

MARC 21과 BIBFRAME 2.0의 변환에 관한 기초 연구*

이 성 숙, 이 지 원†

충남대학교 문헌정보학과, 대구가톨릭대학교 도서관학과

기존의 MARC 레코드를 BIBFRAME으로 변환하기 위한 과제중 하나는 둘 간의 변환 명세를 마련하는 것이다. 이 연구의 목적은 LC에서 2017년 발표한 ‘MARC 21과 BIBFRAME 2.0과의 변환 명세’를 살펴봄으로, BIBFRAME 2.0에 대한 이해를 높이고, KORMARC 레코드 변환을 위한 시사점을 도출하기 위한 것이다. 본문에서 BIBFRAME 2.0의 주요 클래스인 저작, 인스턴스, 아이템을 중심으로 속성과 클래스를 범주화하였다. 연구결과 BIBFRAME 1.0은 53개의 클래스와 289개의 속성이 선언되었고, BIBFRAME 2.0에서는 1.0의 일부 속성이 클래스로 변경되어 186개의 클래스, 195개의 속성이 정의되었다. 이 연구에서 확장과 추론이 가능한 새로운 형태의 목록데이터를 생성할 수 있는 RDF 기반 BIBFRAME 2.0 어휘 중 속성값으로 문자열이 오는 사례와 속성값으로 클래스가 오는 사례를 확인하였다. 또한 21개의 파일로 제공되는 ‘MARC 21과 BIBFRAME 2.0과의 변환 명세’를 토대로, BIBFRAME 2.0의 클래스/속성과 매핑되는 MARC 21 표시기호/식별기호/지시기호의 매핑을 검토한 결과, BIBFRAME과 MARC 21과의 1대다 매핑 사례, MARC 21과 BIBFRAME과의 1대다 매핑 사례, BIBFRAME 기준으로 매핑되지 않는 클래스/속성 사례, MARC 21 기준으로 매핑되지 않거나 고려되지 않는 표시기호/식별기호/지시기호 사례를 도출하였다. 도출된 연구결과는 KORMARC와 BIBFRAME의 매핑을 위한 기초자료로 활용될 수 있을 것이다.

주요어: 서지프레임워크, 기계가독형목록, 변환 명세, 매핑, 크로스워크, 저작, 인스턴스, 아이템

* 본 연구는 2015년 충남대학교 교내연구비 지원에 의해 수행되었음.

† 교신저자(Corresponding Author) : 이지원, 대구가톨릭대학교 도서관학과 부교수, 경북 경산시 하양읍 하양로 13-13, E-mail : jiwon@cu.ac.kr

■ 최초투고일 : 2018년 6월 1일 ■ 심사마감일 : 2018년 6월 26일 ■ 게재확정일 : 2018년 6월 26일

1. 서론

도서관을 둘러싼 오늘날의 정보환경은 공유와 연계 환경으로 변화하고 있다. 이러한 환경의 대표적인 흐름은 LOD(Linked Open Data)이다. 이에 도서관들은 기 구축한 서지데이터나 전자데이터를 링크드 데이터로 변환하여 서비스하고 있다. 예컨대 LC(Library of Congress, 미국 국회도서관)는 LDS(Linked Data Service)를 통해 LC가 작성한 전자데이터를 제공한다. 또한 영국 국립도서관은 'Linked Open BNB'에서 영국의 전국 서지 약 280만건을 제공하고, 프랑스 국립도서관은 전용 Open Data 포털 사이트 'data.bnf.fr'을 통해 서지데이터와 전자데이터, 전자도서관 Gallica의 메타 데이터를 제공한다. 이외에도 독일과 스페인 국립 도서관도 기존 목록 데이터를 LOD로 변환하여 제공한다.

도서관의 또 다른 움직임은 목록 업무에 사용되는 포맷의 데이터 모델로 RDF(Resource Description Framework)기반의 Linked Data를 사용하는 것이다. LC가 2012년 발표한 BIBFRAME(Bibliographic Framework)은 MARC 21을 대체하는 데이터 포맷의 일종이지만, 그 데이터 모델의 기반은 Linked Data를 채용하고 있다. BIBFRAME은 기존의 MARC(Machine Readable Cataloging) 레코드와의 호환성을 유지하면서 MARC을 대체할 수 있는 새로운 서지기술을 위한 데이터모델을 제시하기 위한 것이며, 서지적 기술 표준이 링크드 데이터 모델로 발전하기 위한 것으로서, 이는 도서관 내부뿐만 아니라 오픈 웹 환경에서 서지 정보를 보다 쉽게 접근하기 위한 것이다(LC Bibliographic Framework Initiative, 2016).

LC는 2008년 MARC에 대한 보고서에서 MARC는 40여 년 전의 기술에 기반을 두기 때문에 오늘날의 프로그래밍 환경에 적합하지 않다고 평가

하였다. 2011년 LC는 공식적으로 Bibliographic Framework Initiative를 시작하였고, 2012년부터 BIBFRAME과 관련된 몇몇 사업을 진행하였다. 구체적으로 2012년 LC는 데이터 관리회사인 Zepheira와 계약을 맺고 MARC를 대체하기 위한 링크드 데이터 모델을 개발하였다. LC는 MARC Resources라 부르는 새로운 모델을 발표하였고, 그해 11월 BIBFRAME으로 재명명된 보다 구체적인 모델의 초안을 발표하였다. 또한 2013년 초기 실험을 위한 어휘와 변환도구를 테스트하였고, 2014년 구현 테스트베드를 진행하였고, 2015년 BIBFRAME 1.0기반으로 한 pilot 1 프로젝트를 진행하였다. 2016년 BIBFRAME 2.0 모델과 어휘를 발표하였고, 2017년 BIBFRAME 2.0 기반으로 한 pilot 1 프로젝트를 진행하였다(Xu, Hess & Akerman, 2017).

우리나라의 경우, 학술정보에 대한 LOD 서비스는 국립중앙도서관(<https://lod.nl.go.kr/>)과 KERIS(<http://data.riss.kr/>), KISTI(<http://lod.ndsl.kr/>) 등이 제공하고 있으며, 국회도서관, KISTI, 한국연구재단, 성균관대학교 도서관이 참여하여 국가 학술정보융합데이터 LOD서비스(<http://cloud.nanet.go.kr/lod/>)도 제공하고 있다. 이러한 학술정보의 LOD 서비스 확장과 함께 기존 포맷인 MARC의 새로운 접근인 BIBFRAME을 중요하게 검토하는 것이 학술정보 LOD 서비스의 통일성과 정확성을 확보하기 위한 선결조건일 것이다. 더욱이 도서관 분야에서 그동안 MARC가 미친 영향을 생각하면, 향후 BIBFRAME의 파급효과가 클 것으로 예상된다.

BIBFRAME에 대한 국내 선행연구는 BIBFRAME 1.0의 특징을 언급한 박진호(2013)와 박지영(2013)의 연구가 있다. 또한 BIBFRAME 1.0을 시범 적용하여 온톨로지를 구축한 박옥남(2015)의 연구와 공공도서관 서지데이터와 서비스데이터를 연계

한 박지영(2016)의 연구가 있다. BIBFRAME 1.0과 MARC 매핑 관련 연구로는 박옥남과 오정선(2014)의 연구가 있으며, 이 연구에서는 당시 MARC과 BIBFRAME 변환 매핑표가 발표되지 않은 관계로 LC가 공개한 MARC 변환도구의 XQuery 핵심 모듈 소스 코드와 실제 레코드 변환 결과를 분석한 것을 토대로, 핵심 MARC 데이터 필드와 서브 필드가 BIBFRAME의 클래스와 속성과 어떻게 매핑되는지 살펴보았다. BIBFRAME 2.0에 대한 연구로는 이미화(2017)의 연구가 있으며, 이 연구에서는 BIBFRAME 2.0의 특징을 변경사항을 중심으로 분석하고 이를 바탕으로 국내 구축시 고려사항을 제시하였다. 또한 최근 BIBFRAME 환경 하에서 목록사서의 대응 전략에 관한 연구(이미화, 2018)와 서지데이터 공유 및 활용을 위한 BIBFRAME 2.0 수용 방안 연구(이혜원, 2018)도 진행되었다.

선행연구 분석 결과 서지분야에서 MARC가 차지했던 비중을 고려하면, BIBFRAME에 대한 추가 연구가 필요한 것으로 판단된다. 특히 BIBFRAME 2.0이 발표되었음에도 국내 선행연구는 BIBFRAME 1.0 중심으로 진행되었고, BIBFRAME 2.0에 대한 연구는 소수이기 때문이다. 또한 MARC 레코드를 BIBFRAME의 링크드 데이터화하기 위해서는 MARC과 BIBFRAME간의 변환이 우선되어야 하지만, 그동안 MARC와 BIBFRAME간의 직접적인 변환 테이블이 부재하여 박옥남과 오정선(2014)의 연구에서는 MARC 변환도구의 XQuery 핵심 모듈 소스 코드와 실제 레코드 변환 결과를 분석하고, BIBFRAME 1.0과 MARC과의 관계를 유추하였다. 그러나 LC는 2016년 BIBFRAME 2.0 모델과 어휘를 발표하였고, 2017년 3월 MARC 21과 BIBFRAME 어휘의 변환 명세(conversion specifications)인 ‘MARC 21 to BIBFRAME 2.0 Conversion Specifications’

(LC Bibliographic Framework Initiative, 2017)를 공개하였다.

기존의 MARC 레코드를 BIBFRAME의 링크드 데이터로 변환하기 위한 우선과제는 MARC 21과 BIBFRAME과의 변환 명세를 마련하는 것이다. 이 연구에서는 BIBFRAME 2.0 모델과 어휘를 이해하고, KORMARC 포맷의 대안적인 포맷으로서 자원 기술 데이터 모델의 변환 명세 마련을 위한 시사점을 도출하기 위해, LC에서 공개한 MARC 21과 BIBFRAME 2.0과의 변환 명세를 살펴보고자 한다. 다만 MARC 21과 BIBFRAME 2.0의 변환 명세의 범위가 광범위함으로, 이 연구에서는 기초연구의 성격으로 MARC 21과 BIBFRAME 2.0 요소간의 매핑에 중점을 두었다.

이 연구는 BIBFRAME 2.0에 대한 문헌조사와 MARC 21과 BIBFRAME 2.0의 변환 명세를 분석하는 문헌조사를 통해 이루어졌다. 변환 명세를 이해하기 위해서는 우선 BIBFRAME 2.0 모델에 대한 이해가 필수이므로 2장에서는 BIBFRAME 2.0 모델에 대해 살펴보고, 3장에서는 구체적인 변환 대상이 되는 BIBFRAME 2.0 어휘를 살펴보고, MARC 21과 BIBFRAME 2.0의 변환 명세를 개괄적으로 살펴보았다. 4장에서는 MARC 21과 BIBFRAME 2.0의 변환 관계를 분석하고 시사점을 도출하였다.

2. BIBFRAME 2.0 모델

MARC 표준은 지난 40년간 컴퓨터에 의한 서지레코드 생성에 기여해왔고, MARC 레코드는 도서관 목록에서 검색된 자료에 대한 OPAC(Online Public Library Catalog) 레코드로 활용되었다. MARC은 LC가 MARC의 웹 표준인 MARCXML

을 만들 때까지 그 위치가 견고하였다. MARCXML은 MARC의 필드와 서브필드를 XML 스키마로 재포맷한 것으로, 새로운 모델이 아니기 때문에 널리 사용되지 못하였다. LC는 MARC가 오늘날의 프로그래밍 환경에 적합하지 않다고 평가하고, Zepheria와 함께 BIBFRAME라 불리는 새로운 서지 프레임워크를 개발하기 시작하였다. 이는 도서관 레코드를 웹 표준을 따르도록 하는 것이다. BIBFRAME은 대규모 웹에 도서관 레코드를 접근 가능하도록 하는 첫번째 링크드 데이터 모델이다(Hansen & Mueller, 2015).

LC는 목록레코드의 데이터화를 위해 RDF 기반의 링크드 데이터 모델인 BIBFRAME 모델 및 어휘를 발표하였다. BIBFRAME은 도서관 자원을 기술할 때 RDF화하기 위한 것이다. BIBFRAME의 이점은 구조화된 형태로 웹에 정보를 출판하기 위한 표준인 ‘링크드 데이터’를 사용한다는 것이다. 예컨대 BIBFRAME은 도서관 데이터를 웹에 보다 가시적으로 만들고, 도서관 데이터를 보다 광범위하고 효과적으로 공유하게 하며, 도서관 외부 세계의 이용자가 다루기에 접근가능한 형식을 제공한다(McGrath, 2017). 또 다른 특징으로는 링크드 데이터 원칙을 이용함으로써 개념적 콘텐츠와 물리적 구현물의 구분이 가능하고, 정보 개체를 식별하기 위한 중의성 해소에 도움이 되며, 개체 간의 관계 정보 표현 방안을 제시한다는 것이다(박지영, 2013). BIBFRAME은 MARC를 대체하기 위한 것이다. 그러나 그동안 도서관 분야에 MARC이 여러 가지 일을 수행하였기 때문에 MARC의 어떤 기능을 대체하는지 명확하지 않다는 비판도 있다(McGrath, 2017).

LC는 2012년 BIBFRAME 1.0을 발표하였으며, BIBFRAME 1.0은 4개의 주요 클래스(classes)를 정의하였다. 그 후 LC는 BIBFRAME 1.0에 대한 LC pilot, 외부 전문가와 파트너, 이메일 리스트

등으로부터 많은 의견을 받았다. 이에 LC는 다양한 의견을 수용하고, 특히 시청각 자료(AVPreserve)에 적합한 모델을 위해, 또한 링크드 데이터 모범 사례(best practices)를 보다 준수하여 어휘를 재 개발하였고, 2016년 BIBFRAME 2.0을 발표하였다(LC Bibliographic Framework Initiative, 2018a; McGrath, 2017).

BIBFRAME 2.0에는 저작(Work), 인스턴스(Instance), 아이템(Item), 에이전트(Agents), 사건(Events), 주제(Subjects)의 클래스(엔티티)가 선언되었다. BIBFRAME 2.0 추상화의 핵심 수준은 저작, 인스턴스, 아이탬이다(LC Bibliographic Framework Initiative, 2018b). 책과 같은 어떤 자원을 저작, 주제, 출판형태, 복본 형식 등의 정보 요소를 포함해서 편목할 때, BIBFRAME 2.0은 저작, 인스턴스, 아이탬 등 추상화의 기본 단계로 정보를 조직한다. 또한 BIBFRAME 2.0은 핵심적인 클래스와 관련된 부가적인 핵심 개념으로 에이전트, 사건, 주제 등을 정의하였다. BIBFRAME 2.0에서는 1.0의 전거와 주석 클래스를 삭제하고, 아이탬과 사건을 추가하였다.

3. MARC 21과 BIBFRAME 2.0의 변환 명세

1) BIBFRAME 2.0 어휘

BIBFRAME 어휘는 자원을 기술하기 위한 핵심적인 것이다. BIBFRAME 2.0에서 핵심클래스와 부가적인 클래스와 함께 속성(properties)이 정의되었다. 속성은 자원자체의 특징이나 자원 간의 관계를 기술하는 것이다. 즉 BIBFRAME 어휘는 클래스와 속성으로 구성된다. BIBFRAME 2.0 어휘는 Category view, List view, RDF

View 측면에서 제시되고 있다(LC Bibliographic Framework Initiative, 2018c). BIBFRAME 2.0에서 정의된 클래스는 <표 1>, 속성은 <표 2>와 같다.

MARC 포맷이 요소(elements)와 속성(attributes)의 셋으로 정의된 것처럼, BIBFRAME 어휘는 클래스와 속성으로 정의된다. MARC 필드가 단일 개념을 묶을 수 있는 것처럼 클래스는

BIBFRAME 자원의 타입을 식별하기 위한 것이고, MARC 서브필드가 개념의 측면을 식별하기 위한 것처럼 속성은 BIBFRAME 자원을 기술하기 위한 수단으로 사용된다. RDA와 BIBFRAME 2.0의 매핑표에서도 확인할 수 있듯이(University of Washington Libraries, 2018), BIBFRAME에서 정의된 어휘는 RDA와 MARC를 상당부분 수용하였고, 필요한 일부 요소들을 추가하였다.

<표 1> BIBFRAME 2.0의 클래스

AbbreviatedTitle*, AccessPolicy*, AcquisitionSource*, AdminMetadata*, Agent, Ansi*, AppliedMaterial*, Archival, Arrangement, AspectRatio*, Audio, AudioIssueNumber*, AudioTake*, Barcode*, BaseMaterial*, BookFormat*, BroadcastStandard*, Capture*, Carrier*, Cartographic*, CartographicDataType*, CartographicObjectType*, Cartography, Chronology*, Classification, ClassificationDdc*, ClassificationLcc*, ClassificationNlm*, ClassificationUdc*, Coden*, Collection, CollectiveTitle*, ColorContent*, Content*, ContentAccessibility*, Contribution*, CopyrightNumber*, CopyrightRegistration*, CoverArt, Dataset, DescriptionAuthentication*, DescriptionConventions*, DigitalCharacteristic*, Dissertation*, DissertationIdentifier*, Distribution*, Doi*, Ean*, Electronic, Emulsion*, EncodedBitrate*, EncodingFormat*, Enumeration*, EnumerationAndChronology*, Event, Extent*, Family, FileSize*, FileType*, Fingerprint*, FontSize*, Frequency*, Generation*, GenerationProcess*, GenreForm*, GeographicCoverage*, GrooveCharacteristic*, Gtin14Number*, Hdl*, Identifier, Illustration*, ImmediateAcquisition*, Instance, IntendedAudience, Isan*, Isbn*, Ismn*, Isni*, Iso*, Isrc*, Issn*, IssnL*, Issuance*, Iste*, Iswc*, Item*, Jurisdiction, KeyTitle*, Language, Layout*, Lccn*, LcOverseasAcq*, Local*, Manufacture*, Manuscript, MatrixNumber*, Media*, Meeting, MixedMaterial, Mount*, MovementNotation*, MovingImage, Multimedia, MusicDistributorNumber*, MusicEnsemble*, MusicFormat*, MusicInstrument*, MusicMedium*, MusicNotation*, MusicPlate*, MusicPublisherNumber*, MusicVoice*, Nbn*, NotatedMovement, NotatedMusic, Notation*, Note*, Object* ObjectCount*, Organization, ParallelTitle*, Person, Place, PlaybackChannels*, PlaybackCharacteristic*, PlayingSpeed*, Polarity*, Chronology*, PostalRegistration*, PresentationFormat*, Print, Production*, ProductionMethod*, Projection*, ProjectionCharacteristic*, ProjectionSpeed*, ProvisionActivity*, Publication*, PublisherNumber*, RecordingMedium*, RecordingMethod*, ReductionRatio*, RegionalEncoding*, ReportNumber*, Resolution*, RetentionPolicy*, Review, Role*, Scale*, Script*, ShelfMark*, ShelfMarkDdc*, ShelfMarkLcc*, ShelfMarkNlm*, ShelfMarkUdc*, Sici*, SoundCharacteristic*, SoundContent*, Source*, Status*, StillImage, StockNumber*, Strn*, StudyNumber*, Sublocation*, Summary, SupplementaryContent*, SystemRequirement*, TableOfContents, Tactile, TactileNotation*, TapeConfig*, Temporal, Text, Topic, TrackConfig*, Unit*, Upc*, Urn*, UsageAndAccessPolicy*, UsePolicy*, VariantTitle*, VideoCharacteristic*, VideoFormat*, VideoRecordingNumber*, Work

주) 클래스 뒤에 있는 *는 BIBFRAME 2.0에서 신규 클래스로 선언된 어휘임.

LC의 BIBFRAME 2.0 홈페이지와 BIBFRAME 1.0(<http://bibframe.org/vocab-list/>) 어휘를 비교하여 작성하였음.

<표 2> BIBFRAME 2.0의 속성

absorbed, absorbedBy, accompaniedBy, accompanies, acquisitionSource, acquisitionTerms, adminMetadata, agent, appliedMaterial, arrangement, ascensionAndDeclination, aspectRatio, assigner, awards, baseMaterial, bookFormat, capture, carrier, cartographicAttributes, changeDate, classification, classificationPortion, code, colorContent, content, contentAccessibility, continuedBy, continuedInPartPartBy, continues, continuesInPart, contribution, coordinates, copyrightDate, copyrightRegistration, count, coverArt, creationDate, credits, custodialHistory, dataSource, date, degree, derivativeOf, derivedFrom, descriptionAuthentication, descriptionConventions, descriptionLanguage, descriptionModifier, digitalCharacteristic, dimensions, dissertation, duration, edition, editionEnumeration, editionStatement, electronicLocator, emulsion, ensemble, ensembleType, enumerationAndChronology, equinox, eventContent, eventContentOf, exclusionGRing, expressionOf, extent, findingAid, findingAidOf, firstIssue, fontSize, frequency, generation, generationDate, generationProcess, genreForm, geographicCoverage, grantingInstitution, hasDerivative, hasEquivalent, hasExpression, hasInstance, hasItem, hasPart, hasReproduction, hasSeries, hasSubseries, heldBy, hierarchicalLevel, historyOfWork, identifiedBy, identifies, illustrativeContent, immediateAcquisition, index, indexOf, instanceOf, instrument, instrumentalType, intendedAudience, issuance, issuedWith, itemOf, itemPortion, language, lastIssue, layout, legalDate, mainTitle, media, mergedToForm, mergerOf, mount, musicFormat, musicKey, musicMedium, musicOpusNumber, musicSerialNumber, musicThematicNumber, natureOfContent, notation, note, noteType, organization, originalVersion, originalVersionOf, originDate, originPlace, otherEdition, otherPhysicalFormat, outerGRing, part, partName, partNumber, partOf, pattern, physicalLocation, place, polarity, precededBy, preferredCitation, productionMethod, projection, projectionCharacteristic, provisionActivity, provisionActivityStatement, qualifier, reductionRatio, referencedBy, references, relatedTo, replacedBy, replacementOf, reproductionOf, responsibilityStatement, review, role, scale, schedulePart, separatedFrom, seriesEnumeration, seriesOf, seriesStatement, shelfMark, soundCharacteristic, soundContent, source, spanEnd, splitInto, status, subject, sublocation, subseriesEnumeration, subseriesOf, subseriesStatement, subtitle, succeededBy, summary, supplement, supplementaryContent, supplementTo, systemRequirement, table, tableOfContents, tableSeq, temporalCoverage, title, translation, translationOf, unit, usageAndAccessPolicy, variantType, version, videoCharacteristic, voice, voiceType

BIBFRAME 2.0과 1.0을 비교하면, BIBFRAME 1.0은 53개의 클래스와 289개의 속성으로 구성되며, BIBFRAME 2.0은 186개의 클래스, 195개의 속성으로 클래스 수는 늘었고, 속성은 줄었다. BIBFRAME 2.0에서 신규로 정의된 클래스(<표 1>에서 *표시)가 많다. BIBFRAME 1.0에서 속성으로 선언된 어휘 중 일부가 2.0에서는 클래스로 선언되었다. BIBFRAME 2.0의 변경 어휘는 1.0에 대한 각계의 리뷰 의견이 반영된 것이다. 또한 1.0에 대해 링크드 데이터 모범 사례의 관점에서 분석한 보고서(Sanderson, 2015)의 결과가

반영된 것이다.

2) MARC 21과 BIBFRAME 2.0의 변환 명세

BIBFRAME은 2012년 개발되었지만, LC는 상당한 시간이 흐른 2017년 MARC과 BIBFRAME 어휘의 변환 명세인 ‘MARC 21 to BIBFRAME 2.0 Conversion Specifications’를 발표하였다(LC Bibliographic Framework Initiative, 2017). LC의 변환 명세는 MARC 21의 각 요소를 고려 하되, LC 레코드에서 변환을 시도하지 않은 요소

(변환 명세에서 'nac'(no attempt to convert) 사용)나 MARC 환경의 외부와 관련이 없는 요소(변환 명세에서 ignore 사용)는 매핑 대상에서 제외하였다. 또한 MARC 전거와 관련해서는 표제 전거변환명세만 포함하였다. LC는 BIBFRAME 2.0이 수정되는 동안, 이 변환 명세가 지속적으로 수정될 것이라고 밝히고 있다. 2017년 발표된 이 변환 명세에는 21개의 파일(MARC 서지변환명세 20개, 표제전거변환명세 1개 포함)이 제공되며, 21개의 파일을 합치면 총 5,000줄 이상을 포함한 분량이며, RDF 문장(statement)으로 표현된 내용을 담고 있다. 이는 MARC에서 BIBFRAME 으로의 전환은 단순한 레코드 형식의 변화에 그치는 것이 아니라 RDF 문장으로 표현되는 서지 기술 방식 자체의 변화를 의미하기 때문이다.

BIBFRAME의 기술방식에 대해 간단히 살펴보면, 하나의 MARC 레코드는 하나의 저작에 대한 RDF 문장의 집합으로 변환된다. 예컨대 MARC에서 변환된 BIBFRAME RDF 문서는 URI로 식별되는 저작을 주어(subject)로 하는 트리플(triple)의 집합으로 구성되며, 저작을 정의역(domain)으로 갖는 속성 중 주어진 MARC 레코드에서 데이터를 추출하여 속성값(object)을 구성할 수 있는 속성들이 모두 사용된다. 저작(bf:Work)의 속성들은 대부분 상응하는 공역(range)으로 갖기 때문에 URI나 공백노드(blank node)를 사용하여 기술한다. MARC 서브필드에 포함된 혹은 지시 기호로 표현된 데이터는 공백노드에 연결된 속성을 이용하여 기술한다(박옥남, 오정선, 2014).

변환 명세는 ① Fields 001-005, 007 Control, physical description - R1(06/07/2017), ② Fields 006, 008-Fixed fields-R0(03/07/2017), ③ Fields 010-048 - Identifiers, etc. - R1(06/02/2017), ④ Field 048-Instrument codes-R0(03/07/2017), ⑤ Fields 050-088-Class/call numbers, etc.-R1

(01/02/2018), ⑥ Fields 1XX, 6XX, 7XX, 8XX: X00, X10, X11 - Names-R1(01/02/2018), ⑦ Fields 200-24X, except 240 - Titles - R1(06/08/2017), ⑧ Fields 240, X30, etc. - Uniform titles - R2(01/02/2018), ⑨ Fields 250-270 - Edition, imprint, etc. - R2(01/24/2018), ⑩ Fields 3XX - Physical description, etc. - R1(06/07/2017), ⑪ Fields 490, 510, 530-535 - Links - R1(06/09/2017), ⑫ Fields 5XX - Notes - R2(01/02/2018), ⑬ Fields 647-662 - Subjects - R2(02/01/2018), ⑭ Fields 720, 740-758 - Added entries - R1(02/09/2018), ⑮ Fields 760-788-Linking entries-R0(07/2017), ⑯ Fields 841-887 - Holdings, location, alternate graphics, etc. - R2(01/02/2018), ⑰ Leader-R0(03/07/2017), ⑱ Numeric Subfield Conversions - R2(01/02/2018), ⑲ Process notes - R2(07/27/2017), ⑳ Subject Thesaurus indicator-R0(03/07/2017), ㉑ Authority Info: Name/Title and Title-R0(07/28/2017)이다(LC Bibliographic Framework Initiative, 2017).

4. MARC 21과 BIBFRAME 2.0의 매핑

4장의 매핑 분석은 다음과 같은 방법으로 진행하였다. LC가 발표한 변환 명세는 <표 3>의 예시와 같이, MARC 21의 필드를 기준으로 지시 기호, 식별기호마다 대응하는 BIBFRAME 어휘를 제시하고 있다. 연구자는 먼저 총 21개의 파일을 합쳐 하나의 파일로 모아 작업의 기초가 되는 통합파일을 생성하였다. 다음으로 통합파일을 토대로 BIBFRAME 2.0 어휘(<http://id.loc.gov/ontologies/bibframe.html>) 중 속성과 클래스를 엑

<표 3> LC의 MARC 21과 BIBFRAME의 매핑표<일부>

Fields 200-24X, except 240 - Other title fields - R1, 06/08/2017	Conversion 1
210 - ABBREVIATED TITLE (R)	I-title-VariantTitle-AbbreviatedTitle
Indicators	
First - Title added entry	ignore
0 - No added entry	
1 - Added entry	
Second - Type	
# - Abbreviated key title	## - source - Source - rdf:value - "issnkey"
0 - Other abbreviated title	ignore
Subfield Codes	##-rdfs:label [Make title label by extracting \$ab, keep order in field, substitute blank for each subfield code]
\$a - Abbreviated title (NR)	## - mainTitle - literal
\$b - Qualifying information (NR)	## - qualifier - literal
\$2 - Source (R)	## - source -Source - rdfs:label - "content of \$2"
생략	생략

셀 ‘찾기 및 바꾸기’ 기능을 활용하여 ‘모두 찾기’를 한 후, 육안으로 의미있는 매핑 어휘를 확인하고 MARC 21의 해당 필드의 지시기호와 식별기호를 확인하였다.

또한 이 연구의 매핑 결과의 서술 방식은 BIBFRAME 2.0 어휘에 대한 이해를 높이기 위해, 변환 명세에 나열된 MARC 21의 필드 순보다는 4장 각 절의 그림과 같이 BIBFRAME 어휘 중심으로 기술하였다. 아울러 LC의 홈페이지에서 제공한 BIBFRAME 2.0의 속성과 클래스를 살펴본 후, 매핑표에 나열된 알파벳 순 나열식보다는 BIBFRAME 2.0의 주요 클래스에 사용되는 어휘별로 그룹핑하여 분석하는 것이 효과적인 것으로 판단하였다. 어휘별 그룹핑은 Miller(2017)의 구분을 따랐다. Miller는 LC 홈페이지에서 제공하는 각 속성의 설명에서 ‘Used with’을 기준으로 ① 저작 사용 어휘, ② 저작과 인스턴스의 공통사용 어휘, ③ 인스턴스 사용 어휘, ④ 아이tem 사용 어휘, ⑤ 저작, 인스턴스, 아이tem 공통사용 어휘, ⑥ 불특정(unspecified) 사용 어휘로 구

분하였다. Miller는 MARC과의 매핑은 고려하지 않았기 때문에, 연구자는 Miller의 구분만을 채용하였다.

4장 각 절의 그림 작성 방법을 살펴보면, 그림에서 표시기호만 표기된 것은 nac나 ignore를 제외하고 필드 전체가 매핑되는 경우를 의미하며, 표시기호와 식별기호, 지시기호가 따로 표기된 것은 해당 표시기호 내에서 식별기호, 지시기호 마다 다른 BIBFRAME 2.0 어휘에 매핑되는 것을 의미한다. 각 절의 그림에는 BIBFRAME 2.0과 MARC 21과의 매핑 결과, 속성값이 문자열(literal, 글자로 된 스트링 또는 정수 등)을 갖는 경우와 속성값이 클래스를 갖는 경우, BIBFRAME 2.0 저작 기준으로 1대 다 매핑 사례를 포함하였다. 다만 MARC 21 기준으로 1대 다의 매핑 사례는 [부록 1]에, MARC 21 기준으로 BIBFRAME 어휘와 매핑되지 않는 어휘는 [부록 2]에 수록하였다. 그림에서 매핑되는 MARC 21 필드가 없는 것은 해당되는 적합한 필드가 없는 경우도 있고, 변환을 시도하지 않은 nac와 ignore에 해당되는 경우도

있다.

1) 저작 어휘의 매핑

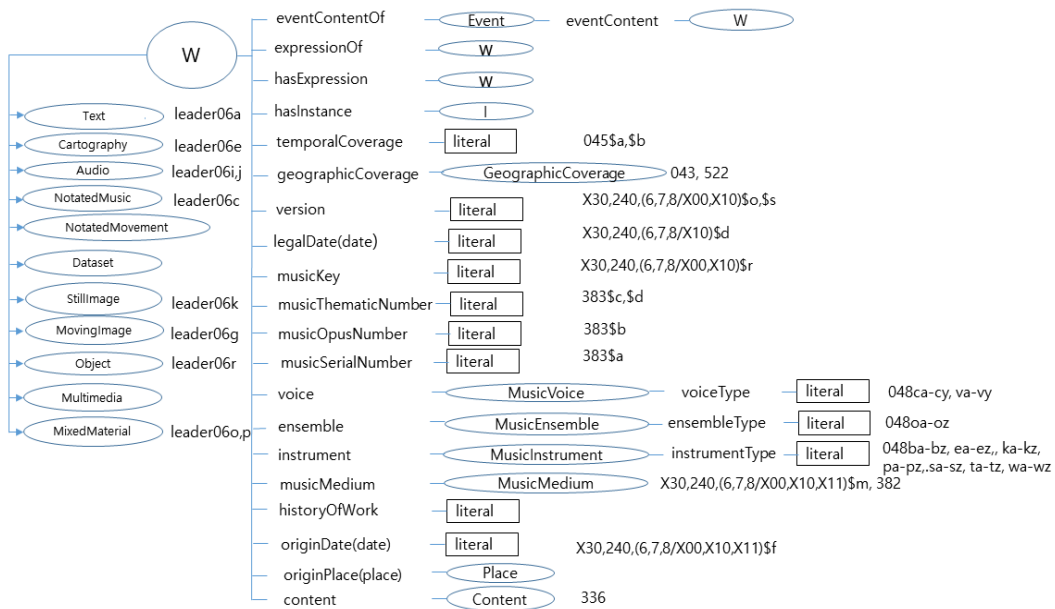
BIBFRAME 어휘 중 저작에만 사용되는 속성과 클래스를 중심으로 한 매핑 결과는 <그림 1>과 같다.

BIBFRAME 2.0 어휘 중 bf:Work의 하위 클래스는 bf:Text, bf:Cartography, bf:Audio, bf:NotatedMusic, bf:Notated Movement, bf:Dataset, bf:Stillimage, bf:MovingImage, bf:Object, bf:Multimedia, bf:Mixed Material이다. LC 변환 명세의 한 예를 살펴보면, MARC 21 리더 06(레코드 유형) a(문자자료)는 BIBFRAME 2.0 어휘 bf:Work - rdf:type - bf:Text로 변환한다. RDF에서 rdf:type은 각각의 인스턴스를 클래스에 연결시키는 서술부(predicate)이며, rdf:type의 주어부에는 묘사하고자 하는 인스턴스가 오고, 목적부

는 인스턴스가 속한 클래스가 자리한다. 즉 ‘저작이라는 자원의 타입이 텍스트이다’를 나타낸다.

저작의 하위클래스인 bf:Text는 MARC의 리더 06 a문자자료, bf:Cartography는 e지도자료(구체포함), bf:Audio는 i녹음자료(음악 이외), j녹음자료(음악), bf:NotatedMusic는 c필사악보 이외의 악보, bf:Stillimage는 k평면비영사자료, bf:MovingImage은 g평면영사자료, bf:Object는 r입체자료(실물), bf:Mixed Material는 o키트, p혼합자료와 매핑되어, 저작의 하위클래스들은 대부분 리더 06(레코드 유형)과 매핑되는 것을 알 수 있다. MARC 21과 매핑되지 않는 하위클래스는 bf:Notated Movement, bf:Dataset, bf:Multimedia이다.

저작 클래스에만 사용되는 속성은 bf:eventContentOf, bf:expressionOf, bf:hasExpression, bf:hasInstance, bf:temporalCoverag, bf:geographic Coverage, bf:version, bf:legalDate(date), bf:musicKey, bf:musicThematicNumber, bf:musicOpusNumber, bf:musicSerialNumber, bf:voice, bf:ensemble, bf:instrument, bf:musicMedium, bf:historyOfWork, bf:originDate(date), bf:originPlace(place), bf:content



<그림 1> BIBFRAME 저작 어휘와 MARC 21의 매핑 결과

:musicKey, bf:musicThematicNumber, bf:musicOpusNumber, bf:musicSerialNumber, bf:voice, bf:ensemble, bf:instrument, bf:MusicMedium, bf:historyOfWork, bf:originDate(date), bf:originPlace(place), bf:content 총 20개이다. 이 중 bf:eventContentOf, bf:expressionOf, bf:hasExpression, bf:hasInstance, bf:historyOfWork, bf:originPlace(place)는 MARC 21과 매핑되지 않는다.

속성을 속성값에 따라 구분하면 다음과 같다. 첫째, 저작에 사용되는 속성의 속성값이 문자열에 해당하는 경우이다. 문자열 노드는 다른 트리플의 주어가 될 수 없으며, 그래프는 그 상태에서 더 이상 연결되지 않는다. 여기에 해당하는 속성 bf:temporalCoverag는 MARC 21의 045\$a연대부호, \$b기원전 9999년부터 서기까지의 형식화된 연대부호와 매핑되며, bf:version은 X30, 240(6,7,8/X00,X10)\$o편곡, \$s판과 매핑되며, bf:legalDate(date)는 X30, 240(6,7,8/X10)\$d조약 체결일자과 매핑되며, bf:musicKey는 X30, 240(6,7,8/X00,X10)\$r음악자료의 장단조와 매핑되며, bf:musicThematicNumber는 383음악저작번호표시\$c주제색인번호, \$d주제색인부호와 매핑되며, bf:musicOpusNumber는 383\$b작품번호와 매핑되며, bf:musicSerialNumber는 383\$a일

련번호와 매핑되며, bf:originDate(date)는 X30, 240,(6,7,8/X00,X10)\$f저작년도와 매핑된다.

둘째, 속성값으로 클래스를 갖는 경우이다. RDF에서 속성값이 클래스를 갖는 것은 예를 들어 주소, 거리와 시, 주, 우편번호, 각기 별개로 구성된 구조로 기록할 필요가 있을 때, 이와 같이 구조화된 정보를 기술 대상인 통합된 대상(주소)으로 취급되어 RDF에서 하나의 자원으로 표현하고, 이 새로운 자원에 관한 선언문을 작성하게 된다. 예컨대 043지역부호의 변환 명세는 <표 4>와 같다. 043의 식별기호는 클래스의 URI로 변환되거나(\$a), rdfs:label로 변환되거나(\$b), 또 6절에서 서술할 또 다른 속성과 속성값으로 연결되도록(\$c)로 변환된다. 여기에서 rdfs:label(인간이 읽기 위한, 주어에 대응하는 이름)은 긴 URI를 사람이 읽기 편한 이름으로 나타내고자 할 때 쓰인다.

또한 048은 저작의 속성값으로 클래스를 가지고, 그 클래스의 속성이 속성값으로 문자열을 갖는다. MARC 21의 048 악기나 성악부호 ca-cy합창, va-vy성악은 bf:Work - bf:voice - bf:MusicVoice - bf:voiceType - literal로 변환된다. 또한 이와 유사한 패턴을 보이는 속성 bf:ensemble은 048 oa-oz대합주곡과 bf:instrument는 048의 나머지 부호와 매핑된다.

<표 4> MARC 21 043 필드의 BIBFRAME 2.0 어휘로의 변환

043 - GEOGRAPHIC AREA CODE	W-geographicCoverage-GeographicCoverage
Subfield Codes	
\$a - Geographic area code	##-URI : Use URI from id.loc.gov/vocabulary/geographicAreas
\$b - Local GAC code	##-rdfs:label : add source - Source rdfs:label "content of \$2"
\$c - ISO code	##-geographicCoverage-GeographicCoverage-source : Add source - Source rdfs:label "ISO 3166"
\$0 - Authority record control number or standard number	See Subfield \$0 spec
\$2 - Source of local code	See \$b above
\$6 - Linkage (NR)	ignore

속성 bf:geographicCoverage의 속성값 bf:GeographicCoverage는 043지역부호, 522지리적범위주기, 속성 bf:musicMedium의 속성값 bf:MusicMedium은 X30, 240,(6,7,8/X00,X10,X11)\$m연주수단, 382연주매체와 매핑되며, bf:content의 bf:Content는 336내용유형에 매핑된다. 이와 같이 BIBFRAME에서 속성 값이 클래스를 갖는 것은 자원에 대한 정보가 RDF 문장들과 그 연결 관계로 표현되며, 링크드 데이터 원칙에 따라 개체와 개념을 URI로 기술하므로 확장과 추론이 가능한 새로운 형태의 목록 데이터를 생성할 수 있게 된다.

2) 저작과 인스턴스 공통사용 어휘의 매핑

이 절부터는 앞 절과 동일한 특징은 각 절의 그림으로 대신하고자 한다. BIBFRAME 어휘 중 저작과 인스턴스에 공통으로 사용되는 속성과 MARC 21과의 매핑결과는 <그림 2>와 같다.

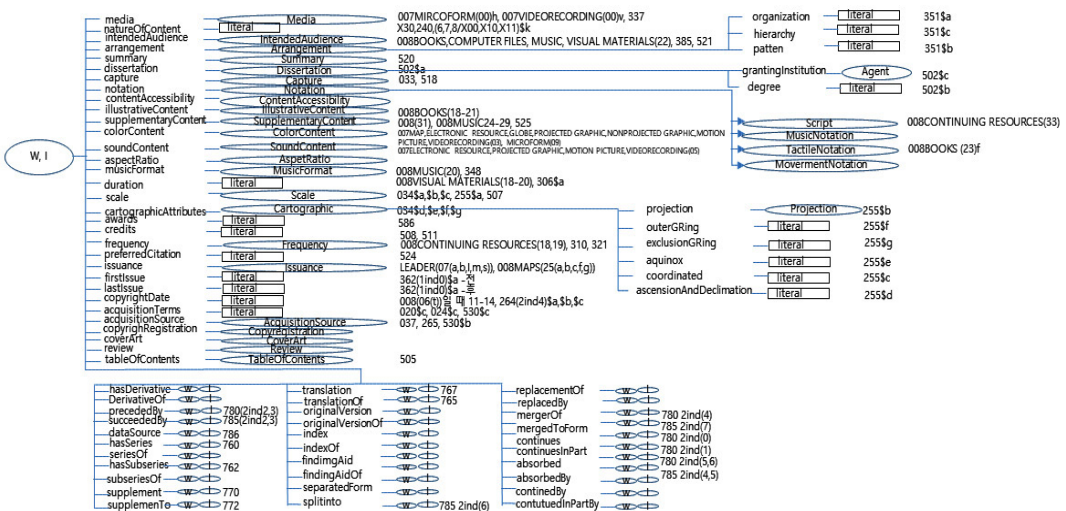
저작과 인스턴스에 공통으로 사용되는 하위클래스는 없으며, 속성은 bf:media부터 bf:continuedIn

PartBy 까지 총 63개이다. 이 중 MARC 21과 매핑되지 않는 속성은 bf:notation의 하위클래스 bf:MusicNotation과 bf:MovementNotation, bf:contentAccessibility, bf:aspectRatio, bf:copyrightRegistration, bf:coverArt, bf:review, bf:hasDerivative, bf:derivativeOf, bf:seriesOf, bf:subseriesOf, bf:originalVersion, bf:originalVersionOf, bf:index, bf:indexOf, bf:findingAid, bf:findingAidOf, bf:separatedForm, bf:replacementOf, bf:replacedBy, bf:continuedBy, bf:continuedInPartBy이다.

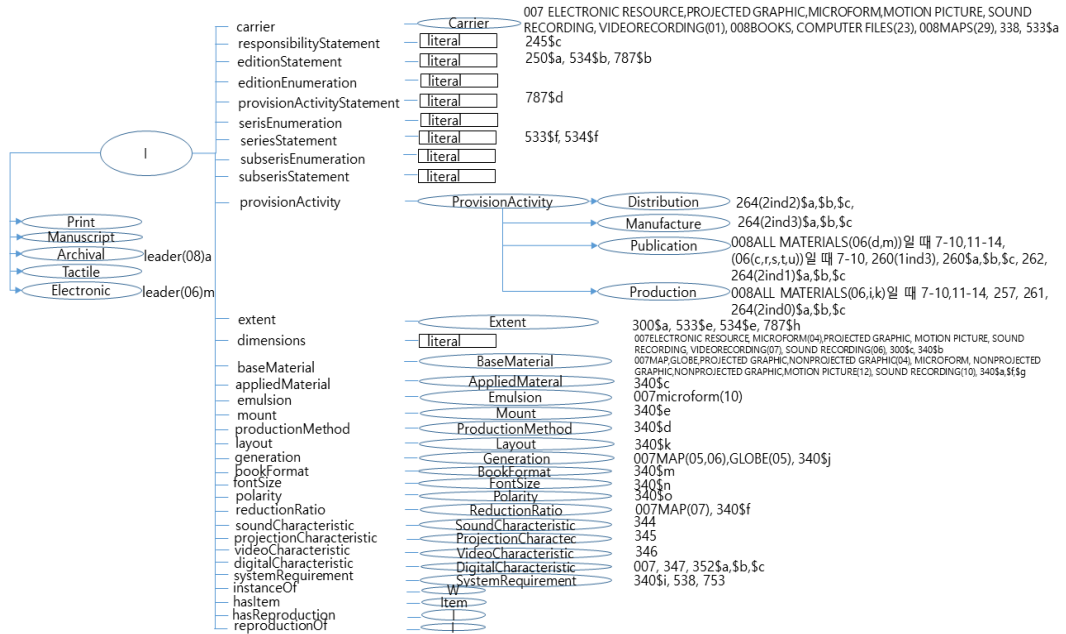
3) 인스턴스 어휘의 매핑

BIBFRAME 어휘 중 인스턴스에만 사용되는 속성과 MARC 21과의 매핑결과는 <그림 3>과 같다.

인스턴스의 하위클래스는 bf:Print, bf:Manuscript, bf:Archival, bf:Tactile, bf:Electronic이다. bf:Archival 문서는 리더(08제어유형)a문서, bf:Electronic는 리더(06레코드유형)m전자자료와 매



<그림 2> BIBFRAME 저작/인스턴스 공통어휘와 MARC 21의 매핑 결과



<그림 3> BIBFRAME 인스턴스 어휘와 MARC 21의 매핑 결과

핑되며, 하위클래스인 bf:Print, bf:Manuscript, bf:Tactile는 MARC 21과 매핑되지 않는다. 속성은 bf:carrier부터 bf:reproductionOf까지 총 32개이며, 이중 bf:editionEnumeration, bf:series Enumeration, bf:subseriesEnumeration, bf:subseries Statement, bf:instanceOf, bf:hasItem, bf:has Reproduction, bf:reproductionOf는 매핑되지 않는다.

4) 아이템 어휘의 매핑

BIBFRAME 어휘 중 아이টে에만 사용되는 속 성과 MARC 21과의 매핑 결과는 <그림 4>와 같다.

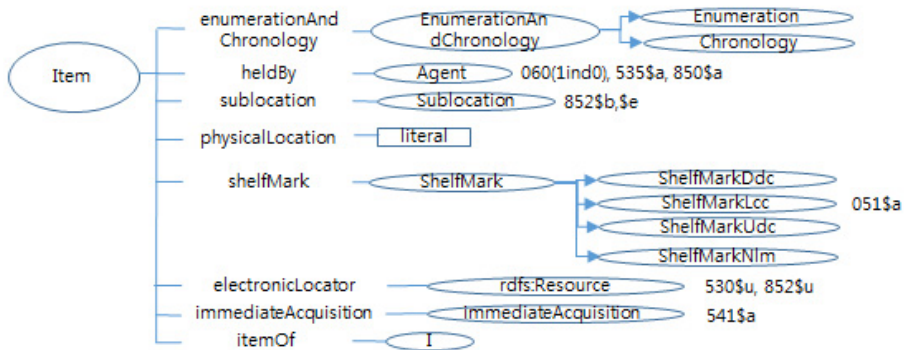
아이টে에만 사용되는 하위클래스는 없으며, 속 성은 bf:enumerationAndChronology, bf:heldby, bf: sublocation, bf:physicalLocation, bf:shelf Mark, bf:electronicLocator, bf:immediateAcquisition,

bf:itemOf 총 8개이다. 이중 MARC 21과 매핑되 지 않는 속성은 bf:enumerationAndChronology 의 속성값 클래스 bf:EnumerationAndChronology 의 하위클래스 bf:Enumeration과 bf:Chronology, 속성값이 문자열인 bf:physicalLocation, bf:shelf Mark의 속성값 클래스 bf:ShelfMark의 하위클 래스 bf:ShelfMarkDdc, bf:ShelfMarkUdc, bf: ShelfMarkNlm, 그리고 bf:itemOf이다.

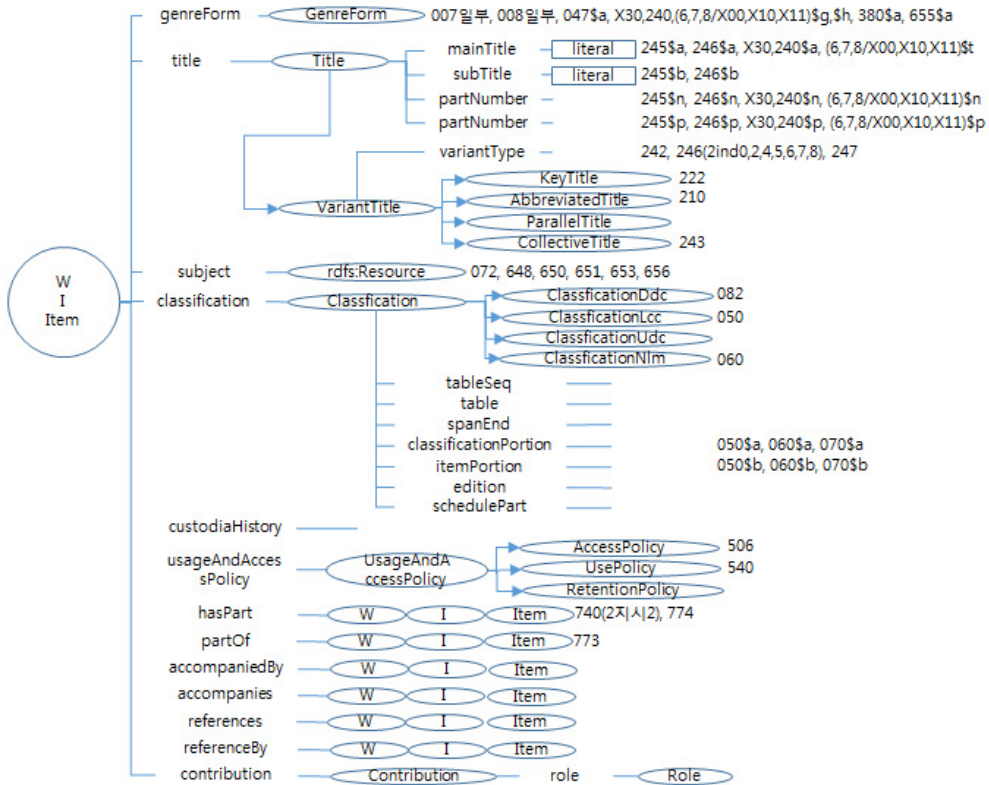
5) 저작, 인스턴스, 아이템 공통사용 어휘의 매핑

BIBFRAME 어휘 중 저작, 인스턴스, 아이টে에 공통으로 사용되는 속성과 MARC 21과의 매핑 결 과는 <그림 5>와 같다.

저작, 인스턴스, 아이টে에 공통으로 사용되는 하 위클래스는 없으며, 속성은 bf:genreForm, bf:title, bf:subject, bf:classification, bf:custodiaHistory,



<그림 4> BIBFRAME 아이템 어휘와 MARC 21의 매핑 결과



<그림 5> BIBFRAME 저작, 인스턴스, 아이템 공통사용 어휘와 MARC 21의 매핑 결과

bf:usageAndAccessPolicy, bf:hasPart, bf:partOf, bf:accompaniedBy, bf:accompanies, bf:references, bf:referencedBy, bf:contribution

으로 총 13개이다. 이 중 MARC 21과 매핑되지 않는 속성은 문자열을 속성값으로 갖는 속성 bf:custodialHistory, 속성 bf:title의 속성값 Title의

하위클래스 bf:VariantTitle의 속성값 중에서 클래스 bf:ParalleTitle, 속성 bf:classification의 속성값 bf:Classification의 하위클래스 bf:ClassificationDdc, bf:ClassificationUdc와 속성 bf:tableSeq, bf:table, bf:spanEnd, bf:edition, bf:schedulePart, 속성 bf:usageAndAccessPolicy의 속성값 bf:UsageAndAccessPolicy의 하위클래스 bf:RetentionPolicy, 속성 bf:contribution의 속성값 bf:Contribution의 속성인 bf:Role의 bf:속성값 role 그리고 클래스 Work, Instance, Item와 상호 연결되는 속성들인 bf:accompaniedBy, bf:accompanies, bf:references, bf:referenceBy는 MARC 21과는 매핑되지 않는다.

클래스 bf:Work의 속성으로서 bf:genreForm은 속성값으로 클래스 bf:GenreForm을 가지는데, <표 5>와 같이 007필드와 008필드 중 일부 항목 그리고 047\$a작곡형식부호, X30, 240,(6,7,8/X00,X10,X11)\$g기타정보, \$h자료유형표시, 380\$a 저작의 형식, 655\$a자료 또는 핵심어의 유형(장르)/형식과도 매핑된다. bf:Instance의 속성으로서 bf:genreForm 역시 속성값으로 클래스 bf:Genre Form을 가지고 008필드 CONTINUING RESOURCES (21계속자료의 유형)와 매핑된다.

6) 불특정 어휘의 매핑

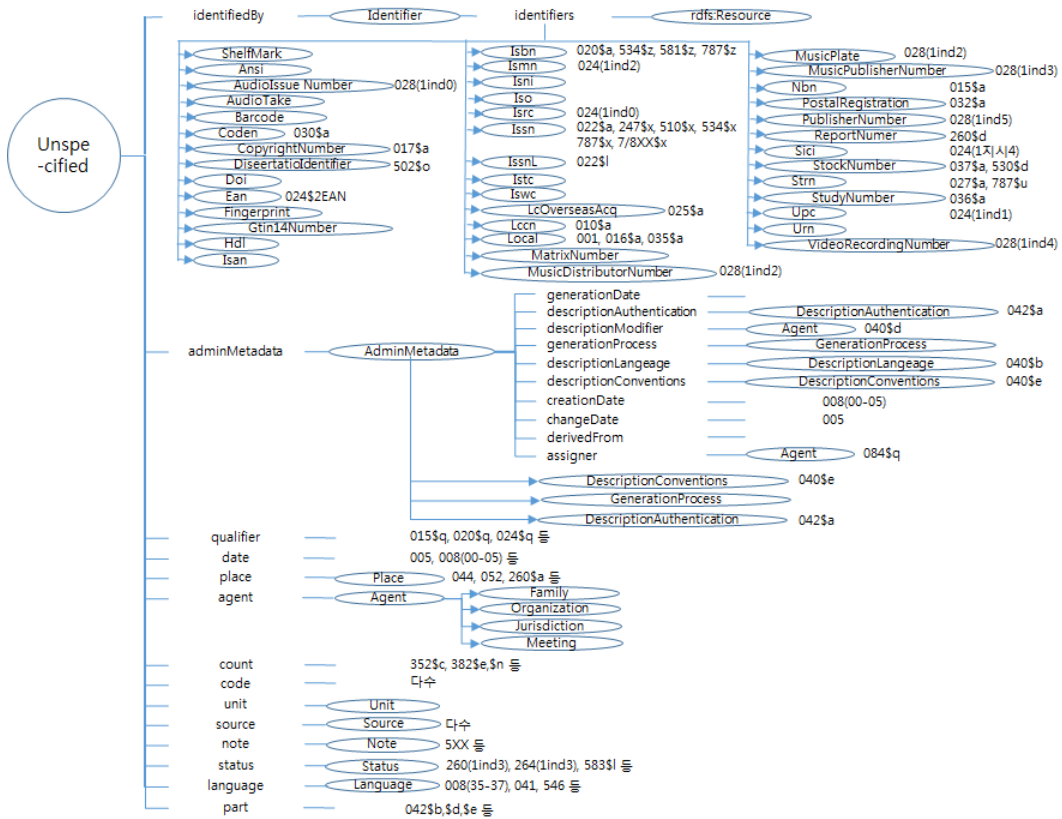
BIBFRAME 어휘 중 다양한 클래스와 연결되

어 사용되는 불특정한 성격의 속성과 MARC 21과의 매핑 결과는 <그림 6>과 같다. 이들 속성은 저작, 인스턴스, 아이템 클래스와 연결된 다양한 클래스의 속성값으로서 조합되어 매핑에 매우 많이 활용된다. bf:source의 경우 1절 <표 4>의 예와 같이 043 필드(W-geographicCoverage-GeographicCoverage-source)를 비롯하여 015 필드(I-identifiedBy-Nbn-source), 041 필드(W-language-Language-source) 등 다수의 필드와 매핑된다.

여기에 포함되는 속성은 bf:identifiedBy, bf:adminMetadata, bf:qualifier, bf:data, bf:place, bf:agent, bf:count, bf:code, bf:unit, bf:source, bf:note, bf:status, bf:language, bf:part로 총 14개이다. 이중에서 클래스를 속성값으로 갖는 bf:unit, 속성 bf:identifiedBy의 속성값 bf:Identifier는 속성 bf:identifies의 41개 하위클래스 가운데 bf:ShelfMark, bf:Ansi 등 15개, bf:adminMetadata의 속성값 bf:AdminMetadata의 속성 bf:generationDate, bf:generationProcess, bf:derivedFrom과 하위클래스 bf:GenerationProcess은 MARC과 매핑되지 않는다. 특이한 사항으로는 agent의 속성값 Agent는 4개의 하위클래스 Family, Organization, Jurisdiction, Meeting을 가지는데, 이는 단순 매핑의 방법이 아닌 별도의 프로그램에 의하여 변환 처리된다.

<표 5> W - genreForm - GenreForm과 매핑되는 007필드, 008필드 내역

007필드	008필드
<ul style="list-style-type: none"> ▪ MAP(01특정자료종별) ▪ GLOBE(00자료범주표시, 01특정자료종별) ▪ NONPROJECTED GRAPHIC(01특정자료종별) ▪ MOTION PICTURE(09제작요소) 	<ul style="list-style-type: none"> ▪ BOOKS(24-25내용형식, 29회의간행물, 30기념논문집, 33문학형식, 34전기) ▪ COMPUTER FILES(26컴퓨터파일의 유형) ▪ MAPS(33-34형태상의 특징) ▪ MUSIC(18-19작곡형식, 30-31녹음자료의 내용) ▪ VISUAL MATERIALS(33시청각자료의 유형)



<그림 6> BIBFRAME 불특정 어휘와 MARC 21의 매핑 결과

5. 결론

이 연구의 연구결과는 다음과 같다. 첫째, BIBFRAME 2.0에서는 1.0의 일부 속성이 클래스로 변경되어 186개의 클래스, 195개의 속성이 정의되었다. 또한 BIBFRAME 2.0에서 속성값이 문자열을 갖는 사례와 속성값이 클래스를 가지면서 또 다른 속성값을 가지는 사례를 확인하였다 (<그림 1> - <그림 6> 참고). 도출된 결과와 같이 BIBFRAME 2.0의 어휘들은 RDF 기반으로 기술되기 때문에 확장과 추론이 가능한 새로운 형태의 목록데이터를 생성할 수 있다. KORMARC 레코드도 확장과 추론의 목록으로 변화하기 위해

서는 <그림 1> - <그림 6>에서와 같이 RDF 기반으로 표현하되, 속성값으로 문자열을 갖는 어휘, 클래스를 갖는 어휘를 고려하고, KORMARC의 특징이 잘 반영되도록 해야 한다.

둘째, MARC 21과 BIBFRAME 2.0 간의 1대다 매핑사례를 도출하였다. 먼저 BIBFRAME 어휘를 기준으로 한 MARC 21과의 1대다 매핑 사례는 <그림 1>부터 <그림 6>에서 확인할 수 있다. 1대다 매핑사례를 범주화하면, BIBFRAME 어휘에 007, 008 필드와 같이 MARC 21의 부호화한 값과 서지기술 필드가 매핑되는 경우, 주기사항 이외의 서지기술 필드와 주기사항에 해당하는 필드가 함께 매핑되는 경우, 주기사항 이외의

서지기술 필드가 함께 매핑되는 경우이다. 예를 들어 <그림 2>의 bf:IntendedAudience는 저작 혹은 인스턴스에 공통으로 사용되는 어휘로서 bf:Work 혹은 bf:Instance - bf:intendedAudience - bf:IntendedAudience로 변환되는데, MARC 21 중 008 BOOKS, COMPUTER FILES, MUSIC, VISUAL MATERIALS의 (22) 이용 대상자 수준은 저작에, 385는 저작에, 521은 인스턴스에 매핑된다. MARC 21에서 상이한 필드로 분산되어 있지만 유사한 의미를 갖는 표시기호/식별기호/지시기호가 BIBFRAME에서 한 어휘와 매핑된다. 또한 MARC 21을 기준으로 한 BIBFRAME과의 1대다 매핑 사례로는 [부록 1]과 같다. 예를 들어 리더 06 레코드 유형 필사악보(d)는 bf:Work - rdf:type - bf:NotatedMusic과 bf:Instance - rdf:type - bf:Manuscript에, 필사지도(f)는 bf:Work - rdf:type - bf:artography과 bf:Instance - rdf:type - bf:Manuscript에 필사문자자료(t)는 bf:Work - rdf:type - bf:Text와 bf:Instance - rdf:type - bf:Manuscript로 변환된다. 이는 MARC 21의 필사자료로 처리된 부호값이 BIBFRAME의 관점에서는 중복된 의미를 가지므로 저작이나 인스턴스로 처리가 가능하다. 따라서 MARC 21과 BIBFRAME 2.0 간의 1대다 매핑사례를 면밀히 검토할 필요가 있다. 아울러 KORMARC을 위한 변환 명세에서 어떻게 처리할 것인가에 대한 분명한 가이드라인이 제시되어야 한다. 이러한 부분을 충분히 검토하지 않으면, 메타데이터 매핑 시 1대다 매핑에 따라 야기되는 완벽한 상호 변환, 변환 후 재변환이 어렵다.

셋째, MARC 21과 BIBFRAME 2.0 간에 매핑되지 않는 사례를 도출하였다. BIBFRAME 기준으로 매핑되지 않는 클래스와 속성의 사례는 <그림 1>부터 <그림 6>과 같다. 그림에서 매핑되지

않은 대부분의 어휘는 추상적인 수준인 저작과 물리적인 수준인 인스턴스로 명확하게 구분하고, 클래스 간에 존재하는 관계를 명확하게 선언함으로 발생한 것으로, 이 중 일부 어휘는 RDA에서 파생된 것이며, BIBFRAME에서 새롭게 추가된 어휘도 있다. 따라서 BIBFRAME 2.0 어휘와 RDA의 적합성에 대한 연구가 진행되어야 할 것이다. 또한 매핑되지 않는 어휘를 KORMARC 레코드의 데이터 모델 설계 시 어떻게 처리할 것인가에 대한 고민이 필요하다.

특히 MARC 21 기준으로 매핑되지 않거나 고려되지 않는 표시기호/식별기호/지시기호의 사례는 [부록 2]와 같다. 매핑되지 않는 요소 중 'nac'라는 기호의 요소들은 LC 레코드에서 사용되지 않는 요소들이다. 또한 'ignore'에는 다수의 표시기호/식별기호/지시기호가 포함된다. 이 중 일부는 리더 레코드길이(00-04), 문자부호화체계(09), 지시기호 자리수(10), 식별기호 자리수(11), 데이터 기본번지(12-16), 엔트리맵(20-23) 등과 같이 자동 생성되거나 MARC 레코드 가독을 위해 필요한 데이터, 007 MAP(01)특정자료종별의 u 자료를 특정화하지 않음, z 기타, | 부호화하지 않음과 같이 변환의 의미가 없는 경우도 있다. 그러나 nac나 ignore로 표시된 요소 중 일부는 어떤 기관에게는 중요한 요소일 수도 있다. 미래에 사용되지 않을 것으로 예상되어, 생략된 요소의 변환 부분도 마찬가지이다. 예를 들어 변환 명세에서는 031이 생략되었지만, LD4P(Linked Data for Production)의 Linked Data for Performed Music에서는 이 부분의 변환 필요성을 언급하였다(Lorimer, 2018). 이와 같이 현재 매핑대상으로 고려되지 않은 필드들을 검토하여 추가적인 BIBFRAME 클래스와 속성의 개발이 필요하지 살펴보아야 한다. 또한 KORMARC 레코드를 기준으로 변환되지 않을 표시기호/지시기호/식별기호의 검토가 필요하다.

아울러 BIBFRAME 구조와 매핑 자체의 문제도 몇가지 발견되었다. 예를 들어 BIBFRAME 구조와 매핑은 특정 총서 사항, 식별자 등과 총서 번호와의 연결 방법을 제공하지 않고 있으므로, 다수 총서로 발행된 인스턴스는 데이터와 혼동되는 결과를 가져온다. 또한 MARC 852에서 bf:Item으로의 변환과 연속간행물을 위한 MARC 86X의 변환은 아직도 수작업 처리과정을 거쳐야 한다. 또한 로컬 도서관에서 정의한 9xx, x9x, xx9 필드를 위한 변환 명세도 마련되어야 한다. bf:Work과 관련된 bf:adminMetadata 부분도 더 고민해야 한다.

이 연구에서 살펴본 MARC 21과 BIBFRAME 2.0의 변환 명세에 대한 기초연구를 통해, BIBFRAME 2.0과 MARC 21 간의 매핑에 대해 일부 이해할 수 있었다. 그러나 변환 명세를 구체적으로 이해하기 위해서는 더 심도있는 연구가 진행되어야 한다. 또한 변환도구, 실제적인 데이터 변환 등 다양한 관점에서의 추가적인 논의도 필요하다. 특히 LC의 변환 명세는 LC의 레코드를 대상으로 한 변환 명세이기 때문에, KORMARC 레코드를 대상으로 한 변환 명세에 대한 연구가 필요하다. 궁극적으로는 KORMARC를 위한 데이터 모델 설계에 대한 연구가 필요하다.

참 고 문 헌

- 국립중앙도서관 (2018). Linked Open Data. [On-line], Available: <https://lod.nl.go.kr/>
- 국회도서관 (2018). 국가학술정보융합데이터 서비스. [On-line], Available: <http://cloud.nanet.go.kr/lod/>
- 박옥남 (2015). 고전문학의 대중화를 위한 온톨로지 설계에 관한 연구. <한국비블리아학회지>, 26(3), 187-290.
- 박옥남 · 오정선 (2014). 링크드 데이터 환경에서의 서지기술형식 BIBFRAME과 그 활용에 대한 고찰. <한국비블리아학회지>, 25(4), 235-263.
- 박지영 (2013). 차세대 서지 기술 형식으로서의 BIBFRAME 모형 연구. <한국정보관리학회 학술대회 논문집>, 101-104.
- 박지영 (2016). 서지프레임워크를 활용한 공공도서관 서지데이터와 서비스 데이터의 연계. <정보관리학회지>, 3(1), 293-316.
- 박진호 (2013). 국가서지정보 Open Public Data Platform 구축. 공공정보 활용 사업화 전략 세미나 발표자료. 2013년 3월 29일, 대전시청.
- 이미화 (2017). BIBFRAME 2.0 특징 분석 및 BIBFRAM 구축시 고려사항에 관한 연구. <한국도서관 · 정보학회지>, 48(4), 108-127.
- 이미화 (2018). BIBFRAME 환경 하에서 목록사서의 대응 전략에 관한 연구. <한국도서관 · 정보학회 하계학술대회>, 171-195.
- 이혜원 (2018). 서지데이터 공유 및 활용을 위한 BIBFRAME 2.0 수용 방안 연구. <한국도서관 · 정보학회 하계학술대회>, 73-90.
- Hansen, C., & Mueller, V. (2015). From MARC to BIBFRAME: An Introduction. May 13, 2015. [On-line], Available: <http://www.ala.org/alcts/confevents/upcoming/webinar/051315>

- KERIS (2018). RISS Linked Open Data. [On-line], Available: <http://data.riss.kr/>
- KISTI (2018). KISTI Linked Open Data. [On-line], Available: <http://lod.ndsl.kr/>
- LC Bibliographic Framework Initiative (2012). BIBFRAME 1.0 Vocabulary. [On-line], Available: <http://bibframe.org/vocab-list/>
- LC Bibliographic Framework Initiative (2016). Overview of the BIBFRAME 2.0 Model. [On-line], Available: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>
- LC Bibliographic Framework Initiative (2017). MARC 21 to BIBFRAME 2.0 Conversion Specifications. [On-line], Available: <http://www.loc.gov/bibframe/mtbf/>
- LC Bibliographic Framework Initiative (2018a). LC BIBFRAME FAQs. [On-line], Available: <https://www.loc.gov/bibframe/faqs/>
- LC Bibliographic Framework Initiative (2018b). Overview of the BIBFRAME 2.0 Model. [On-line], Available: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>
- LC Bibliographic Framework Initiative (2018c). BIBFRAME vocabulary. [On-line], Available: <http://id.loc.gov/ontologies/bibframe-category.html>
- Lorimer, N. (2017). Linked Data for Performed Music, [On-line], Available: <https://www.loc.gov/bibframe/news/bibframe-update-mw2017.html>
- Lorimer, N. (2018). Performed Music Ontology. [On-line], Available: <https://wiki.duraspace.org/display/LD4P/Performed+Music+Ontology>
- McGrath, K. (2017). A Brief Introduction to BIBFRAME 2.0. Oregon Library Association Conference April 20, 2017. [On-line], Available: http://pages.uoregon.edu/kelleymp/publications/McGrath_Bibframe_OLA_2017.pdf
- Miller, M. (2017). Diagramming Bibframe 2.0. [On-line], Available: <https://medium.com/@thisismattmiller/diagramming-bibframe-2-0-4c1696653052>
- Sanderson, R. (2015). Analysis of the BIBFRAME Ontology for Linked Data Best Practices. [On-line], Available: https://docs.google.com/document/d/1dly-FgQsH67Ay0T000ulhyRiKjpf_I0AVQ9v8FLmPN0/edit
- University of Washington Libraries (2018). Mapping of RDA Core to BIBFRAME 2.0. [On-line], Available: <https://www.lib.washington.edu/msd/pubcat/ld/rda-bibframe-mapping>
- Xu, A., Hess, K., & Akerman, L. (2017). From MARC to BIBFRAME 2.0: Crosswalks. *Cataloging & Classification Quarterly*, Published online: 13 Nov 2017, [On-line], Available: <https://doi.org/10.1080/01639374.2017.1388326>

[부록 1] MARC 21 필드와 BIBFRAME 어휘간의 1대대 매핑

Fields	Conversion 1	Conversion 2
leader 06 - Type of record		
d - Manuscript notated music	W - rdf:type - NotatedMusic	I - rdf:type - Manuscript
f - Manuscript cartographic material	W - rdf:type - Cartography	I - rdf:type - Manuscript
t - Manuscript language material	W - rdf:type - Text	I - rdf:type - Manuscript
X30,240,etc.-Uniformtitles		
\$k - Form subheading	W - natureOfContent - literal	W - genreForm - GenreForm - rdfs:label- literal
\$l - Language of a work	W - language - Language - rdfs:label - literal	W - translationOf
787 - OTHER RELATIONSHIP ENTRY		
\$a - Main entry heading	W-contribution-bflc:PrimaryContribution-agent-Agent-rdfs:label-literal	I-contribution-bflc:PrimaryContribution-agent-Agent-rdfs:label-literal
\$c - Qualifying information	W - title - Title - qualifier - literal	I - title - Title - qualifier - literal
\$i - Relationship information	W - bflc:relationship - bflc:Relationship - <relationship property based on tag and indicator as indicated above> - URI : bflc:relation - rdfs:label content of \$i	I- bflc:relationship - bflc:Relationship - <relationship property based on tag and indicator as indicated above> - URI : bflc:relation - rdfs:label content of \$i

[부록 2] BIBFRAME 어휘와 매핑되지 않는 MARC 21 표시기호/식별기호/지시기호

<nac>

leader(19), 007ELECTRONIC RESOURCE(10, 11, 12, 13), 007MOTION PICTURE(13, 14, 15), 007SOUND RECORDING(13), 007REMOTE-SENSING IMAGE, 007TACTILE MATERIAL, 007KIT, 007NOTATED MUSIC, 007text(00), 007UNSPECIFIED, 008ALL MATERIALS(39), 008MUSIC(21, 33), 013, 031, 038(\$e,\$2), 050(lind, 2ind4), 052(lind1,\$2), 055(lind, 2ind3, 4, 5, 6, 7, 8, 9, \$2), 060(2ind4), 061, 070(lind), 071, 080, 082(\$m), 083, 085, 258, 307, 340(\$0,\$3), 342, 343, 344(\$0), 345(\$0), 346(\$0), 347(\$0), 350(\$b, \$6), 355, 357, 363, 365, 366, 370, 377, 380(\$0), 381, 382(\$0), 383(\$e, \$2), 384, 388, 535(\$g), 505(lind, \$u), 506(lind, \$2), 514, 516(lind), 518(\$0, \$2), 520(lind, \$2), 521(lind3, 4, 8), 522(lind), 524(lind, \$2, \$3, \$6), 526, 535(\$g), 538(\$i), 540(\$b), 541(lind, \$b, \$6), 542, 544(lind), 552, 556(\$z), 561(lind, \$u), 562, 563(\$u, \$3, \$6), 565, 567, 581(lind), 583(lind, \$b, \$d, \$e, \$f, \$i, \$j, \$n, \$o, \$x, \$6), 584, 648(\$3, \$6), 650(lind, \$g, \$4), 651(\$g, \$3, \$4), 653(lind), 654, 655(lind0, \$b, \$c), 657, 658, 662(\$e, \$4), 751, 754, 841, 842, 843, 844, 845, 851, 852(lind, 2ind, \$c, \$d, \$f, \$g, \$h, \$i, \$j, \$k, \$l, \$m, \$p, \$q, \$s, \$t, \$2, \$3, \$6), 853, 854, 855, 863, 864, 865, 866, 867, 868, 876, 877, 878, 882, 883, 884, 886, 887

<ignore>

leader(00-04, 05(d), 06(b,h,n), 08(#), 09, 10, 11, 12-16, 20-23), 007MAP(01(u, z, |), 03(|), 04(q, r, s, t, u, y, z, |), 05(n, u, |), 06(u, z, |), 07(n, |)), 007ELECTRONIC RESOURCE(01(u, |), 03(n, u, z, |), 04(n, z, |), 05(u, |), 06-08(nnn, ---, ||), 09(u, |)), 007GLOBE(01(u, z, |), 03(|), 04(u, z, |), 05(n, u, z, |)), 007PROJECTED GRAPHIC(01(u, z, |), 03(n, u, z, |), 04(u, z, |), 05(u, |), 06(#, u, z, |), 07(u, z, |), 08(#, u, z, |)), 007MICROFORM(01(u, z, |), 03(u, |), 04(u, z, |), 05(u, |), 09(u, z, |), 10(n, u, z, |), 11(u, |), 12(n, u, z, |)), 007NONPROJECTED GRAPHIC(01(u, z, |), 03(u, z, |), 04(u, z, |), 05(#, u, z, |)), 007MOTION PICTURE(01(u, z, |), 03(n, u, z, |), 04(u, z, |), 05(u, |), 06(#, u, z, |), 07(u, z, |), 08(n, u, z, |), 09(u, z, |), 10(n, u, z, |), 11(u, z, |), 12(n, u, z, |), 16(n, u, |)), 007SOUND RECORDING(01(u, z, |), 02, 03(u, z, |), 04(u, z, |), 05(n, u, z, |), 06(n, u, z, |), 07(n, u, z, |), 08(n, u, z, |), 09(n, u, z, |), 10(u, z, |), 11(n, u, |), 12(n, u, z, |)), 007VIDEORECORDING(01(u, z, |), 02, 03(n, u, z, |), 04(u, z, |), 05(u, |), 06(#, u, z, |), 07(u, z, |), 08(u, z, |)), 008ALL MATERIALS(06(b, n, |), 38), 008BOOKS(18-21(#, |), 22(#, |), 23(#, |), 24-27(#, |), 28(#, u, z, |), 29(0, |), 29(0, |), 30(0, |), 31(0, |), 32, 33(0, u, |), 34(#, |)), 008COMPUTER FILES(26(u, z, |), 29-34), 008MAPS(18-21(#, z, |), 22-23(##, zz, |), 24, 25(u, z, |), 26-27, 30, 32, 33-34(#, z, |)), 008MUSIC(18-19(zz, ||), 20(n, u, z, |), 24-29(#, z, |), 30-31(#, n, z, |)), 008CONTINUING RESOURCES18((u, z, |), 19(|), 21(#, |), 33(#, u, z, |), 34(|)), 008VISUAL MATERIALS(18-20(---, nnn, ||), 33(z, |), 34(n, u, z, |)), 008MIXED MATERIALS(18-22, 24-34), 010(\$b), 015(\$6), 016(1차시#7, \$6), 017(2ind, \$z, \$2, \$6), 018, 020(\$6), 022(lind, \$6), 024(2ind, \$6), 026(\$6), 027(\$q, \$6), 028(2ind, \$6), 030(\$6), 031, 033(2ind#, 0, \$0, \$6), 034(lind, \$a(b, c), \$h, \$j, \$k, \$m, \$n, \$p, \$r, \$s, \$t, \$x, \$y, \$z, \$0, \$2, \$6), 035(\$6), 036(\$6), 037(lind#, \$6), 038(\$6), 040(\$6), 041(1차시#, 0, \$6), 043(\$6), 044(\$6), 045(lind#, 0, 1, \$c, \$6), 047(2ind7), 048(\$2), 050(\$3, \$6), 052(lind7), 055(\$6), 060(lind1), 066, 072(2ind7) 082(lind7, 2ind4), 086(lind#), 088(\$6), 100/600/700/800(lind0, 1, 2, 2ind#), 110/610/710/810(lind0, 2, 2ind#), 111/611/711/811(lind0, 1, 2, 2ind#), 210(lind, 2ind0), 242(\$c, \$h), 243(\$d, \$f, \$g, \$h, \$k, \$l, \$m, \$n, \$o, \$p, \$r, \$s), 245(\$f, \$g, \$k, \$s), 246(2ind#, 1, 3, \$g, \$h, \$i), 247(\$b, \$h), 260(1차시#, 2), 262(\$k, \$l, \$6), 263(\$6), 264(lind#, 2), 265(\$6), 270, 340(\$h), 352(\$d, \$e, \$f, \$g, \$i), 362(\$z), 382(lind, 2ind), 490(lind, \$1), 505(2ind), 510(lind, \$u, \$3), 511(\$0), 521(lind#), 533(\$7), 534(\$1, \$o), 545(lind#), 556(lind), 586(lind), 588(lind), 648-662(2ind4), 655(lind#), 656(2ind), 720(lind#, 2, \$6), 740(lind#, 2차시0, 1), 753(\$6), 787(\$j, \$o, \$p, \$q, \$4, \$7)

Study on Conversion between MARC 21 and BIBFRAME 2.0

Sung-Sook Lee, Ji-Won Lee

Dept. of Library and Information Science, Chungnam National University

Dept. of Library Science, Daegu Catholic University

Providing a conversion specification is the first process to converting MARC format records to linked data based on BIBFRAME. This study examined BIBFRAME 2.0 determine the implications for converting KORMARC records. the 'MARC 21 to BIBFRAME 2.0 conversion specification' by LC published 2017 was reviewed and the properties and classes were categorized based on the work, instance, and item, which are the BIBFRAME 2.0 core class. BIBFRAME 1.0 defined 53 classes and 289 properties and BIBFRAME 2.0 performed 186 classes and 195 properties. This study identified some of the properties of the BIBFRAME 1.0 that changed to classes of BIBFRAME 2.0. The property values that were defined as 'literal' or 'class' in the vocabularies of BIBFRAME 2.0 based on RDF, which could produce new type of bibliographic records that could be extended and inferred, were also identified. The mapping results between BIBFRAME 2.0 classes/properties and MARC 21 tags/subfields/indicators were investigated using 21 files of 'MARC 21 to BIBFRAME 2.0 conversion specification'. This study found that of the one-to-many mapping cases of BIBFRAME to MARC 21 and the one-to-many mapping cases of MARC 21 to BIBFRAME, multiple classes/properties were not mapped on a BIBFRAME basis. Furthermore, multiple tags/subfields/indicators were not mapped on a MARC 21 basis. These results can be used as basic data for the mapping KORMARC to BIBFRAME.

Keywords: BIBFRAME, MARC 21, Conversion Specification, Mapping, Crosswalk, Work, Instance, Item