC passen die Inhalte der Unterfelder \$c und \$d von Feld 080 normalerweise auch nicht zum erwarteten Inhalt dieser HUNMARC-Unterfelder. Das Eigenartige am SZTE-Lokalfeld 090 ist, dass die zwei Unterfelder, \$a und \$b, jeweils mit verschiedenen Werten gibt. Es ist bei Betrachtung der Daten ebenfalls interessant, dass ein signifikanter Anteil der SZTE-Daten in Feld 084 nicht in MOKKA vorhanden ist. \$!-Werte wurden von MTAK nicht nach MOKKA importiert, \$#-Werte wurden aus anderen Quellen übernommen. Natürlich kann das mit der Qualität des beim Import verwendeten Kontrollprogramms zusammenhängen, worüber den Autoren keine Informationen vorliegen. All das ist von relativ geringer Bedeutung, andererseits ist es doch insgesamt enttäuschend und hängt, wie wir schon angedeutet haben, in erster Linie nicht an den Katalogisierungssystemen, sondern an ungenügenden Datenvalidierungsfähigkeiten der jeweiligen Katalogisierungssoftware.

## 4 Die Feldergruppe 6XX für Schlagwort-eintragungen

### 4.1 Feld 600 Schlagworteintragung – Personennamen

Bei den für Schlagwörter vorgesehenen Feldern vergleichen wir nur die Felder 600, 651, 653 *Indexierungsterm – nicht normiert* und 655 *Indexierungsterm – Genre/Formschlagwort*. Die Eigennamenfelder 610 *Schlagworteintragung – Körperschaftsname*, 611 *Schlagworteintragung – Kongressname*, und 630 *Schlagworteintragung – Einheitstitel* haben ähnliche Eigenschaften. Die Felder 647 *Historisches Einzelereignis* und 654 *Facettierte Sachbegriffe* werden in den untersuchten Bibliotheken nicht verwendet.

**Tab. 9: Verwendung von Feld 600 im Vergleich**

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
600 Schlagworteintragung – Personenname										
\$a Personenname	433 676	2,28	79 207	4,49	172 001	16,35	64 228	5,06	632 725	7,12
\$b Zählung	88 640	0,05	4 343	0,25	5 582	0,53	3 788	0,30	30 695	0,35
\$c Titulaturen und andere Ordnungsnamen	39 799	0,21	16 501	0,94	29 909	2,84	8 245	0,65	173 296	1,95
\$d Mit dem Namen verknüpfte Daten	351 243	1,85	61 884	3,51	139 159	13,22	6 592	0,52	490 838	5,52
\$e Funktionsbezeichnung	–	–	67	0,00	23 725	2,25	2	0,00	845	0,01
\$f Erscheinungsjahr eines Werkes	82	0,00	15	0,00	2	0,00	–	–	61	0,00
\$g Sonstige Informationen	292	0,00	80	0,00	722	0,07	3	0,00	3 118	0,04
\$h Medium	–	–	–	–	1	0,00	–	–	–	–
\$j Zuschreibungsvermerk	–	–	–	–	6	0,00	–	–	2 328	0,03
\$k Formales Unterschlagnwort*	–	–	36	0,00	–	–	–	–	9	0,00
\$l Sprache eines Werkes	–	–	61	0,00	1	0,00	–	–	35	0,00
\$m Medium der Musikaufführung**	212	0,00	15	0,00	238	0,02	–	–	416	0,00
\$n Zählung des Teils/der Abteilung eines Werkes	819	0,000	186	0,01	4	0,00	1	0,00	19	0,00
\$o Angabe des Musikarrangements	–	–	–	–	11	0,00	–	–	–	–

Fortsetzung **Tab. 9:** Verwendung von Feld 600 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$p Titel eines Teils/einer Abteilung eines Werkes	478	0,00	147	0,01	-	-	-	-	28	0,00
\$q Vollständigere Namensform	-	-	2 297	0,13	3	0,00	-	-	43	0,00
\$s Version	21	0,00	6	0,00	2	0,00	2	0,00	33	0,00
\$t Titel eines Werkes	36 496	0,19	9 398	0,53	10 790	1,03	50	0,00	21 474	0,24
\$0 IDN des Normdatensatzes	378 007	1,99	3 860	0,22	3	0,00	1	0,00	12	0,00
\$1 URI	-	-	1	0,00	-	-	-	-	6	0,00
\$2 Quelle der Ansetzung oder des Terms	433 870	2,27	79 207	4,49	1	0,00	63 060	4,97	75 881	0,85
\$3 Materialspezifikation	-	-	1	0,00	-	-	81	0,01	487	0,01
\$4 Funktionsbezeichnung	-	-	49	0,00	3	0,00	-	-	575	0,01
\$6 Verlinkung	-	-	280	0,00	164	0,01	-	-	19	0,00
\$8 Feldverknüpfung und Reihenfolge	18 733	0,10	-	-	1	0,00	-	-	4	0,00
unbekannte Unterfelder										
\$#	-	-	-	-	-	-	-	-	447	0,01
\$D	-	-	-	-	-	-	-	-	2	0,00
\$M	-	-	-	-	-	-	-	-	2	0,00
\$i	-	-	-	-	-	-	-	-	3	0,00
\$'	-	-	-	-	-	-	-	-	1	0,00

\* Nicht dasselbe wie 655 Indexierungsterm – Genre/Formschlagwort.

\*\* In HUNMARC Verwandtschaftszusatz.

Die Werte in den mit Großbuchstaben bezeichneten Unterfeldern in MOKKA stellen unverständliches Rauschen dar.

Das relativ neue Element \$0 *IDN des Normdatensatzes* kommt in ungarischen Katalogen kaum vor. Bei der DNB hingegen haben die meisten Personennamen bereits eine ID eines Autoritätsdatensatzes und sind in der Gemeinsamen Normdatei (GND) enthalten, die folglich auch als \$2 *Quelle* der meisten Einträge in Feld 600 fungiert. Die Situation ist wohl bei GENT ähnlich. Für die in SZTE katalogisierten normierten Personennamen werden als Quellen die bibliografischen Aufnahmen der eigenen Bibliothek angegeben.

Das neue Element \$1 *Persistenter Link zum Datensatz* wird in keinem Katalog verwendet. Zugegebenermaßen sind beide Unterfelder ziemlich neu in MARC 21, und es ist nicht sicher, ob die untersuchten Bibliotheken sie in ihre Katalogisierungssysteme aufnehmen können. Die Systeme wurden möglicherweise so aufgesetzt, dass Felddefinitionen „fest verdrahtet“ sind und von der jeweiligen Bibliothek nicht modifiziert werden können.

Dasselbe trifft auch auf die anderen thematischen 6XX-Felder zu. Weiterhin wird \$0 von der DNB am häufigsten verwendet, und diese Praxis sollte nur mit einer anderen Nationalbibliothek wie der NSZL verglichen werden. Es ist merkwürdig, dass \$2 *Quelle* bei der DNB etwas öfter vorkommt als \$a *Personenname*. In den betreffenden 273 Titelaufnahmen fehlt der Personenname, die Normdateieinträge können jedoch über das vorhandene Feld \$0 *IDN des Normdatensatzes* identifiziert werden. Auch für diese Unstimmigkeit könnte vielleicht die Software verantwortlich sein, wenn sie dies zulässt, ohne die Katalogisierer:innen zu warnen. Personennamen werden üblicherweise nicht anhand von Namens- oder Schlagwortlisten, sondern anhand der zu erschließenden Ressource erfasst (natürlich in einheitlich standardisierter Form). In solchen Fällen fehlt der Nachweis einer externen Quelle, so dass die Anzahl der Quellenwerte üblicherweise geringer ist als die Anzahl der Namenwerte.

Bei MTAK ist die Verwendung des Personennameneintrags verhältnismäßig am höchsten (\$a kommt auf 16,35%). Diese Bibliothek liegt auch mit einigen anderen Unterfeldern über dem Durchschnitt (mit dem Namen verknüpfte \$c *Titularien* und \$d *Daten*). Darüber hinaus sind die Werte in \$2 *Quelle* im Vergleich am konsistentesten. Übrigens werden \$c und \$d in ungarischen Bibliotheken intensiv verwendet. Solche Auffälligkeiten sind interessant und es stellt sich die Frage, wie sie sich zur Katalogisierungspraxis der jeweiligen Bibliothek verhalten.

## 4.2 Feld 648 Schlagworteintragung – Zeitschlagwort

Die separate Erschließung mit Zeitschlagwörtern (Jahr, Jahrzehnt, Jahrhundert, Jahrtausend usw.) steht in engem Zusammenhang mit 045 *Zeitabschnitt der Eintragung* und 046 *Speziell kodierte Datumsangaben* sowie deren koordiniertem Gebrauch.

Tab. 10: Verwendung von Feld 648 im Vergleich

MARC-21-Da- tenelement	DNB	%	GENT	%	MTAK %	SZTE %	MOKKA %
648 Schlagworteintragung – Zeitschlagwort							
\$0 IDN des Normdaten- satzes	–	–	147	0,01	–	–	–
\$2 Quelle der Anset- zung oder des Terms	439 379	2,31	4 976	0,28	–	–	2
\$a Zeit- schlagwort	439 379	2,31	4 976	0,28	–	–	472

Die Verwendung dieses Felds ist in der DNB und der Finnischen Nationalbibliothek, die in dieser Untersuchung nicht berücksichtigt wird, ungefähr gleich. Die Unterfelder \$2 und \$a werden gleichermaßen befüllt. Die Verwendung ist für nicht-ungarische wissenschaftliche Bibliotheken nicht typisch und für ungarische noch weniger, was damit zusammenhängt, dass sie das Unterfeld \$y *chronologischer Unterbereich* in den anderen Schlagworteintragungen verwenden. Die Werte in MOKKA stammen aus einer Forschungsbibliothek mit historischem Schwerpunkt. Bei der NSZL müssen die Zeitschlagwörter aus dem Universalthesaurus *Köztaurusz* entnommen werden, wo sie von vornherein als Zeitschlagwörter qualifiziert sind, somit müssen sie zwangsweise in Feld 648 erfasst werden.

Vielleicht liegt der Grund für diese geringe Verwendung darin, dass Katalogisierer:innen Zeitschlagwörter nicht separat behandeln, sondern sie mit anderen Schlagwörtern kombinieren. Oder das Katalogisierungswerkzeug lässt möglicherweise diese Unterscheidung nicht zu.

Mit Ausnahme der Finnischen Nationalbibliothek, bei der neben 648\$a *Zeitschlagwort* auch das Feld 045 *Zeitabschnitt der Eintragung* relativ signifikant verwendet wird (4,9%), fehlen in den hier untersuchten Katalogen sowohl das



Feld 045 als auch das Feld 046 *Speziell kodierte Datumsangaben* ganz und gar. Dabei wären diese Elemente für die inhaltliche Suche sehr wichtig.

Die fehlende separate Erfassung der Schlagwörter in 648 *Zeitschlagwort* und in 647 *Historisches Einzelereignis* sowie die sehr seltene Verwendung von 045 *Zeitabschnitt der Eintragung* liegt in erster Linie nicht daran, dass die zu erschließenden Ressourcen keine zeitlichen Aspekte beinhaltet haben. Vermutlich mangelte es bisher an Sensibilität gegenüber den Möglichkeiten des MARC-21-Formats, chronologische Einordnungen vorzunehmen, natürlich nur, falls auch zeitliche Aspekte in den zu erschließenden Ressourcen vorliegen. Darauf deutet auch die Tatsache hin, dass Zeitschlagwörter mit IDNs der Normdatensätze nur bei GENT vorkommen.

Das gänzliche Fehlen von Einträgen in 647 *Historisches Einzelereignis* lässt ebenfalls darauf schließen, dass der zeitlichen Dimension bei der Sacherschließung keine besondere Bedeutung beigemessen wird. Auch die Einschränkung von Suchergebnissen nach Schlagwortkategorien ist auch bei der Suche nützlich. Da im Laufe der Zeit die automatischen Erschließungssysteme merklich verbessert werden, wird den Zeitaspekten und damit den Zeitschlagwörtern eine bedeutendere Rolle zukommen.

### 4.3 Feld 650 Schlagworteintragung – Sachschlagwort

Dieses Feld ist das zentrale Element der Sacherschließung. An ihm lässt sich über ihre Qualität eine Menge ablesen. Es ist von seiner Bedeutung her vergleichbar mit den 1XX-Feldern für die Formalerschließung, die in dieser Untersuchung nicht behandelt wird.

Tab. 11: Verwendung von Feld 650 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
650 Schlagworteintragung – Sachschlagwort										
\$a Schlagwort oder geografischer Name als Eintragungselement	3 612 054	18,98	689 532	39,08	140 030	13,31	733 809	57,82	3 658 732	41,16
\$b Sachbegriff, der auf ein Schlagwort folgt, das einen geografischen Namen enthält	–	–	50	0,00	12	0,00	26	0,00	17 231	0,19
\$c Ort des Ereignisses	–	–	2 690	0,15	13	0,00	17	0,00	9 057	0,10
\$d Zeitraum	–	–	3 363	0,19	126	0,00	35	0,00	3 503	0,04
\$e Funktionsbezeichnung	–	–	9	0,00	4	0,00	3	0,00	22	0,00
\$g Sonstige Informationen	208 211	1,09	29	0,00	1	0,00	22	0,00	56	0,00
\$0 IDN des Normdatensatzes	2 844 292	14,95	67 743	3,84	–	–	–	–	81	0,00
\$1 Persistenter Link zum Datensatz	–	–	–	–	–	–	–	–	54	0,00
\$2 Quelle der Ansetzung oder des Terms	3 612 054	18,98	689 532	39,08	64	0,01	733 857	57,82	1 709 757	19,23
\$3 Spezifizierte Materialien	–	–	–	–	–	–	1 922	0,15	8 952	0,10
\$4 Funktionsbezeichnung (Code)	–	–	–	–	–	–	–	–	196	0,00
\$6 Verlinkung	–	–	337	0,02	–	–	–	–	2	0,00
\$8 Feldverknüpfung und Reihenfolge	209 766	1,10	2	0,00	–	–	–	–	–	–

Fortsetzung **Tab. 11:** Verwendung von Feld 650 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
undefinierte Unterfelder										
\$#	-	-	-	-	-	-	-	-	7 891	0,09
\$\$	-	-	-	-	-	-	-	-	5	0,00
\$5	-	-	-	-	-	-	-	-	10	0,00
\$7	-	-	-	-	-	-	-	-	2	0,00
\$A	-	-	-	-	-	-	-	-	1	0,00
\$B	-	-	-	-	-	-	-	-	2	0,00
\$H	-	-	-	-	-	-	-	-	1	0,00
\$M	-	-	-	-	-	-	-	-	1	0,00
\$P	-	-	-	-	-	-	-	-	1	0,00
\$S	-	-	-	-	-	-	-	-	5	0,00
\$T	-	-	-	-	-	-	-	-	6	0,00
\$V	-	-	-	-	-	-	-	-	2	0,00
\$X	-	-	-	-	-	-	-	-	1	0,00
\$Z	-	-	-	-	-	-	-	-	3	0,00
\$f	-	-	-	-	3	0,00	7	0,00	36	0,00
\$h	-	-	-	-	-	-	-	-	11	0,00
\$i	-	-	-	-	-	-	2	0,00	40	0,00
\$j	-	-	-	-	1	0,00	1	0,00	12	0,00
\$k	-	-	-	-	-	-	9	0,00	58	0,00
\$l	-	-	-	-	-	-	5	0,00	43	0,00
\$m	-	-	-	-	-	-	1	0,00	31	0,00

Fortsetzung **Tab. 11:** Verwendung von Feld 650 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
undefinierte Unterfelder										
\$n	-	-	-	-	-	-	3	0,00	33	0,00
\$o	-	-	-	-	1	0,00	2	0,00	4	0,00
\$p	-	3	0,00	-	-	-	6	0,00	49	0,00
\$r	-	-	-	-	-	-	6	0,00	16	0,00
\$s	-	3	0,00	-	-	-	5	0,00	43	0,00
\$t	-	1	0,00	-	-	-	26	0,00	108	0,00
\$u	-	-	-	-	-	-	-	-	5	0,00

Bei der DNB hat das Unterfeld 650\$0 *IDN des Normdatensatzes* häufiger einen Wert (14,95%) als das Feld 600\$0. Das trifft auch für \$a *Sachschlagwort oder geografischer Name als Eintragungselement* zu, dessen Werte vollständig mit denen im Unterfeld \$2 *Quelle der Ansetzung oder des Terms* übereinstimmen. Dies hängt sicherlich mit der Zuständigkeit der DNB für die Gemeinsame Normdatei zusammen. Die Unterfelder \$b–\$e sind dagegen völlig leer.

Mit Ausnahme von MTAK passen die Werte der Unterfelder \$2 *Quelle der Ansetzung oder des Terms* vollständig zu denen in \$a *Sachschlagwort oder geografischer Name als Eintragungselement*. Nach unserer Erfahrung wird in der NSZL die Konsistenz dieser Unterfelder durch das Katalogisierungssystem gewährleistet, da alle thematischen Deskriptoren obligatorisch dem *Köztaurusz* und geografische Namen einer seiner Komponenten, dem *Geotaurusz*, entnommen werden. Dieser Universalthesaurus wird auch in vielen anderen ungarischen Bibliotheken eingesetzt.

Es ist erwähnenswert, dass für SZTE die Anzahl der Werte sowohl in 080\$a *Universelle Dezimalklassifikation* als auch in 650\$a *Sachschlagwort oder geografischer Name als Eintragungselement* relativ hoch ist (81,85% bzw. 57,82%). All das deutet auf eine sorgfältige Sacherschließung hin. Diese kann in der Finnischen Nationalbibliothek und unserer Erfahrung nach in der NSZL beobachtet werden.

Dagegen enttäuschen die hohe Anzahl unverständlicher Unterfelder im Falle von SZTE und MOKKA, und dafür gibt es sicherlich nachvollziehbarere Erklärungen als Unaufmerksamkeit und Zufall.

#### 4.4 Feld 651 Schlagworteintragung – Geografischer Name

Bei der Suche werden Sachschlagwörter oft mit geografischen Schlagwörtern ergänzt.

Was wir zu Feld 650 beschrieben haben, trifft auch auf Feld 651 grob zu. Jedoch besteht weder für MTAK noch für SZTE Konsistenz zwischen \$2 *Quelle der Ansetzung oder des Terms* und \$a *Geografischer Name*. Bzgl. NSZL können wir bekräftigen, dass sich der Gebrauch von *Köztaurusz* auf die Felder 648, 650 und 651 gleichermaßen auswirkt, da für jedes dieser Felder Schlagwörter aus *Köztaurusz* verpflichtend sind.

Das bei HUNMARC vorkommende Unterfeld \$c *Zusatzelement* dient der Unterscheidung zwischen homonymen Deskriptoren und homonymen geografischen Namen (was in MARC21 nicht vorgesehen ist). Das Unterfeld wird in der NSZL verwendet.

Tab. 12: Verwendung von Feld 651 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
651 Schlagworteintragung – Geografischer Name										
\$a Geografischer Name	1 114 929	5,86	84 056	4,76	7 325	0,70	50 893	4,01	882 109	9,92
\$b Geografischer Name einzelner Ortsangabe folgend (nicht mehr gültig)	–	–	4	0,00	4	0,00	8	0,00	4 190	0,05
\$e Funktionsbezeichnung	–	–	–	–	4	0,00	–	–	1	0,00
\$g Sonstige Informationen	78 322	0,41	19	0,00	124	0,01	–	–	1	0,00
\$0 IDN des Normdatensatzes	984 088	5,17	12 199	0,69	–	–	–	–	12	0,00
\$1 Persistenter Link zum Datensatz	–	–	1	0,00	–	–	–	–	8	0,00
\$2 Quelle der Ansetzung oder des Terms	1 114 929	5,86	84 056	4,76	1	0,00	35 122	2,77	68 448	0,77
\$3 Spezifizierte Materialien	–	–	–	–	–	–	113	0,01	1 563	0,02
\$4 Funktionsbezeichnung (Code)	–	–	–	–	–	–	–	–	64	0,00
\$6 Verlinkung	–	–	57	0,00	–	–	–	–	1	0,00
\$8 Feldverknüpfung und Reihenfolge	41 490	0,22	–	–	–	–	–	–	7	0,00
unbekannte Elemente	–	–	–	–	–	–	–	–	–	–
\$5 Institution, die dieses Feld verwendet	–	–	–	–	–	–	–	–	3	0,00

Fortsetzung **Tab. 12:** Verwendung von Feld 651 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
\$7	-	-	-	-	-	-	-	-	1	0,00
\$9	728	0,00	-	-	-	-	-	-	-	-
\$@	-	-	-	-	-	-	-	-	2	0,00
\$M	-	-	-	-	-	-	-	-	2	0,00
\$c Zusatzelement	-	-	-	-	6	0,00	2	0,00	13 244	0,15
\$d	-	-	-	-	2	0,00	2	0,00	19	0,00
\$f	-	-	-	-	-	-	1	0,00	1	0,00
\$h	-	-	-	-	-	-	-	-	1	0,00
\$i	-	-	-	-	-	-	1	0,00	2	0,00
\$k	-	-	1	0,00	-	-	1	0,00	8	0,00
\$l	-	-	-	-	-	-	-	-	4	0,00
\$m	-	-	-	-	-	-	3	0,00	5	0,00
\$n	-	-	-	-	1	0,00	-	-	-	-
\$p	-	-	-	-	-	-	-	-	1	0,00
\$s	-	-	-	-	-	-	-	-	2	0,00
\$t	-	-	-	-	-	-	-	-	23	0,00

## 4.5 Feld 653 Indexierungsterm – nicht normiert

In diesem Feld können Indexierungsterme frei hinzugefügt werden ohne jegliche terminologische Kontrolle und uneinheitlich innerhalb des Katalogs. Mit anderen Worten, Schlagwörter eines kontrollierten Vokabulars, wie z. B. bei der NSZL dem Köztaurusz, dürfen nicht in 653\$a *Indexierungsterm – nicht normiert* erfasst werden. In der Praxis erfassen Katalogisierer:innen in diesem Feld natürlich Deskriptoren auf Grundlage irgendeines Handbuchs, vielleicht einem Verschlagwortungsschema, aber die Verwendung einer bestimmten Quelle ist extrem beliebig, die Quelle nicht wirklich entscheidend. Daher hat dieses Feld nur ein wichtiges Unterfeld, \$a.

Tab. 13: Verwendung von Feld 653 im Vergleich

MARC-21-Da- tenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
653 Indexierungsterm – nicht normiert										
\$a Nicht-normierter Term	6 116 045	32,14	–	–	68 550	6,51	22	0,01	331 743	3,73
\$6 Verlinkung	–	–	–	–	307	0,03	–	–	–	–
undefinierte Felder										
\$0 IDN des Normdatensatzes*	–	–	–	–	–	–	4	0,00	2	0,00
\$2 Quelle	–	–	–	–	–	–	4	0,00	5	0,00
\$3	–	–	–	–	1	0,00	–	–	15	0,00
\$A	–	–	–	–	1	0,00	–	–	1	0,00
\$c	–	–	–	–	1	0,00	–	–	1	0,00
\$d	–	–	–	–	1	0,00	–	–	2	0,00
\$j	–	–	–	–	3	0,00	–	–	–	–
\$l	–	–	–	–	1	0,00	–	–	2	0,00
\$m	–	–	–	–	1	0,00	–	–	1	0,00
\$n	–	–	–	–	1	0,00	–	–	1	0,00
\$t	–	–	–	–	1	0,00	–	–	1	0,00
\$x	–	–	–	–	1	0,00	–	–	344	0,00
\$y	–	–	–	–	1	0,00	–	–	2	0,00
\$z	–	–	–	–	1	0,00	–	–	5	0,00

\* Wie können Identifier von Normdatensätzen und Quelle (z. B. kontrolliertes Vokabular, Verschlagwortungsschema usw.) für einen unkontrollierten Ausdruck angegeben werden?



Angesichts der vorangegangenen Bemerkungen ist es unverständlich, dass ein solches Element einen Identifier einer Normdatei (da es keine Normdaten sind), eine Quellenangabe usw. haben kann – ganz zu schweigen von den weiteren unverständlichen Unterfeldern. Wenn ein Deskriptor einem Schlagwortschema entnommen wird, welches von einer anderen Institution gepflegt wird, dann muss er in den Feldern 650, 651 oder 655 erfasst werden, wobei die Quelle im Unterfeld \$2 *Quelle* gekennzeichnet werden muss. Natürlich könnten solche thematischen Angaben dann auch ein Unterfeld \$0 *IDN des Normdatensatzes* haben.

#### 4.6 Feld 655 Indexierungsterm – Genre/Formschlagwort

In Feld 655 werden Indexierungsterme erfasst, die das Genre, die Art und die Form von Dokumenten beschreiben. Somit können die Werte in Feld 008 eine wichtige, wenn nicht sogar eine noch wichtigere Bereicherung der inhaltlichen Suche sein. Zweifelsohne sind die Codewörter im Feld 008 sprachunabhängig codiert, jedoch einerseits sehr allgemein und andererseits aus klassifikatorischer Sicht milde gesagt sehr oberflächlich zusammengebastelt. Demgegenüber können im Feld 655 ganz spezielle Formschlagwörter bzw. -angaben erfasst werden (Einführung, Konferenzschrift, usw.).

Es gibt Ausnahmen, darunter eine bemerkenswerte: In der LoC wurden über eine lange Zeit mit formalen Aspekten verbundene Deskriptoren nicht von thematischen unterschieden – vielleicht gerade mit der Begründung, sie seien gleich wichtig.

Die Feldverwendung zeigt ein ausgeglicheneres Bild; nur bei MTAK – wie wir es bei Unterfeld \$2 *Quelle des Terms* im Zusammenhang mit anderen 6XX-Feldern bereits beobachtet haben – stimmen in Feld 655 die Werte von \$a *Genre/Form oder fokussierter Term* und \$2 *Quelle* wieder nicht überein.

Tab. 14: Verwendung von Feld 655 im Vergleich

MARC-21-Datenelement	DNB	%	GENT	%	MTAK	%	SZTE	%	MOKKA	%
655 Indexierungsterm – Genre/Formschlagwort										
\$a Genre/Form oder fokussierter Term	5 098 603	26,80	197 768	11,21	19 226	1,83	1 493	0,12	595 078	6,69
\$b Nicht-Fokusbegriff	-	-	3	0,00	-	-	-	-	1	0,00
\$c Facette/Hierarchische Bezeichnung	-	-	4	0,00	-	-	-	-	4	0,00
\$0 IDN des Normdatensatzes	5 087 226	16,74	15 908	0,85	-	-	-	-	6	0,00
\$2 Quelle des Terms	5 098 459	26,80	197 768	11,21	62	0,01	1 493	0,12	367 537	4,13
\$3 Spezifizierte Materialien	-	-	-	-	-	-	-	-	1	0,00
undefinierte Unterfelder										
\$4	-	-	-	-	1	0,00	-	-	1	0,00
\$i	-	-	-	-	-	-	-	-	12	0,00
\$s	-	-	-	-	-	-	-	-	10	0,00
\$t	-	-	-	-	-	-	-	-	4	0,00

## 5 Zusammenfassung

### 5.1 Katalogisierungscharakteristika

#### 5.1.1 Verwendung von Feldern und Unterfeldern

Es gibt Felder wie 045 *Zeitabschnitt der Eintragung*, 046 *Speziell kodierte Datumsangaben* und 648 *Schlagworteintragung – Zeitschlagwort*, die entweder gar nicht oder nur in sehr geringem Umfang verwendet werden. Eine Überprüfung dieser Praxis würde sich lohnen.

Falsche Unterfelder bei 653 *Indexierungsterm – nicht normiert* zeigen vielleicht an, dass es Missverständnisse im Hinblick auf die Verwendung dieser Deskriptoren gibt. Solche Deskriptoren stammen im Prinzip nicht aus einer Quelle – denn wenn es doch so wäre, sollten sie in einem anderen für Schlagworteintragungen vorgesehenen Feld (und mit Angabe der jeweiligen Quelle) erfasst werden.

Die Verwendung der Unterfelder \$0 *IDN des Normdatensatzes* und \$1 *Persistenter Link zum Datensatz* wird künftig immer wichtiger werden. Die einzelnen Kataloge unterscheiden sich in etlichen Fällen recht deutlich darin, ob und in welcher Menge bestimmte Unterfelder vorkommen. Die Gründe dafür sind am besten den Expert:innen der jeweiligen Bibliotheken bekannt.

#### 5.1.2 Unterfeldprobleme

Die Gründe für das Auftreten von Problemen mit Unterfeldern werden unten in Abschnitt 5.2 zur Qualität von Katalogisierungssystemen diskutiert. Wir möchten hier erwähnen, dass bei MTAK beim Umstieg von HUNMARC zu MARC 21 sehr sorgfältig gearbeitet wurde und dazu ein eindrucksvolles Dokument erstellt wurde (Bilicsi 2018). Den Expert:innen ist es sicherlich nachvollziehbar, warum trotzdem noch falsche Unterfelder und falsche Werte in 008 vorkommen.

#### 5.1.3 Vollständigkeit der Felder und Unterfelder

In manchen Katalogen ist der Vollständigkeitsgrad bei Feldern und Unterfeldern auffällig uneinheitlich. Die Unterschiede lassen sich nicht getrennt von der etablierten Praxis und den Traditionen der Länder und der einzelnen Bibliotheken betrachten. Selbst in anscheinend präzisen formalisierten Fällen, wie

der Praxis der Normierung von Personennamen, kann es zu ziemlich großen Unterschieden kommen. Davon kann man sich leicht überzeugen, indem man in VIAF, wo bislang nur Personennamen verzeichnet werden, die Ansetzungsformen für Victor Hugo in den einzelnen nationalen Systemen vergleicht.<sup>13</sup>

Tatsächlich hat das Problem auch viel mit den Eigenschaften der jeweiligen Landessprache zu tun. Dies zeigt sich besonders deutlich bei den Benennungen der Begriffe für die Sacherschließung. Der inhaltliche Umfang der Benennungen ist nicht in allen Sprachen genau gleich. Insofern wird ihre Harmonisierung in mehrsprachigen Suchmaschinen der Zukunft also eine schwierige Aufgabe sein.

#### **5.1.4 Kohärenz verwandter Unterfeldern**

Die Gründe für Fehler in diesem Bereich werden im Abschnitt zur Qualität von Katalogisierungssystemen diskutiert.

#### **5.1.5 Verwendung lokal definierter Unterfelder**

Die Seltenheit lokal definierter Unterfelder ist auffällig. Außerdem ist pro Feld immer nur ein lokales \$9-Unterfeld zulässig, obwohl oftmals mehr als eins gebraucht wird. Bei einem Datenaustausch sollten sie außer Acht gelassen werden. Dies wird dadurch erschwert, dass außer \$9 weitere lokale Unterfelder definiert wurden (z. B. \$7), deren Verwendung zudem nicht öffentlich gemacht wird.

#### **5.1.6 Felder und Unterfelder, die in MARC 21 und in Versionen von MARC nicht definiert sind**

Nicht-definierte Felder und Unterfelder kommen relativ häufig vor. Viele sind wahrscheinlich durch Fehler entstanden, die sich durch eine Katalogisierungssoftware mit Funktionalitäten zur Qualitätsprüfung vermeiden ließen. Die kontinuierliche binäre Anzeige von MARC 21 ist ebenfalls ein IT-Problem, da ein falsches Unterfeldzeichen erzeugt werden kann, wenn ein Zeichen beim Lesen während der Konvertierung verrutscht, oder z. B. ein \$-Zeichen im Text vorkommt.

---

<sup>13</sup> <http://viaf.org/viaf/9847974/> (27.12.2020).

Es ist aber auch möglich, dass es sich bei einem kleinen Teil davon tatsächlich um lokale Unterfelder handelt. Lösungen für diese noch verborgenen Bedarfe zu finden, ist eine Aufgabe für die Weiterentwicklung von MARC 21. Die Veröffentlichungen der Dokumentation lokal definierter Datenelemente könnte die Interpretation der Kataloge verbessern.

## 5.2 Qualität des jeweiligen Katalogisierungssystems

Die meisten und die größten Probleme haben ihre Ursachen in der jeweils eingesetzten Software. Einige Datenelemente, die Feldnummern und Unterfeldcodes sowie die vorgeschriebenen Ein-Zeichen-Indikatoren können formal und weitestgehend vollständig überprüft werden. Auch sollten vollständige Beschreibungen der lokal definierten Felder und Unterfelder in der Software nachgehalten werden. Die Software sollte es autorisierten Katalogisierer:innen erlauben, neue standardisierte Felder, Unterfelder, Indikatoren und Berechtigungen einzufügen – ohne dass dafür weitreichende IT-Kenntnisse benötigt werden, da sich das MARC-21-Format ändert. Wie inhaltsbezogene Datenelemente aus Kontrollfeld 008 und aus der Feldergruppe 01X–09X in Kombination mit Schlagwörtern der Feldergruppe 6XX für Nutzer:innen bei der Suche ausgewertet werden können, ist eine bislang kaum beantwortete Frage. Dasselbe trifft auf die Datenelemente aus 008 in Abhängigkeit vom Dokumententyp zu. Ebenso verhält es sich mit der Kunstsprache der Notationen von Klassifikationssystemen (UDK, DDC, usw.): Die Suche danach ist nicht mit der natürlich-sprachigen Suche nach verbalen Deskriptoren koordiniert. Das heißt, dass ein Abgleich der von Menschen lesbaren Benennungen der Notationen der UDK, der DDC usw. mit verbalen Deskriptoren fehlt. Wichtig ist, dass etliche dieser Klassifikationen nicht als Open-Access-Ressourcen verfügbar sind, was die Entwickler:innen von Open-Source-Software daran hindert, ihre Werkzeuge in dieser Hinsicht zu verbessern. Einige Katalogisierungssysteme sind nicht so beschaffen, dass die Stimmigkeit zwischen verwandten Feldern und Unterfeldern adäquat geprüft werden kann (beispielsweise zwischen den Werten in 008/35-37 *Sprache* und 041\$a *Sprache*). Die Aufzählung ist bei weitem nicht vollständig.

## 5.3 Vergleich ausgewählter Kataloge

Es zeigt sich, dass die Anzahl undefinierter Unterfelder bei ungarischen Bibliotheken am höchsten ist. Der Hauptgrund dafür ist nicht eine weniger solide Ausbildung der ungarischen Expert:innen, sondern vielmehr die im vorherigen

Abschnitt diskutierten Defizite der Katalogisierungssoftware, und last but not least, wie wir meinen, die finanzielle Situation.

## 6 Ein allgemeiner Ausblick für MARC 21

Jason Thomale (2010) und anschließend Karen Coyle (2011) haben ihrem Verständnis nach in kritischen und in die Tiefe gehenden Analysen von MARC 21 gezeigt, dass es einige Merkmale des Standards gibt, die für jede Form automatisierter Weiterverwendung von Katalogdatensätzen Probleme verursachen. Wir können nicht so ins Detail und in die Tiefe gehen, wie das die beiden genannten Arbeiten charakterisiert. Wir wollen nur einen, den auffälligsten, Atavismus erwähnen, der eigentlich mittlerweile völlig überflüssig geworden ist: In wichtigen Datenfeldern wie den Titelfeldern werden Interpunktionszeichen noch intellektuell erfasst!

Interpunktion bei der physischen Beschreibung:

300\$a 303 pages : \$b illustrations (chiefly color) ; \$c 26 cm

Interpunktion bei historischen Einzelereignissen (hier: Bunker Hill Battle near Boston):

647\$a Bunker Hill, Battle of \$c (Boston, Massachusetts : \$d1775) \$2 fast \$0 (OCoLC) fst01710024

Bei der Anzeige im ISBD-Format werden natürlich die Interpunktionszeichen aus den MARC-Feldern und -Unterfeldern verwendet; deren Übertragung in getrennte Unterfelder ist also formal gewährleistet. In 040\$e können Katalogisierer:innen Beschreibungskonventionen eingeben, woraus (im Prinzip) hervorgeht, ob mit Interpunktionszeichen gerechnet werden muss oder nicht. Eine andere Frage ist aber, wie sorgfältig dieses Datenelement erfasst wird. Beispielsweise werden in SZTE Interpunktionszeichen verwendet.

Es ist problematisch, dass für das Merkmal *Form der Formalerschließung* (Position 18 im *Leader*) nur eine begrenzte Menge von erlaubten Werten definiert wurde (*non-ISBD*, *AACR 2*, *ISBD*) und zusätzliche Werte (*ISBD Interpunktion ausgelassen*, *ISBD Interpunktion einbezogen*, *nicht-ISBD Interpunktion ausgelassen*, *unbekannt*), die mehr Flexibilität bringen, erst in den letzten zehn Jahren eingeführt wurden und normalerweise auf ältere Titelaufnahmen nicht angewandt worden sind. Anhand von Position 18 im *Leader* und 040\$e sollten Programmierer:innen erkennen können, was bei der Weiterverwendung der Daten zu erwarten ist, also ob mit Interpunktionszeichen zu rechnen ist oder nicht und falls ja, wo.

Wenn Programmierer:innen ohne Bibliothekserfahrung versuchen, MARC-Daten formal zu bearbeiten, werden sie sofort auf dieses Problem stoßen. Es ist besonders verwirrend, wenn die Daten außerhalb ihres Ursprungskontexts, z. B. eines Bibliothekssystems, gebraucht werden (etwa in der LOD-Cloud, in Wikidata, bei Zitierformaten, in Europeana usw.). Sogar im Fall von MOKKA (oder von anderen Verbundkatalogen) stellt sich die Frage, ob bei der Anzeige einer Titelaufnahme Position 18 des *Leader* berücksichtigt wird, bei dem die Regeln streng einzuhalten sind. Dabei ist nicht einmal klar, ob die Eingaben überhaupt zutreffen und die Aufnahme tatsächlich der angegebenen Form der Formalerschließung folgt.

Die Interpunktionspraxis, deren Geschichte in angelsächsische „Urzeiten“ zurückreicht, steht eigentlich im Widerspruch zum zentralen MARC-Prinzip, alle Datenelemente separat im Rahmen der Feld-Unterefeld-Indikator-Struktur zu behandeln. Dass es zu dieser Praxis überhaupt kommen konnte, lässt vermuten, dass Bibliothekar:innen selbst das Prinzip nicht wirklich verstanden haben. Henriette Avram (1975) erwähnt in ihrem Werk über die frühe Geschichte des MARC-Formats, dass es ursprünglich darum ging, Katalogkarten zu bedrucken. Dem würden wir hinzufügen, dass, wie üblich, die Praxis die Prinzipien überlagert hat und spätestens als die ersten Online-Kataloge auf den Bildschirmen erschienen, diese Praxis zum Standard wurde.

Mit HUNMARC wurde diese Praxis in den 1990ern vollständig aufgegeben. Darüber hinaus hat es in neuerer Zeit weitere starke Abschaffungsbemühungen gegeben (Király und Büchler 2018). MARC 21 wird immer noch weiterentwickelt und so sind im letzten Jahrzehnt eine Reihe neuer Datenelemente eingeführt worden (Ungváry 2011). Hinsichtlich seiner Möglichkeiten zur Datenhaltung ist es ein wunderbares Format. Diese Tatsache wird wahrscheinlich in Zukunft noch mehr geschätzt werden, wenn Titelaufnahmen mit reichhaltigen Daten genauere Suchen ermöglichen.

Aus einem anderen Problem ergibt sich eine größere Aufgabe: Das MARC-Format verfügt nicht über ein Qualitätssicherungssystem. Wir gehen hier darauf nicht weiter ein, obwohl es eine längere Diskussion wert wäre und es sich dabei nicht um ein unlösbares Problem handelt.

Wegen dieser wenigen Probleme wird oft die ganze Existenzberechtigung von MARC in Frage gestellt. Ein Beispiel dafür ist Roy Tennants Artikel *MARC must die* (Tennant 2012), in dem er schrieb: „The very nature of the MARC record is, to some degree, an anachronism.“ Obwohl Tennant einige stichhaltige Argumente anführt, halten wir sein Urteil für unberechtigt. Die Feld-Unterefeld-Indikator-Struktur ist eine klassische logische Struktur, die für IT-Lai:innen relativ leicht zu verstehen ist. Ein Compiler oder eine Schnittstelle können auf diese logische Struktur aufgesetzt werden, um MARC in eine für Programmierungs-

zwecke besser geeignete Struktur zu überführen. Letztlich sind die XML-Version bzw. die JSON-Version von MARC genau das.

Zudem meint Tennant (2012): „[M]etadata situations like this are completely indefensible in the world of the web.“ So sei MARC für gewisse (triviale und weniger triviale) Anwendungsfälle heute nicht geeignet und bräuchte dazu ein anderes Metadatenformat. Das ist genau das Schnittstellensystem, das die besagte logische Struktur von MARC umwandelt. Weder MARC 21 noch die Katalogisierer:innen haben damit etwas zu tun; es ist ein reines IT-Problem, wie wir meinen.

Übrigens wird schon seit zwanzig Jahren darüber nachgedacht, wie MARC durch Aufhebung der binären Form verbessert werden könnte. Die Veröffentlichungen von Coyle (2011) und Thomale (2010) belegen das. Aber wie ein zukünftiges MARC++ auch immer aussehen könnte, es muss immer noch ein logisches Text-Format haben, welches die anwendenden Bibliothekar:innen verstehen und im Rahmen ihres Nicht-IT-Berufs interpretieren können. Mit anderen Worten: Sie müssen die Datenelemente in ihren Zusammenhängen verstehen können, also auch die Beziehung zwischen übergeordneten und untergeordneten Datenelementen, wie sie sich aktuell in der logischen Feld-Unterfeld-Struktur von MARC 21 widerspiegelt. Dies ist das logische Niveau und Format, das Katalogisierer:innen in Zukunft brauchen werden. Und das lässt sich erreichen.

Die wirklich großen Probleme verursachen die Bibliotheksanwendungen. Selbst die exzellentesten Bibliotheken (Library of Congress, Harvard, Deutsche Nationalbibliothek usw.) arbeiten nicht fehlerfrei. Das ist eine schockierende Erfahrung. Unsere vergleichende Analyse zeigte, dass der Grund für die Probleme nicht an der Struktur von MARC liegen, sondern zum einen an den Schwächen seiner praktischen Anwendung, zum anderen – und das ist ein viel größeres Problem –, überwiegend an den Katalogisierungssystemen, die mit MARC arbeiten.

## 7 Ausblick

In diesem Beitrag haben wir uns mit einfachen Metriken zur Messung der Vollständigkeit beschäftigt. Es gibt andere komplexere Metriken, die wir hier nicht diskutiert haben, von denen wir aber zwei erwähnen wollen:

- Delsey (2002) hat ein Mapping zwischen zwölf in FRBR-Dokumenten definierten Funktionen (wie *search*, *identify*, *sort*, *display* usw.) und ungefähr 2000 MARC-Datenelementen erarbeitet. Ausgehend von den Faktoren der Vollständigkeit dieser Datenelemente und des Auftretens von Problemen in



ihnen (diskutiert in Abschnitt 2) können wir berechnen, wie ein Katalog diese Funktionen unterstützt.

- Es gibt Modelle, die Datenelemente je nach ihrer Wichtigkeit unter bestimmten Blickwinkeln unterschiedlich gewichten. Carlstone (2017) legte den Fokus darauf, die besten Aufnahmen für elektronische Reihen zu finden, Thompson und Traill (2017) die besten Titelaufnahmen für e-Books, während Booth (2020), ausgehend von einer Umfrage, bei der sie die Einschätzung von 50 Bibliotheken im Vereinten Königreich eingeholte, ein allgemeines Wichtigkeits-Ranking für Metadatenelemente aufstellte. Die Gewichtungsfaktoren können in derselben Weise wie das FRBR-Funktionsmapping dazu verwendet werden, eine komplexe Messgröße für einzelne Titelaufnahmen sowie für ihre Verteilung über einen bestimmten Katalog zu berechnen.

Wir können uns einen Katalog auch als Graph vorstellen, wobei übereinstimmende Werte (z. B. dasselbe Erscheinungsjahr) die Aufnahmen miteinander verbinden. Besonders spannend ist eine solche Vorstellung im Fall der Sacherschließung, die in der Regel mit kontrollierten Vokabularen oder anderen Wissensorganisationssystemen (*Knowledge Organisation Systems*, KOS) operiert. Es liegen KOS-Kategorisierungen vor, die sich meistens auf intrinsische Eigenschaften, z. B. auf die Komplexität ihrer Struktur, auf die Typen von Beziehungen zwischen den Deskriptoren, und auf ihre „Aussagestärke“ beziehen. In den letzten 20 Jahren hat die Netzwerkforschung eine solide Methodologie sowie Metriken geliefert, die für die Beschreibung und das Verständnis der Eigenschaften von Netzwerken hilfreich sind. Zuletzt stellen Wills (2017) und Phillips (2020) herausragende Beispiele für die Anwendung dieser Methodologie im Bibliothekswesen vor, ohne aber dabei MARC-21-Daten zu analysieren. Auf der anderen Seite hat die KOS-Community einige Vorschläge für die Bewertung kontrollierter Vokabulare zu bieten. Wir präsentieren (Király 2019b) einige Details, wie wir diese Forschungsrichtungen gern im Rahmen der Forschung zu Bibliothekskatalogen miteinander kombinieren würden.

Für einen allgemeinen Forschungsüberblick zu der Thematik empfehlen wir die *Metadata Assessment Bibliography*.<sup>14</sup>

---

14 [https://www.zotero.org/groups/488224/metadata\\_assessment](https://www.zotero.org/groups/488224/metadata_assessment) (25.12.2020).

## 8 Literaturverzeichnis

- Avram, Henriette D.: MARC, its history and implications. Washington: Library of Congress 1975. <https://babel.hathitrust.org/cgi/pt?id=mdp.39015034388556> (24.12.2020).
- Bilicsi, Erika: A MARC21 szerinti katalógizálás bevezetése az MTA könyvtár és információs központban. Budapest: Könyvtár és Információs Központ 2018. <https://doi.org/10.14755/MTAKIK.KOZL.2018.MARC21>.
- Booth, Emma: Quality of Shelf-Ready Metadata. Analysis of survey responses and recommendations for suppliers. Pontefract (UK): National Acquisitions Group, 2020. [https://nag.org.uk/wp-content/uploads/2020/06/NAG-Quality-of-Shelf-Ready-Metadata-Survey-Analysis-and-Recommendations\\_FINAL\\_June2020.pdf](https://nag.org.uk/wp-content/uploads/2020/06/NAG-Quality-of-Shelf-Ready-Metadata-Survey-Analysis-and-Recommendations_FINAL_June2020.pdf) (25.12.2020).
- Carlstone, Jamie: Scoring the Quality of E-Serials MARC Records Using Java. In: *Serials Review* (2017) Bd. 43 Nr. 3–4. S. 271–277. <https://doi.org/10.1080/00987913.2017.1350525>.
- Coyle, Karen: MARC21 as Data: A Start. In: *Code4Lib Journal* (2011) Bd. 14. <https://journal.code4lib.org/articles/5468> (24.12.2020).
- Delsey, Tom: Functional analysis of the MARC 21 bibliographic and holdings formats. Updated and Revised by the Network Development and MARC Standards Office Library of Congress: April 6, 2006. Washington DC: Library of Congress 2002. <https://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html> (25.12.2020).
- Harper, Corey et al.: Metadata assessment. Group Library. 2016–. [https://www.zotero.org/groups/488224/metadata\\_assessment](https://www.zotero.org/groups/488224/metadata_assessment) (25.12.2020).
- HUNMARC. A bibliográfiai rekordok adatcsere forkmátuma. KSZ 4/1. Budapest: Széchényi Könyvtár 2002. <https://ki.oszk.hu/sites/default/files/dokumentumtar/hunmarc.pdf> (24.12.2020).
- Internationale Organisation für Normung: ISO 3166-2:2020. Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code <https://www.iso.org/standard/63546.html> (25.12.2020).
- Király, Péter: QA catalogue – a metadata quality assessment tool for MARC records. Source code. 2017–. <https://github.com/pkiraly/metadata-qa-marc> (25.12.2020).
- Király, Péter: QA catalogue for analysing library data [Web interface]. <http://gent.qa-catalogue.eu/metadata-qa/>. (25.6.2021).
- Király, Péter und Marco Büchler: Measuring completeness as metadata quality metric in Europeana. In: *Proceedings: 2018 IEEE International Conference on Big Data (Big Data)*. Seattle, WA: 2018. S. 2711–2720. <https://doi.org/10.1109/BigData.2018.8622487>.
- Király, Péter (2019a). Validating 126 million MARC records. In: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATECH2019)*. New York, NY: Association for Computing Machinery 2019. S. 161–168. <https://doi.org/10.1145/3322905.3322929>.
- Király, Péter (2019b): Measuring subject term usage in bibliographic records. 2019. <http://pkiraly.github.io/2019/10/30/measuring-subject-term-usage-in-bibliographic-records/> (25.12.2020).
- Király, Péter: Empirical evaluation of library catalogues. In: *Europeana Tech* (2020) Nr. 15. <https://pro.europeana.eu/page/issue-15-swib-2019#empirical-evaluation-of-library-catalogues> (25.12.2020).

- Library of Congress: MARC21 Format for Bibliographic data. Washington DC: Library of Congress 1999 Edition, Update No. 1 (October 2000) through Update No. 31 (December 2020). <https://www.loc.gov/marc/bibliographic/> (25.12.2020).
- Phillips, Mark Edward: Exploring the use of metadata record graphs for metadata assessment. Dissertation. Texas: University of North Texas 2020. <https://digital.library.unt.edu/ark:/67531/metadc1707350/> (25.12.2020).
- PCC Standing Committee on Applications: Draft PCC Guidelines for Minimally Punctuated MARC Bibliographic Records. Prepared by PCC Standing Committee on Applications. Approved by PCC Policy Committee, Sept. 2019. <http://www.loc.gov/aba/pcc/documents/PCC-Guidelines-Minimally-Punctuated-MARC-Data-draft.docx> (24.12.2020).
- Szabó, Julianna: MOKKA inicializálás. Az OSZK rekordok szintaktikai ellenőrzése. Egykötetes és többkötetes közös adatokat leíró rekordok. 07.02.2003 [Manuskript].
- Tennant, Roy: MARC must die. In: *Library Journal* (2010). <https://www.libraryjournal.com/?detailStory=marc-must-die> (25.12.2020).
- Thomale, Jason: Interpreting MARC: Where's the Bibliographic Data? In: *Code4Lib Journal* (2010) Bd. 11. <https://journal.code4lib.org/articles/3832> (25.12.2020).
- Thompson, Kelly und Stacie Trill: Implementation of the scoring algorithm described in Leveraging Python to improve ebook metadata selection, ingest, and management. In *Code4Lib Journal* (2017) Bd. 38. <http://journal.code4lib.org/articles/12828> (25.12.2020).
- Ungváry, Rudolf: A besorolási adatsere-formátum bővülése. A legutóbbi két évtized fejleményei. [Erweiterung des Formats MARC-21 für Normdaten. Entwicklungen der letzten zwei Jahrzehnte] In: *Tudományos és Műszaki Tájékoztatás* (2011) Jg. 58 Nr. 9. S. 371–386. <https://tmt.omikk.bme.hu/tmt/article/view/868/10649> (25.12.2020).
- Ungváry, Rudolf: Besorolási, szabványosított, normatív vagy „autorizált“. [Normdaten, standardisierte Daten, oder „autorisierte Daten“?] In: *Tudományos és Műszaki Tájékoztatás* (2019) Jg. 66 Nr. 6. S. 328–342. <https://tmt.omikk.bme.hu/tmt/article/view/12309/14064> (25.12.2020).
- Ungváry, Rudolf: MARC21 tartalmi adatmezők használata jelentősebb nagykönyvtárakban. Egy elemzés tanulságai. [Verwendung von MARC-21-Inhaltsdatenfeldern in großen Bibliotheken. Lehren aus einer Analyse] Hungarnet, Networkshop 2020. szeptember 2–4. Pécsi Tudományegyetem, (Online) konferencia előadásainak közleményei. <http://real.mtak.hu/119192/1/ungvary.pdf> (13.1.2021).
- Ungváry, Rudolf: Ismeretszervező-könyvtári rendszerek tartalmi feltárásának összehasonlító vizsgálata MARC21 környezetben. [Vergleichende Untersuchung zur inhaltlichen Erforschung von Wissensorganisationssystemen in der MARC-21-Umgebung]. In: *Tudományos és Műszaki Tájékoztatás* (2020) Jg. 67. Nr. 11. S. 655–680. <https://tmt.omikk.bme.hu/tmt/article/view/12776> (13.1.2021).
- Wills, Josh: Analyzing co-occurrence networks with GraphX. In: Ryza, Sandy, Uri Laserson, Sean Owen and Josh Wills: *Advanced Analytics with Spark*. 2. Aufl. O'Reilly 2017. S. 141–171.

