

SECTION 3. INFORMATION AND COMMUNICATION TECHNOLOGIES IN EDUCATION

DOI: 10.46299/ISG.2024.MONO.PED.4.3.1

3.1 Специфікації переходу авторитетних записів із MARC21 до UNIMARC

Актуальність теми дослідження. У сучасному інформаційному суспільстві одним із ключових завдань бібліотек і інформаційних центрів є забезпечення ефективної організації та доступу до даних. Формати MARC21 та UNIMARC є провідними міжнародними стандартами для опису бібліографічної інформації. Проте, з огляду на різницю у філософії форматів, структурі полів та підполів, перехід даних між ними становить значні технічні та методологічні виклики. Актуальність цього дослідження зумовлена кількома факторами. Зі зростанням співпраці між бібліотеками різних країн важливим стає забезпечення інтероперабельності бібліографічних даних. Конвертація авторитетних записів між MARC21 та UNIMARC сприяє гармонізації стандартів та інтеграції національних бібліотечних систем у міжнародне інформаційне середовище.

Попри широке використання MARC21 у бібліотеках Північної Америки, UNIMARC залишається основним стандартом у багатьох європейських країнах. Розбіжності у структурі форматів, наприклад, відсутність у UNIMARC спеціальних підполів для URL або певних метаданих, вимагають пошуку ефективних підходів до збереження максимальної повноти даних під час конвертації. Впровадження цифрових технологій і зростання ролі електронних ресурсів посилюють необхідність розвитку форматів. Очікувані оновлення UNIMARC, такі як введення підполя \$R для зберігання URL, демонструють динамічність стандартів і потребу в адаптації бібліотечних систем до цих змін. Конвертація записів із максимально можливим збереженням даних зменшує витрати часу та ресурсів на їх ручну обробку, що особливо актуально для великих бібліотечних фондів. Таким чином, дослідження переходу авторитетних записів із MARC21 до UNIMARC є важливим для розвитку бібліотечної справи, підтримки міжнародної співпраці та вдосконалення інформаційного обміну.

Воно дозволяє не лише визначити поточні виклики та обмеження форматів, але й сприяє формуванню рекомендацій для майбутніх удосконалень стандартів.

Основною **метою** цього дослідження є розробка систематичного та практичного підходу до переходу авторитетних записів персональних імен із формату MARC21 до UNIMARC.

У рамках дослідження поставлено три ключові завдання:

1. **Розробка специфікації переходу** – створення детального зіставлення тегів та підполів MARC21 із відповідними елементами UNIMARC з акцентом на полях, що стосуються авторитетних записів персональних імен.

2. **Оцінка охоплення переходу** – аналіз ступеня відповідності елементів авторитетних записів MARC21 формату UNIMARC з використанням статистики використання, отриманої з авторитетних записів Бібліотеки Конгресу США.

3. **Вивчення відмінностей форматів** – дослідження випадків, що вимагають ручного втручання, та причин цього.

Аналіз літературних джерел. Для аналізу теми переходу авторитетних записів із MARC21 до UNIMARC важливо розглянути кілька основних напрямів літератури, які висвітлюють історію розвитку форматів MARC, їхню функціональність, стандартизацію та особливості конвертації між ними. Основні теоретичні та практичні аспекти MARC були закладені у роботах, які висвітлюють історію його розробки та впровадження, зокрема MARC21 (Machine-Readable Cataloging) та UNIMARC. Праці, присвячені MARC21, розглядають його широке використання у Північній Америці та за її межами, акцентуючи на деталізації підполів і можливості адаптації до сучасних технологій (наприклад, роботи Бібліотеки Конгресу США). Література з UNIMARC більше зосереджена на його міжнародному застосуванні, зокрема у Європі. Значну роль у стандартизації відіграють документи, підготовлені IFLA, які містять специфікації UNIMARC. Існує низка праць, присвячених відмінностям між MARC21 та UNIMARC, що мають вирішальне значення для розробки методів конвертації. Наприклад, роботи Рендалла К. Баррі (Randall K. Barry) описують співвідношення між форматами та пропонують приклади

їхнього узгодження. Офіційні рекомендації Бібліотеки Конгресу США щодо перетворення MARC21 у UNIMARC також є важливим джерелом для аналізу механізмів конвертації. Більшість попередніх досліджень зосереджені на конвертації бібліографічних записів, у той час як авторитетні записи отримали менше уваги. Однією з ключових праць у цій сфері є семінари та презентації, що стосуються конвертації даних на основі реальних бібліотечних записів (наприклад, використання MarcEdit). Аналіз досліджень, які оцінюють використання специфікацій IFLA для UNIMARC, демонструє, як вирішуються практичні питання, пов'язані з розбіжностями у структурах форматів.

У літературі наголошується на важливості інтероперабельності між системами та стандартизації, що дозволяє обмінюватися даними на міжнародному рівні. Наприклад, дослідження Європейської комісії з питань гармонізації бібліографічних стандартів висвітлюють практичні труднощі, які виникають через відсутність єдиного підходу до деяких полів (таких як зберігання URL у MARC21, але не в UNIMARC). Інструменти, як-от MarcEdit, слугують базою для практичної реалізації конвертації між форматами. Опис їхнього функціонала та приклади використання є важливою частиною літературного аналізу.

Проблеми, визначені в літературі полягають у відсутності однозначного відповідника для деяких підполів у UNIMARC, наприклад, для URL чи певних метаданих, таких як гендерна інформація. Необхідність ручного доопрацювання для деяких полів через відмінності у філософії форматів. Складність переходу для великих баз даних, що потребує мінімізації втрат даних і зменшення ручної роботи. Аналіз літератури демонструє, що тема переходу авторитетних записів між MARC21 та UNIMARC є недостатньо дослідженою, особливо у контексті авторитетних записів. Це створює підґрунтя для нових досліджень, спрямованих на розробку ефективних підходів до конвертації та вдосконалення існуючих стандартів.

Основні результати дослідження. Формати MARC (Machine-Readable Cataloging) тривалий час залишаються домінуючим стандартом для кодування

бібліографічних та авторитетних даних. MARC21, який широко використовується у Північній Америці та інших регіонах, і UNIMARC, поширений у Європі та деяких міжнародних контекстах, є двома основними варіантами цього стандарту. Хоча обидва формати мають схожі цілі, вони відображають різні філософії та організаційні структури, що часто призводить до суттєвих розбіжностей у визначенні полів, використанні підполів та логіці кодування.

З огляду на те, що бібліотеки все частіше прагнуть обмінюватися даними між системами та національними межами, необхідність у розробці ефективних методів переходу між MARC21 і UNIMARC зростає. Такі переходи є критично важливими для створення об'єднаних каталогів, посилення співпраці та забезпечення актуальності бібліографічних даних у глобальному контексті.

Попередні дослідження з переходу між форматами MARC здебільшого зосереджувалися на бібліографічних записах. У 1984 році Бібліотека Конгресу підготувала програму для переходу бібліографічних записів із MARC21 у UNIMARC. Попри ці зусилля, сервіс залучив лише двох клієнтів, що призвело до його припинення у 1995 році. Наразі Бібліотека Конгресу більше не підтримує програми переходу з MARC21 у UNIMARC, але надає рекомендації зі зіставлення на своєму вебсайті. Ці рекомендації залишаються одним із небагатьох офіційних джерел для розуміння зв'язків між форматами.

Більшість досліджень і інструментів зосереджуються на бібліографічних записах, а не на авторитетних. Значний внесок зробив семінар Рендала К. Баррі для студентів Бібліотечно-інформаційного центру Університету Македонії, який пропонує приблизне зіставлення між форматами. Це дослідження базується на таких фундаментальних працях, щоб створити детальну специфікацію переходу оновлених форматів авторитетних записів MARC.

Для аналізу було відібрано 179 авторитетних записів MARC21, які містять поле 100. Вибірка включала 16 записів персональних імен та 162 записи імен/заголовків. Ці записи отримані за допомогою пошукових протоколів Z39.50, а автори представляють як сучасні, так і історичні періоди. Для аналізу

статистики використання підполів використовувалося програмне забезпечення MarcEdit 7.5, що забезпечило детальний розгляд структур записів і їхнього змісту.

Для порівняння форматів та розробки специфікації переходу використовувався багатоджерельний підхід, включаючи ручне порівняння документації стандартів MARC21 і UNIMARC, специфікацій переходу Бібліотека Конгресу та семінару Рендала К. Баррі «Взаємозв'язок формату MARC21 Authority з UNIMARC». Цей всебічний огляд дозволив створити детальну таблицю зіставлення полів та підполів між двома форматами з особливою увагою до складних переходів і потенційних викликів.

Специфікація переходу була розроблена з пріоритетом на повноту та зменшення необхідності ручного втручання. Акцент зроблено на перетворенні основних даних, критичних для ідентифікації та опису персональних імен. Додаткова інформація та контрольні поля отримали нижчий пріоритет.

Нові поля чи підполя поза стандартами MARC21 або UNIMARC не створювалися. Підполя, які не були конвертовані, виділені курсивом у таблиці переходу для чіткого позначення прогалин і полегшення подальших досліджень.

Якщо прямі відповідники були недоступні, дані переносилися з альтернативних полів або об'єднувалися з кількох підполів, якщо це було можливим.

Перехід авторитетних записів із MARC21 до UNIMARC вимагає врахування різниці у структурі форматів, особливостей кодування даних, а також обмежень кожного стандарту. Аналіз специфікацій демонструє кілька ключових аспектів, важливих для ефективного перетворення даних:

1. Перетворення набору символів

UNIMARC не підтримує MARC-8, що робить обов'язковим перетворення даних у формат UCS/Unicode перед початком переходу.

– MARC-8 – це застарілий набір символів, що використовувався в MARC21, який має обмежені можливості щодо відображення нелатинських алфавітів.

- Unicode забезпечує більшу гнучкість у роботі з різними мовами та сценаріями, що є критичним для міжнародних записів UNIMARC.

- Важливим завданням є забезпечення збереження інформації при конвертації, зокрема для складних символів або знаків, що мають неоднозначні відповідники.

2. Заголовки MARC21 та мітки UNIMARC

Обидва формати містять заголовок запису, проте представлені вони по-різному:

- У MARC21 заголовок кодується в полі 000, яке містить інформацію про статус запису, тип запису, тип організації, що його створила, та інші метадані.

- В UNIMARC заголовок також представлений у полі 000, однак структура і значення позицій символів можуть суттєво відрізнятися.

Специфікація переходу повинна враховувати:

- Правильне зіставлення кожної позиції символу в MARC21 з відповідним значенням у UNIMARC.

- Конвертацію кодів статусу запису, типу організації тощо, особливо для значень, які не мають прямого відповідника.

3. Позначення елементів запису

Для уніфікації під час конвертації специфікація використовує стандартизовані символи:

- **Знак долара (\$):** використовується для позначення підполів у записі. Наприклад, у полі 100\$a MARC21 позначає основне поле і його підполе.

- **Коса риска (/):** використовується для звернення до конкретних позицій символів у заголовках запису. Наприклад, 000/05 означає п'яту позицію у полі 000.

- **Решітка (#):** застосовується для позначення пробілів у полях. Це є важливим для дотримання стандартизованого формату, де пробіли мають визначене значення.

4. Розбіжності у підтримці даних

– **Підтримка міток:** MARC21 має мітки записів, які інколи не мають прямого відповідника у UNIMARC, що вимагає ручної адаптації.

– **Типи записів:** MARC21 використовує коди для статусу запису, наприклад, «видалений» чи «оброблений». UNIMARC не підтримує статус «obsolete», що потребує адаптації значень при переході.

5. Обмеження та виклики

– Відсутність підтримки певних специфічних полів у UNIMARC (наприклад, URL у MARC21).

– Неоднозначність у значенні деяких символів або кодів, що потребує уточнення в специфікаціях або подальших досліджень.

– Вимога дотримання стандартів інтероперабельності, що обмежує створення нових полів або підполів.

Специфікація переходу авторитетних записів із MARC21 до UNIMARC повинна враховувати ключові аспекти: правильне перетворення наборів символів, зіставлення полів та підполів, а також адаптацію для розбіжностей у структурі записів. Перехідний процес є складним, але дотримання стандартизованих процедур забезпечує мінімальні втрати даних і збереження цілісності записів.

34 теги MARC21 та 213 підполів успішно зіставлено з 39 тегами UNIMARC та 193 підполями (включаючи компоненти контрольних полів) (див. таблиця 1).

Таблиця 1.

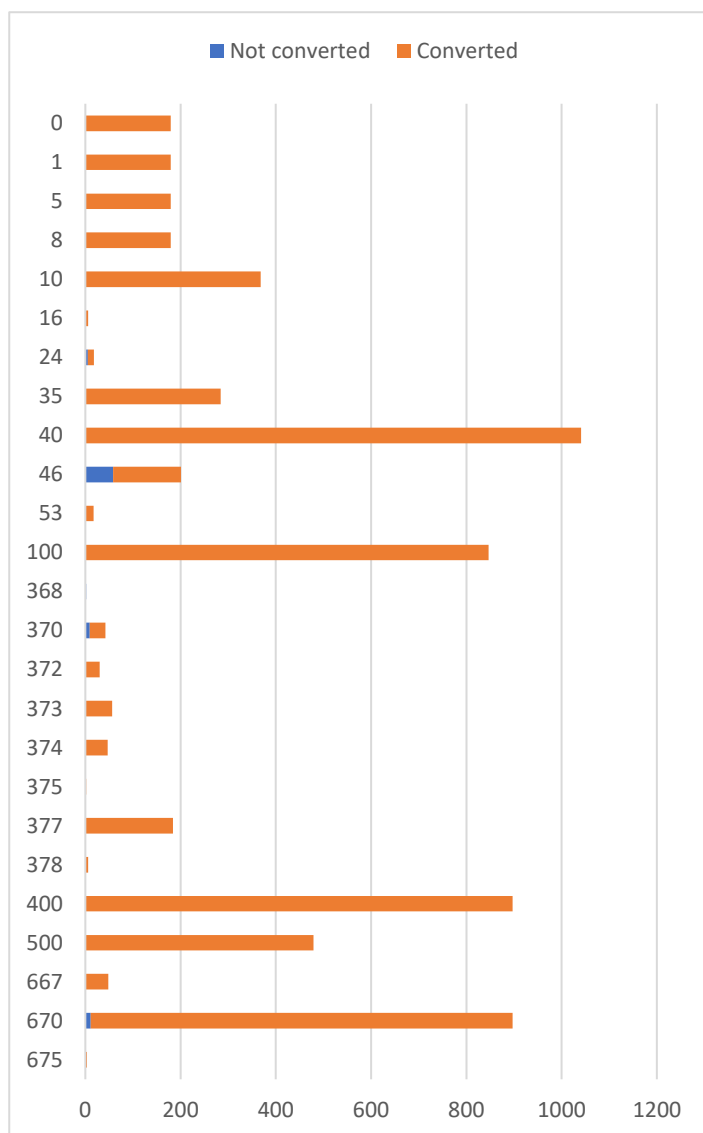
Зразок переходу із формату **UNIMARC** в **MARC21**

UNIMARC			MARC21		
Block	Tags	Subfields	Block	Tags	Subfields
0	3	19	0	14	68
1	8	21	1	1	26
2	3	26	2	0	0
3	6	13	3	6	26
4	3	26	4	1	26
5	3	26	5	1	26
6	5	25	6	10	15
7	3	26	7	1	26
8	5	11	8	0	0
Total:	39	193		34	213

Перехід був зосереджений на забезпеченні максимально повного зіставлення авторитетних записів персональних імен, а також заголовків із прізвищами та заголовків ім'я/заголовок у авторитетних записах, оскільки вони використовують однаковий тег у MARC21. Оскільки поля, що стосуються заголовків або інших типів записів, не входили до сфери цього дослідження, наступні поля, присутні у вибірці, були відсутні у специфікації переходу: 050, 336, 380, 381, 382, 430, 510, 630, 64X. Також не конвертувалися поля локального використання 9XX. Інші не конвертовані поля включають: ідентифікатор контрольного номера (003), інші атрибути особи чи організації (368) та код аутентифікації (042). Поле 368 містило титул особи, який уже використовувався у заголовку. Відповідного поля для титулу в UNIMARC знайдено не було (див. таблиця 2).

Таблиця 2.

Тестування ефективності специфікації переходу в MarcEdit



Tags	Not converted		Converted
	Fields	Count	
000			179
001			179
005			179
008			179
010			368
016			6
024	\$1	5	13
035			284
040			1041
046	\$v, \$2	58	143
053			17
100			847
368	\$d	2	
370	\$2	9	33
372			30
373			56
374			47
375			2
377			184
378			6
400			897
500			479
667			48
670	\$u	11	886
675			3
Total:		85	6106

24\$1 і 670\$u використовуються для URL. На цей момент в UNIMARC немає відповідних підполів, але, згідно з 6-ю Зустріччю користувачів UNIMARC, підполе \$R може бути введено пізніше цього року для зберігання URL.

46\$v (джерело інформації) потенційно може бути конвертоване у якесь поле для текстових приміток.

46\$2 (джерело схеми дат) може бути використане для кращої конвертації дат. Потрібне додаткове дослідження.

UNIMARC більше використовується в міжнародному контексті, і він має більше підполів для опису мови або сценарію (100\$a/21-22), а також навіть напрямую написання (100\$a/23). Оскільки відповідні дані у форматі MARC21 не знайдені, мова і сценарій були визначені на основі мови каталогізації, де це можливо.

Поле 700 у UNIMARC призначене для авторизованої точки доступу іншою мовою та/або сценарієм. Поле 700 у MARC21 може містити еквівалентні імена у багатомовному тезаурусі, але не лише їх, оскільки може також містити заголовки тією ж мовою. Заголовки тією ж мовою слід конвертувати у поле 400 UNIMARC. Водночас поле 400 у MARC21 містить деякі заголовки іншими мовами, які слід конвертувати у поле 700 UNIMARC. Примітно, що відсутність явної інформації про мову у полях 400 та 700 MARC21 вимагає ручної зміни номерів полів 400/700 після конвертації для забезпечення точного відображення у UNIMARC.

UNIMARC 120\$a/0 є простішим і має лише 5 опцій: чоловік, жінка, трансгендер, невідомо, не застосовується. Легко зіставити чоловіка і жінку, але для трансгендерів використовується така логіка: поле MARC21 375 (стать) використовується більше одного разу або містить підполя \$s (початок періоду) чи \$t (кінець періоду).

Країна агентства каталогізації для UNIMARC 801\$a може бути визначена за кодом агентства. Коди ISIL починаються з двосимвольного коду країни. Якщо використовується код організації MARC, тоді можна використовувати таблицю

конвертації, яка містить найуживаніші коди. Якщо ні, то країна не є обов'язковим підполем, і його можна не заповнювати.

Міжнародний стандартний ідентифікатор імен (ISNI). MARC21 не має спеціального підполя для ISNI і зберігає його у полі 024 (інший стандартний ідентифікатор), вказуючи, що джерелом (\$2) є ISNI. Це дозволяє конвертувати його в окреме поле 010 UNIMARC.

Специфікація переходу включає детальне зіставлення полів і підполів між форматами. Неконвертовані елементи даних явно позначені курсивом у специфікації для сприяння порівнянню форматів, полегшення майбутніх досліджень і забезпечення чіткої документації обмежень конвертації. У UNIMARC немає такого статусу запису (000/05), як «застарілий». Це означає, що запис не використовується, але ще не видалений. Для записів, які ще не видалені, конвертований код мав би бути для «виправленого запису» (c), однак конвертований код є для «видаленого запису» (d), оскільки він не використовується, і видалені записи все ще можуть містити всі підполя, згідно з UNIMARC. Такі рішення незначно змінюють записи, роблячи процес конвертації не повністю безвтратним.

Подальші напрями подальших досліджень. Удосконалення конвертації елементів даних. Наприклад, конвертація персональних імен MARC21 не лише в автор/заголовок записи, але й в автор/колективний заголовок. Конвертація контрольного підполя \$6. Використання джерела схеми дат для кращої конвертації дат.

Висновок. Дане дослідження продемонструвало можливість конвертації авторитетних записів MARC21 у формат UNIMARC з акцентом на авторитетні записи персональних імен. Завдяки детальній специфікації конвертації було досягнуто високого рівня успішності зіставлення підполів – 98.6%, що підкреслює надійність і ефективність методології.

Аналіз виявив суттєву узгодженість між двома форматами щодо основних полів даних. Однак також були виявлені ключові виклики, особливо у тих аспектах, де MARC21 і UNIMARC відрізняються у філософії дизайну.

Наприклад, більш багаті можливості UNIMARC для мови, сценарію та контекстних метаданих вимагали інноваційних рішень, включаючи виведені значення та ручні коригування, для забезпечення точного зіставлення.

Поля і підполя, які не вдалося конвертувати, наприклад, ті, що зберігають URL або певні метадані, специфічні для MARC21, вказують на прогалини в поточних стандартах, які можуть вимагати подальшого розвитку або вдосконалення.

Одним із ключових внесків дослідження є дотримання стандартів. Уникаючи створення нових полів або підполів, конвертовані записи залишаються сумісними в рамках існуючої структури UNIMARC. Це забезпечує їх інтеграцію в міжнародні системи каталогізації, зменшуючи ручне навантаження на бібліотекарів і підтримуючи ширшу мету гармонізації бібліографічних даних.

Дослідження також підкреслює необхідність постійного розвитку та оновлення стандартів. Наприклад, відсутність спеціального підполя у UNIMARC для URL (таких як MARC21 24\$1 і 670\$u) обмежувала можливості зіставлення веб-ресурсів. Однак очікуване введення підполя \$R для URL в UNIMARC демонструє, як стандарти еволюціонують для задоволення нових потреб. Майбутні оновлення MARC21 і UNIMARC можуть ще більше узгодити їхні структури, сприяючи більш безперешкодній конвертації.