

디지털화 기록 관리를 위한 메타데이터 요소(안) 설계

A Study on Designing Metadata Elements for the Management of Digitized Records

서지인(Jiin Seo)¹, 노지현(Jee-Hyun Rho)²

Email: archivething@naver.com, jhrho@pusan.ac.kr



1 제 1저자 부산대학교 일반대학원 문헌정보학과 기록관리학전공 박사과정 수료
2 교신저자 부산대학교 사회과학대학 문헌정보학과 교수

논문접수 2022.10.17
최초심사 2022.10.20
게재확정 2022.11.15

ORCID

Jiin Seo
https://orcid.org/0000-0002-3448-8927

Jee-Hyun Rho
https://orcid.org/0000-0002-2760-8732

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

초 록

이 연구는 디지털화 과정을 통해 생산되는 디지털화 기록의 체계적 관리와 품질 유지를 위해 필요한 메타데이터 요소(안)를 제안하는데 목적을 두고 있다. 메타데이터 요소(안)의 개발에 앞서 이 연구에서는 디지털화 기록의 정의와 특성을 확인하고, 원천 기록의 사본이자 대체물로서 디지털화 기록을 규정하여 원천 기록과의 관계를 정립하였다. 다음으로, 국내 기록관리 메타데이터 표준과 미국 및 영국의 사례를 분석하여 디지털화 기록에 대한 메타데이터 요소 비교 및 요소 설계시 고려할 사항을 도출하였다. 이러한 분석결과를 토대로 하여, 디지털화 기록 관리를 위한 메타데이터의 설계방향은 첫째, 원천 기록과 디지털화 기록의 관계 정립, 둘째, 자원 유형별 특성에 따른 자동추출 메타데이터 요소 확대, 셋째, 업무과정 및 프로젝트에 대한 메타데이터 작성 등 3가지로 정리되었다. 이에 근거하여 디지털화 기록의 메타데이터를 디지털화 기록 메타데이터와 디지털화 프로세스 메타데이터로 개체를 구분하고, 해외 사례를 참조하여 각 개체별로 추가되어야 할 요소를 제안하였다.

ABSTRACT

This study aims to design the metadata elements for managing and maintaining digitized records resulting from digitization. The digitized records were first defined and characterized, then assigned as copies and surrogates for source records. Next, the factors to be considered when designing metadata elements for digitized records were determined by comparing standards from different countries like the United States and the United Kingdom.

As a result of the comparison, establishing a relationship between digitized records and source records, expanding the automatically extracted metadata elements following resource characteristics, and creating metadata for digitization processes and projects were set to be the key tasks of metadata design for managing digitized records.

Furthermore, the metadata elements for digital records and digitization processes were designed individually, and the elements to be added for each entity were proposed by referring to standards from other countries.

Keywords: 디지털화, 디지털화 기록, 디지털화 프로세스, 메타데이터, 메타데이터 설계
Digitization, Digitized Records, Digitization Process, Metadata, Metadata Design

1. 서론

1.1 연구배경 및 목적

많은 공공기관들이 소장 중인 비전자기록의 디지털화 및 데이터베이스 구축사업을 추진하고 있다. 디지털화 사업의 주요 목적은 중요 문서의 훼손과 멸실을 방지하여 영구적으로 안전하게 보존하면서도 이를 신속하게 찾을 수 있는 기반을 마련하여 행정효율을 증대시키기 위함이다. 이렇듯 비전자기록의 디지털화는 실물 기록의 열람과 활용을 줄여 안전한 관리환경을 마련함과 동시에 데이터베이스 구축을 통해 공간에 구애받지 않는 자유로운 접근을 가능하게 함으로써 이용가능성을 확보하기 위한 작업이다.

다른 측면에서 비전자기록의 디지털화는 전자기록 중심의 기록관리 체제에 맞춰 여러 유형의 기록을 통합적으로 관리하기 위한 노력이라고도 볼 수 있다. 온-나라 문서시스템과 같은 전자문서시스템나 업무관리시스템을 도입하기 전 공공기관은 다양한 유형의 비전자기록을 생산하였으며, 시스템 도입 이후에도 여전히 비전자기록을 생산하고 있다. 여기에, 각종 행정정보시스템의 개발과 도입으로 데이터세트와 같은 새로운 유형의 기록을 생산하거나 취급하는 등 기록관리환경은 과거보다 많이 복잡해지고 있다. 복잡한 환경 속에서 다양한 기록 유형을 체계적으로 관리하기 위해서는 각 기록의 연결, 즉 결합관계(archival bond)를 유지하는 것이 중요하며, 이를 구현하기 위한 필수적인 도구가 바로 메타데이터라 할 수 있다.

비전자기록의 디지털화 사업을 추진하는 과정에서 공공기관에서는 디지털화 사업 착수 및 용역 계약 시 관련 근거로서 국가표준 및 국가기록원에서 제정한 공공표준을 활용한다. 관련 국가표준은 「KS X ISO TR 13028 문헌정보-기록의 디지털화 이행 지침」이며, 공공표준은 「NAK 23:2017 기록물 목록 작성 및 디지털화 작업 지침」과 「NAK 26:2018(v2.0) 기록물 디지털화 기준」이다. 하지만 세 표준 모두 디지털화 자료의 품질 확보 및 신뢰성 유지에 집중된 것으로, 메타데이터에 관한 내용은 「KS X ISO TR 13028 문헌정보-기록의 디지털화 이행 지침」의 '6. 모범실무지침' 중 '6.3.4 메타데이터' 부분과 <부속서 C>의 '파일명 작성 메타데이터 권고사항' 정도에 불과하다. 「NAK 26:2018(v2.0) 기록물 디지털화 기준」의 개정 전인 「NAK 26:2013(v1.0) 기록물 디지털화 지침-제1부: 종이 및 사진·필름」에서 '디지털화 관련 메타데이터'가 <부속서 B>에 있었으나, 개정 후에는 이마저도 삭제되었다.¹⁾ 이렇듯 2022년 10월 현재 디지털화 기록의 메타데이터 작성에 필요한 실질적인 내용을 다루고 있는 국가표준 및 공공표준은 실질적으로 존재하지 않는다고 볼 수 있다.

이러한 상황을 고려하면 디지털화 기록에 대한 메타데이터는 결국 「NAK 8:2021(v2.2) 기록관리 메타데이터 표준」(이하 「기록관리 메타데이터 표준(v2.2)」)을 따른다고 볼 수 있다. 하지만 이 표준의 목적이 '결재과정에서 생산된 문서류 위주의 기록물'을 대상으로 하되 '시청각기록물, 웹기록물, 행정정보 데이터세트 등도 공통 항목에 대해서는 이 표준을 적용'할 수 있도록 하는 데 있어 디지털화 기록에 적용하기에 매우 제한적이다. 디지털화 작업 절차에서 적절한 메타데이터의 생성은 기록의 품질을 유지하는데 중요한 요인임에도 불구하고, 이렇듯 「기록관리 메타데이터 표준(v2.2)」으로는 디지털화 기록의 특성을 충분히 반영하기에 부족한 것이다.

해외의 경우, 미국 NARA는 「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토(Federal Records Management: Digitizing Permanent Records and Reviewing Records Schedules)」(2020)에서 §1236.54 메타데이터 요구사항(Metadata requirements)의 하위 절로 디지털화 기록에 대한 메타데이터 요구사항을 별도로 다루고 있다. 또한 그 이전인 2004년 6월에 발표된 「전자적 접근을 위한 기록자원 디지털화 기술 가이드라인(Technical Guidelines for Digitizing Archival Materials for Electronic Access)」에서도 디지털화 기록에 대한 메타데이터 구성요소를 언급한 바가 있다. 이 외에도 영국 TNA의 「디지털화 문서 (Digitisation at The National

1) 2018년도 제2차 기록관리 표준 제개정 공지(국가기록원 정책기획과-3091, 2018.12.31.) 공문에 따르면 NAK 26:2013(v1.0)을 비롯한 공공 및 원내표준을 하나의 표준으로 통합하고, 기록관리기관이 수행하기 어려운 기준과 선택의 폭 제한을 해소하기 위해 해당 표준을 개정했음을 밝혔다. 부속서 B(디지털화 관련 메타데이터)는 기록관에서 준수하기 어려운 기준에 해당되어 폐기된 것으로 보인다.

Archives)」(2016)와 캐나다 브리티시 컬럼비아의 「정부정보 디지털화 가이드라인(Digitizing Government Information Guide)」(2020)에서도 디지털화 기록의 메타데이터 영역이 따로 편성되어 있는 것을 확인할 수 있다.

이러한 배경에서, 이 연구는 중요문서 보존 및 활용사업의 결과물로 생산되는 수많은 디지털화 기록의 체계적인 관리를 위해 필요한 메타데이터 요소를 제안하는 데에 목적이 있다. 이를 위해 먼저 문헌분석을 통해 디지털화 기록의 정의와 특성을 확인한 다음 원천 기록과의 관계를 정립하였다. 다음, 국내의 국가표준과 공공표준을 분석하여 디지털화 기록 메타데이터와 관련한 사항을 분석하고, 국내·외 사례 간 상호 비교를 통해 디지털화 기록 관리를 위한 메타데이터 요소 설계에 필요한 시사점을 확인하였다. 마지막으로, 분석 내용을 토대로 디지털화 기록 관리를 위한 메타데이터의 설계방향과 원칙, 메타데이터 요소(안)을 제안하였다.

1.2 연구방법 및 내용

이 연구에서는 디지털화 기록에 대한 메타데이터 표준의 분석대상으로 국내 공공표준인 「기록관리 메타데이터 표준(v2.2)」과 NARA의 「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토」, TNA의 「디지털화 문서」를 선정하였다. 디지털화 기록의 메타데이터에 대한 국내 표준은 「기록관리 메타데이터 표준(v2.2)」이 유일하며, 국외의 경우 디지털화 사업의 주요 대상인 장기보존기록물을 중점적으로 관리하는 국가 기록관의 표준을 분석대상에 포함하였다. 이들 표준의 분석 목적은 디지털화 기록을 관리하는데 필요한 사항을 확인하고, 디지털화 기록과 관련된 메타데이터 기술요소를 상호 비교하는 데 있다. 분석결과는 디지털화 기록 관리를 위한 메타데이터 설계 방향과 원칙을 도출하고, 최종 메타데이터 요소(안)을 제안하는 데 근거자료로 활용하였다.

이 연구는 공공기관의 디지털화 사업으로 생산되는 디지털화 기록을 대상으로 하며, 공공영역 외 디지털화 기록까지는 고려하지 않았다. 또한, 많은 공공기관에서 현재 진행하고 있는 종이문서의 디지털화 사업으로 원천 기록의 유형을 제한하지 않았다. 즉, 이 연구에서 디지털화 기록은 종이를 비롯한 사진, 동영상, 음원 등 다양한 자원유형에 대한 작업의 결과물로 디지털화 파일을 생성하는 사례를 모두 포함하며, 이 연구에서 제안하는 메타데이터 또한 이들 유형을 모두 고려한 것이라 할 수 있다.

1.3 선행연구

디지털화 기록의 관리를 위한 메타데이터의 필요성은 이미 여러 연구에서 제기된 바 있다. 박은경(2005)은 디지털화(디지털라이제이션) 과정에서의 이미지 프로세스와 메타데이터 기술 실무에 대한 필요성을 강조하였다. 이선경(2008)은 기존 디지털화 사업에서 검색을 위한 색인목록을 작성하지만 품질 유지 등을 위하여 관련 메타데이터 요소가 획득되어야 한다고 언급하였다. 한희정, 김태영, 김용(2016)는 무형문화유산의 활용성을 높이고 관련 정보의 상호교환성 확보를 위해 다중개체 기반의 메타데이터 요소를 개발하였다. 이들 연구는 행정영역 외의 기록을 대상으로 하였으나 디지털화 기록의 품질 유지 등을 위해 메타데이터가 필요하다는 사실을 언급한 공통점을 가진다.

다음으로, 행정기록의 디지털화 기록에 관한 연구로는 특정 기관에 대한 사례 연구가 많은 비중을 차지하고 있다. 2005년 국가기록원이 주관한 행정기관 기록 DB 구축 사업을 대상으로 한 이예경(2007)을 시작으로 안지현(2014), 이유정(2016), 김도환(2021)은 지방자치단체의 디지털화 사업을 주제로 연구를 진행하였다. 이 중에서 이유정(2016)은 실제 디지털화 사업에서 메타데이터 기술 시 필수항목의 작성이 누락되는 점을 지적하며 기술항목의 관리 방안을 마련할 것을 제안하였다. 또한, 류재현(2017)은 공공기관의 중요기록 DB사업 프로세스를 분석하고 개선방안의 하나로 원천 기록의 메타데이터와 연계가 필요하다고 보았다. 김도환(2021)은 지방자치단체에서 진행한 디지털화 사업에 대한 만족도를 평가하고 문제점 및 개선방안을 제안하였다.

이처럼 선행연구에서는 디지털화 기록에 대한 메타데이터의 필요성에 대해 공감하는 수준에서 크게 벗어나지

못한 상태이며, 구체적인 메타데이터 적용 방안이나 요소 개발에 대한 연구는 아직까지 진행되지 못하였음을 확인할 수 있다. 이 연구는 디지털화 기록의 특징을 반영하면서도 원천 기록과의 연계를 유지할 수 있는 메타데이터 요소(안)을 제안하는데 목적을 두고 있다는 점에서 선행연구와 차별성을 가진다.

2. 이론적 배경

2.1 디지털화 기록

‘디지털화’란 하드카피 또는 비(非)디지털 기록을 디지털 형태로 변환하는 수단을 의미한다(KS X ISO TR 13028). 비(非)디지털 원천 기록을 디지털 사진으로 촬영하는 것 또는 아날로그 음성 기록을 디지털 매체로 변환하는 작업 등이 이에 해당한다. 「NAK/A 17:2014(v1.0) 녹음·동영상 기록물 디지털화 지침」에서는 ‘디지털화’를 스캐닝 또는 인코딩 장비 등을 이용하여 대상 기록을 디지털 형태로 변환하는 과정으로 정의하였다. 임나영, 남영준(2019)은 대상 기록을 컴퓨터가 인식할 수 있는 확장자를 가진 독립적인 디지털 파일형태로 변환하는 과정으로 정의하면서, 비전자기록뿐만 아니라 디지털 매체에 수록된 파일을 디지털 파일로 변환하는 것까지도 디지털화의 범주에 포함된다고 보았다.

이 연구에서 설정한 디지털화의 범주는 다음과 같다. 첫째, 기록정보 자료의 유형에 상관없이 결과적으로 디지털 형태의 자원을 생산하는 경우를 모두 포함한다. 따라서 종이, 아날로그 음성 파일에서부터 디지털 매체에 수록된 파일까지 다양한 유형의 기록정보 자료가 디지털화 과정에서 원천 기록이 될 수 있다. 이러한 원천 기록은 둘 이상의 매체 유형으로 구성된 혼합 매체 파일(mixed-media files)일 수도 있다. 둘째, 디지털화 기록은 스캐닝 또는 인코딩 장비를 활용한 경우로 한정한다. 즉, 스캐너 등의 장비를 활용하지 않고 원천 기록의 내용을 컴퓨터와 문서작성 소프트웨어를 이용하여 전사한 기록은 포함하지 않는다. 셋째, 변환한 디지털 자원은 기계가 인식할 수 있는 형태로 되어 있어야 하며, 확장자로 구성되어야 한다. 여기서 인식은 기계가 읽을 수 있는 수준이 아닌 기계를 통해 확인이 가능한 수준을 의미한다. 따라서 디지털화 기록은 ‘원천 기록을 대상으로 한 디지털화의 결과물로 기계가 확인할 수 있는 형태로 재생산된 기록’이라고 정의할 수 있다.

디지털화 기록과 유사한 개념으로 「공공기록물 관리에 관한 법률 시행령」 제53조에 따라 생성하는 ‘원본대체사본’도 있다. 원본대체사본은 보존기간 30년 이상인 기록 중 심각한 물리적 훼손으로 복원이 불가능하여 보존매체에 수록하여 원본을 대체하여 보존하는 기록을 말한다. 원본대체사본은 원본의 폐기를 전제로 하기 때문에 원천 기록과 대체사본에 대한 이중적인 관리체계를 필요로 하지 않는다. 이 외에도 공공기록물법 시행령 제29조에서 병행보존, 원본보존, 원본대체보존 등 여러 보존방법을 제시하고 있으나 보존매체를 보존용 전자매체 또는 마이크로필름으로 한정하고 있어 이 연구에서 다루고자 하는 디지털화 기록과는 차이가 있다.

원천 기록과 디지털화 기록의 관계를 정립하기 위해 가상의 디지털화 프로세스 시나리오를 상정한 결과는 다음 <그림 1>과 같다. 원천 기록을 A기관 B부서에서 ‘공무원 교육제도 개선’이라는 제목으로 생산한 10쪽의 종이문서로 가정할 때, 디지털화 프로세스를 통해 디지털화 파일을 획득하더라도 해당 파일 내 문서의 제목은 여전히 ‘공무원 교육제도 개선’이며 생산자는 A기관 B부서이다. 다만 매체 유형은 종이에서 자기디스크로, 크기는 10쪽에서 720KB로 변경될 수 있다. 여기서 알 수 있는 점은 디지털화 과정을 거치더라도 기록의 형태만 변할 뿐 내용이나 생산맥락과 같은 본질적인 요인은 바뀌지 않는다는 사실이다. 이처럼 디지털화 기록은 새로운 개별 기록이 아닌 원천 기록의 사본, 즉 이용과 보존을 위한 대체물이라고 할 수 있다. 따라서 디지털화 기록을 관리하기 위해서는 여러 유형의 원천 기록과 함께 재생산된 기록까지 모두 포괄하는 관리체계를 마련할 필요가 있다.



〈그림 1〉 디지털화 프로세스에서의 원천 기록과 디지털화 기록의 관계

2.2 디지털화 기록 메타데이터

기록 메타데이터란 시간이 지나가도 영역 안과 영역 간 기록의 생산, 관리, 이용을 가능하게 하는, 구조화되었거나 반구조화된 정보를 의미한다(KS X ISO 15489-1:2016 3.12). 기록관리 메타데이터는 기록을 업무맥락 속에 고정하여 기록에 대한 통제를 시작하면서 획득 시점에 기록을 정의하며, 기록의 품질인 진본성, 신뢰성, 이용가능성, 무결성을 보장한다. 또한 전자업무, 보존, 자원 기술, 자원 발견, 권한 관리 등 다양한 목적을 위해 생산, 관리된다(KS X ISO 23081-1:2017). 우리나라 국가표준 및 공공표준에서 디지털화 기록에 대한 메타데이터와 관련된 부분을 살펴보면 다음과 같다.

먼저, 「기록관리 메타데이터 표준(v2.2)」에서는 기록관리 메타데이터가 객체 유형에 상관없이 기록의 품질을 보장하고 적절한 관리와 이해를 위해 매우 중요한 역할을 한다고 밝히고 있다. 이러한 메타데이터는 기록과 기록을 생산·관리·유지·이용하는 사람, 과정과 시스템, 관리정책 등을 확인하고, 관련 활동의 맥락을 파악하는데 사용된다. 「NAK 23:2017(v1.2) 기록물 목록 작성 및 디지털화 작업 지침」에 따라 기록의 디지털화는 (1) 자료준비 단계, (2) 정리단계, (3) 색인목록 등록 단계, (4) 기록 디지털화 단계, (5) 작업 완료 단계 등 5단계로 진행된다. 이를 고려하면 디지털화 기록 메타데이터는 이 5단계와 이러한 과정 전반에 관련된 사람, 시스템 등을 모두 반영해야 한다.

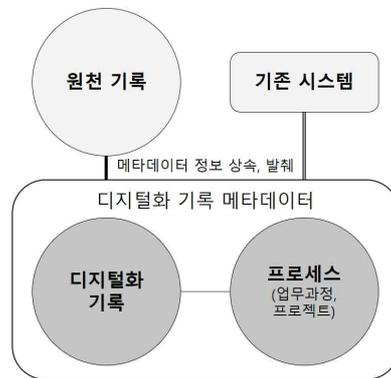
「KS X ISO TR 13028 문헌정보-기록의 디지털화 이행 지침」에서는 디지털화 과정을 문서화하고 지속적인 업무 과정을 지원하기 위해 반드시 모든 디지털화 기록에 메타데이터를 부여할 것을 권고하고 있다. 메타데이터는 디지털화 기록을 관리하고 검색하는데 필수적인 구성요소이며, 이를 위해 다음 두 가지 유형의 메타데이터가 획득되어야 한다고 제시하고 있다: (1) 개별 디지털화 기록과 디지털화 과정을 설명하는 메타데이터, (2) 기록, 수행 중인 업무, 해당 업무와 연관된 행위 주체에 관한 메타데이터. 이 지침에는 이와 관련된 메타데이터 대다수는 디지털화 과정을 관리하는데 사용된 소프트웨어와 하드웨어로부터 자동 추출되어야 하며, 수작업으로 메타데이터를 부여하거나 적용하는 것은 가급적 최소화하여야 한다는 내용이 포함되어 있다. 또한, 메타데이터는 디지털화 기록에 임베디드(embedded)되거나 분리된 시스템에서 관리하거나 혹은 두 가지 방법 모두를 적용할 수 있지만, 어떠한 경우라도 생산된 메타데이터는 디지털화 기록과 직접적으로 연결되거나 연관된 상태로 관리되어야 한다고 강조하고 있다. 이 지침에서는 이외에도 이미지 계층 메타데이터, 업무과정 디지털화 메타데이터, 디지털화 프로젝트²⁾ 메타데이터에서 고려할 사항을 제시하고 있다. 이 가운데 이미지 계층 메타데이터는 상속받은 메타데이터 외에 다음의 요소를 반드시 포함해야 한다고 적시되어 있다: (1) 디지털 이미지 고유식별기호, (2) 디지털화 일시, (3) 디지털화 과정과 연관된 행위 주체의 이름, (4) 획득장치(HW/SW), (5) 최종보정일자.

「NAK 23:2017(v1.2) 기록물 목록 작성 및 디지털화 작업 지침」에서는 색인목록 등록 검증 단계에서 검증담당자가 「NAK 29-2:2016(v1.2) 기록관리시스템 데이터연계 기술규격-제2부 중앙연구기록관리시스템과의 연계」에서

2) 업무과정 디지털화는 일상 업무과정의 일부분으로서의 계속적이고 일상적인 디지털화를 의미하며, 디지털화 프로젝트는 프로젝트 기반으로 구기록(legacy records)을 대량으로 디지털화하는 것을 의미한다(KS X ISO TR 13028).

정하는 '기록철 생산현황'의 항목을 기준으로 등록된 기록철의 메타데이터 정보 발취가 적정한지 검사할 것을 명시하고 있다. 마지막으로, 「NAK 26:2018(v2.0) 기록물 디지털화 기준」에서는 디지털화에 의한 결과물이 가져야 할 기본적인 기준을 제시하고, 기록 유형별 작업방법을 소개하고 있다. 이 기준에서는 디지털화 과정의 기록화를 위해 필요한 최소한의 기록 사항과 이러한 기록이 최초 생성 이후에 변경될 수 없도록 관리할 것을 분명하게 제시하고 있다. 최소한의 기록 사항은 (1) 디지털화 대상, 목적, 일시, 작업자, (2) 디지털화에 사용된 장치(모델, 제조사 등), (3) 디지털화에 사용된 소프트웨어에 관한 사항(모델, 제조사 등), (4) 압축 시 압축방법에 관한 사항(명칭, 표준번호 등), (5) 디지털화 결과의 해시함수 종류, 해시값 등이다.

이상에서 살펴본 바와 같이, 국내 표준에서는 공통으로 모든 디지털화 과정을 기록화할 것을 강조하고 있다. 디지털화 기록 자체에 대한 메타데이터뿐만 아니라 업무 과정과 디지털화 프로젝트에 대한 메타데이터도 생산, 관리되어야 한다고 언급되어 있다. 또한 원천 기록의 메타데이터 정보를 상속하거나 발취하기 위하여 기존 업무시스템 혹은 기록관리시스템과의 연계도 강조하고 있다. 이러한 내용을 토대로 디지털화 기록 메타데이터의 데이터 모델을 도출한 결과는 다음 <그림 2>와 같다. 다음 장에서는 이러한 데이터 모델에 적합한 기록관리 메타데이터 요소를 보다 상세히 분석하고자 한다.



<그림 2> 디지털화 기록 데이터 모델

3. 디지털화 기록 메타데이터 표준과 요소 분석

디지털화 기록에 대한 메타데이터 요소를 도출하기 위하여 국내외 관련 표준을 분석하였다. 국내의 경우, 「KS X ISO 23081-1 문헌정보-기록관리과정-기록 메타데이터-제1부: 원칙」을 기반으로 국내 사정에 맞게 설계한 것이 「기록관리 메타데이터 표준(v2.2)」이므로, 이 표준에서 디지털화 기록과 관련된 요소를 별도로 추출하여 살펴보았다. 국외는 국가 기록관 수준에서 디지털화 기록의 특성을 반영한 표준을 선별하였다. 분석대상에 포함된 국외 표준은 미국 NARA의 「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토」(2020)와 영국 TNA의 「디지털화 문서」(2016년 최종 개정)이다.

3.1 국내 「기록관리 메타데이터 표준(NAK 8:2021(v2.2))」

「기록관리 메타데이터 표준(v2.2)」에 수록된 메타데이터는 23개 상위요소, 66개 하위요소, 58개 세부요소의 3계층으로 구성되어 있다. 국내 표준에 따르면 디지털화 기록 메타데이터는 기록물 '건' 단위로 원천 기록과 연계되어야 하므로 이 연구에서는 기록물 건을 기준으로 메타데이터 요소를 분석하였다. 「기록관리 메타데이터 표준(v2.2)」에

서 기록물 건에 해당하는 메타데이터 요소는 20개 상위요소, 56개 하위요소, 56개 세부요소 등 총 138개이다.

국내의 경우 현재 디지털화 기록에 한정된 메타데이터 표준이 없으므로, 기록물 건 단위 메타데이터 요소를 (1) 공공기록물법 시행령에서 정하는 사본과 원본대체사본 관련 요소, (2) 폐지된 「NAK 26:2013(v1.0) 기록물 디지털화 지침-제1부: 종이 및 사진·필름」의 <부속서 B>에 제시된 디지털화 관련 메타데이터 요소, (3) 앞서 2장 2절에서 정리한 디지털화 기록 메타데이터에 반영되어야 할 요소를 기준으로 하여 디지털화 기록과 관련된 요소만 추출하는 과정을 거쳤다. (1)은 디지털화 기록이 원천 기록의 사본에 해당하기 때문에, (2)는 비록 폐지된 지침이지만 디지털화 관련 메타데이터 요소를 제시했던 유일한 국내 표준에 해당하기 때문에, 그리고 (3)은 앞서 살펴본 것처럼 관련 지침이나 기준 등에서 디지털화 과정을 관리하기 위해 기록해야 할 항목을 분명하게 제시하고 있기 때문에 디지털화 기록에 적용가능한 메타데이터 요소를 추출하고, 나아가 디지털화 기록의 메타데이터 설계에 참조할 사항을 도출하는데 이들을 활용하는 것이 적절하다고 판단되었다. 「기록관리 메타데이터 표준(v2.2)」과 (1), (2), (3)을 비교한 결과는 <표 1>과 같다.

먼저, 공공기록물법 시행령에 따른 사본 관리와 관련된 요소를 선별한 결과, 상위요소는 ‘기록식별자’, ‘전자기록물 여부’, ‘유형’, ‘매체’, ‘크기’, ‘위치’, ‘보존이력’, ‘관계’ 8개이며, 하위요소 13개와 세부요소 8개로 총 29개로 나타났다. 사본 및 원본대체사본의 경우, 메타데이터 생성 시 ‘사본 유형’ 요소를 입력한다. 사본은 ‘사본 유형’에 요소값을 입력하고, ‘관계’ 요소를 활용하여 원천 기록과의 관계를 표시한다. 하지만 원천 기록을 폐기하는 원본대체사본은 개별 기록으로 간주하여, 대부분 원본 메타데이터와 동일한 요소 값을 가지지만 ‘기본식별자’, ‘전자기록물 여부’, ‘매체’, ‘크기’, ‘위치’, ‘보존이력’, ‘관계’에 해당하는 값만 달라진다. 이러한 작성방법을 미루어 보아 사본과 원본대체사본에 대한 메타데이터 스키마가 별도로 구성된 것이 아니며 기존의 메타데이터 요소를 그대로 활용한다는 사실을 알 수 있다. 이러한 점을 고려할 때 디지털화 기록의 메타데이터에서도 원천 기록에서의 정보 발체 및 연계를 위해 기존 메타데이터 요소를 활용하여 최대한 원천 기록의 정보를 획득하고 디지털화 기록의 특성을 추가로 표현할 수 있는 요소를 포함시키는 방향으로 설계되어야 할 것으로 보인다.

다음으로, 「NAK 26:2013(v1.0) 기록물 디지털화 지침-제1부: 종이 및 사진·필름」의 <부속서 B>와 대조해 본 결과, 디지털화 프로젝트 시 ‘기술(description)’, ‘주제’, ‘크기’ 등 일부 요소를 제외한 대부분의 정보가 원천 기록으로부터 획득되어야 하는 것으로 나타났다. 그 중 ‘기록계층’, ‘기록물명’, ‘분류’, ‘생산이력’, ‘보존기간’ 등 일부 요소는 원천 기록의 메타데이터 정보를 그대로 사용하도록 정해져 있었다. 하지만 ‘생산자’와 ‘일시’는 각각 이미지 생산자와 이미지 생산일자를 기재하도록 되어 있어 원천 기록의 정보를 사용하는 요소와의 충돌이 발생할 가능성이 있는 것으로 나타났다. 또한 원천 기록의 생산이력을 그대로 받아오면서 ‘생산자’ 요소에 이미지 생산자를 기재할 경우 원본 생산자와 이미지 생산자가 중복으로 기재되어 혼란을 불러일으킬 수도 있었다. 이와 같은 가능성이 보인 이유는 이 지침에서 원천 기록과 디지털화 기록의 관계를 제대로 정립하지 못한 상태에서 요소를 선정, 도입하였기 때문인 것으로 보인다.

마지막으로, 디지털화 표준이나 지침에 명시된 디지털화 과정에서 기록해야 할 항목을 「기록관리 메타데이터 표준(v2.2)」과 대조한 결과, 상위요소 5개와 하위요소 16개, 세부요소 14개 등 총 35개의 요소가 관련된 것으로 나타났다. 관련 상위요소로는 ‘관력이력’, ‘보존이력’ 등이 있으나 표준에서 언급한 고려사항을 모두 충족하는 것은 아니었다. 디지털화 과정과 연관된 행위주체의 기술에서도 ‘관력이력’이나 ‘보존이력’의 행위자 요소를 활용하여야 하는데, 이 역시 세부요소가 ‘기관명’, ‘부서명’, ‘개인명’ 등으로 설정되어 있어 디지털화 사업을 추진하는 사업체 정보를 작성하기에 적절하지 않은 것으로 나타났다. 또한 표준이나 지침에서 디지털화에 사용한 하드웨어나 소프트웨어의 모델이나 제조사 등에 대한 정보도 관리해야 한다고 규정하고 있으나, 기록물 건 단계에서 ‘포맷’ 요소는 적용되지 않는 것이었다. 따라서 다른 기록계층(컴포넌트)에 적용하는 ‘포맷’ 요소를 임의로 포함시켰다. 그러나 ‘포맷’ 요소에 정의된 하위요소가 ‘포맷명’, ‘포맷버전’, ‘생성 애플리케이션명’, ‘생성 애플리케이션버전’ 등 4개뿐이어서 관련 정보를 충분히 표기하기는 부족한 것으로 나타났다.

〈표 1〉 기록관리 메타데이터 표준 분석

대상	기록관리 메타데이터 표준(v2.2)	(1) 사본 관련 요소 (29)	(2) 폐지된 디지털 화 메타데이터 요소 (118)	(3) 관련 표준이나 지침 요소 (35)	디지털화 기록 에 필요한 요소 (59)	
	상위요소 및 하위요소 ※ 세부요소는 하위요소 뒤 ()안에 개수만 표기					
기록 물 건	1. 생산자		○			
	1.1 생산자 유형		○			
	1.2 기관명		○			
	1.3 기관코드		○			
	1.4 부서명		○			
	1.5 부서코드		○			
	1.6 개인명		○			
	1.7 개인코드		○			
	1.8 직위(직급)명		○			
	1.9 직위(직급)코드		○			
	2. 기록계층			○		
	3. 기록식별자	○	○	○	○	○
	3.1 기본식별자	○	○	○	○	○
	3.2 시스템식별자(2)		○	○	○	○(2)
	3.3 보조식별자(2)		○	○	○	○(2)
	4. 기록물명			○		
	4.1 제목			○		
	4.2 기타제목(2)			○		
	5. 기술					
	5.1 기술 유형					
	5.2 기술 내용					
	6. 주제					
	6.1 주제 유형					
	6.2 주제명					
	7. 전자기록물 여부	○	○			○
	8. 유형	○	○			○
	8.1 기록 유형					
	8.2 사본 유형	○	○			○
	8.3 기타문서 유형					
	10. 매체	○	○			○
	11. 크기	○			○	○
	11.1 용량	○			○	○
	11.2 단위	○			○	○
	13. 일시			○		
	13.1 생산일시			○		
	13.3 등록일시			○		
	13.4 시행일시			○		
	14. 생산이력			○		
	14.1 생산시스템			○		
	14.2 생산부서(4)			○		
	14.3 생산유형			○		
	14.4 생산경로(7)			○		
15. 보존기간(비밀기록물일 경우)			○			
15.1 보존기간			○			
15.2 보존기간책임사유			○			
17. 권한			○			
17.1 비밀(3)			○			
17.2 접근범위			○			
17.3 공개(8)			○			
17.4 공공저작물관리(2)			○			

	18. 위치		○			○
	18.1 소장처		○			○
	18.2 소장위치		○			○
	19. 관리이력			○	○	○
	19.1 관리 유형			○	○	○
	19.2 관리설명			○	○	○
	19.3 관리일시			○	○	○
	19.4 관리행위자(8)			○	○	○(8)
	19.5 변경요소(2)			○	○	○(2)
	20. 이용이력			○		
	20.1 이용 유형			○		
	20.2 이용내용설명			○		
	20.3 이용일시			○		
	20.4 이용자(8)			○		
	21. 보존이력	○		○	○	○
	21.1 보존처리 유형	○		○	○	○
	21.2 보존처리설명	○		○	○	○
	21.3 보존처리일시	○		○	○	○
	21.4 보존행위자(8)	○		○	○	○(8)
	22. 관계	○		○		○
	22.1 관계 유형	○		○		○
	22.2 관계대상식별자	○				○
	22.3 관계설명	○		○		○
	23. 무결성체크			○	○	○
	23.1 무결성체크섬			○	○	○
	23.2 무결성체크값			○	○	○
컴포 넌트	9. 포맷			○	○	○
	9.1 포맷명			○	○	○
	9.2 포맷버전			○	○	○
	9.3 생성 애플리케이션명			○	○	○
	9.4 생성 애플리케이션버전			○	○	○

이러한 분석결과를 토대로, 디지털화 기록과 관련된 메타데이터 요소를 최종적으로 추출하였다. 디지털화 기록의 메타데이터는 <표 1>에서 (1), (2), (3) 중에서 하나라도 언급한 요소를 모두 포함하되, 원천 기록과 동일한 데이터 값을 가지는 요소는 제외하였다. 원천 기록과 동일한 메타데이터는 디지털화 기록 메타데이터가 아니라 원천 기록으로부터 획득하여 활용할 수 있기 때문이다. 또한, 「기록관리 메타데이터 표준(v2.2)」 중 기록물 건에 적용하는 요소를 추출하였지만, 컴포넌트에 적용하는 ‘포맷’ 요소는 포함하였다. ‘포맷’ 요소는 디지털화 기록에 적용해야 하는 요소지만, 「기록관리 메타데이터 표준(v2.2)」에서 기록물 건에 적용되지 않는 요소였기 때문이다. 이러한 과정으로 최종 추출된 디지털화 기록의 메타데이터는 상위요소 ‘기록식별자’, ‘전자기록물 여부’, ‘유형’, ‘포맷’, ‘매체’, ‘크기’, ‘위치’, ‘관리이력’, ‘보존이력’, ‘관계’, ‘무결성체크’였으며, 하위요소, 세부요소를 모두 합쳐 총 59개였다.

이러한 분석 결과를 보면, 디지털화 기록의 관리를 위한 관점에서 「기록관리 메타데이터 표준(v2.2)」는 다음과 같은 문제점을 가지고 있음을 확인할 수 있었다. 첫째, 원천 기록과 디지털화 기록의 관계가 제대로 정립하지 않은 채 메타데이터 표준이 제시되어 있었다. 특히 폐지된 <부속서 B>와 비교하여 ‘생산자’ 요소에 이미지 생산자 정보를 입력하도록 하면서 ‘생산이력’은 원천 기록의 생산이력을 그대로 활용하도록 하는 구조는 메타데이터 작성자의 혼란을 야기할 수 있는 것이었다. 따라서 메타데이터 설계 시 디지털화 기록의 성격과 특성을 제대로 규정한 다음 메타데이터 요소를 구분해서 작성할 수 있도록 설계하여야 할 것이다.

둘째, 관련 표준에서는 디지털화 기록 관리를 위해 디지털화 기록에 대한 정보와 디지털화 업무과정 및 프로젝트에 대한 정보를 구분하여 기록화하여야 한다고 되어 있었다. 하지만 「기록관리 메타데이터 표준(v2.2)」은 기록 자체

에 대한 요소를 중심으로 설계되었을 뿐만 아니라 업무과정과 디지털화 업무에 대한 정보를 포함할 수 있는 요소는 매우 적어 향후 관련 요소가 추가적으로 개발될 필요가 있는 것으로 나타났다.

마지막으로 기록계층별로 메타데이터 요소의 적용을 달리하고 있었다는 점이다. 이 연구에서는 디지털화 기록 메타데이터의 작성계층을 기록물 건 단위로 분석하였는데, 기록물 건에 적용하지 않는 요소로 ‘포맷’, ‘분류’, ‘생산 이력’, ‘보존장소’ 등이 있었다. 하지만 혼합매체파일 등을 관리하여야 하는 상황을 고려할 때, ‘포맷’은 디지털화 기록 관리에 있어 필수적인 요소임에도 불구하고 비적용 요소로 정의되어 있어 이에 대한 검토가 필요한 상황임을 알 수 있었다.

3.2 미국 NARA의 「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토」

미국의 「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토」(2020)는 영구적으로 보존하여야 하는 일천 기록을 처분하고자 하는 부처가 디지털화 과정에서 준수하여야 할 사항을 정해 NARA에서 발표한 규정이다. 규정은 하위 절 36 CFR Parts 1224, 1225와 1236으로 구성되어 있으며, 메타데이터와 관련된 사항은 Part 1236에 있다.

「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토」에서 메타데이터는 디지털화 과정과 결과, 그리고 전자기록의 기술적 속성을 설명하기 위해 필요한 도구로 간주된다. 이 규정에서는 메타데이터를 구성하는 정보를 일천 기록과 디지털화 단계별로 포착하여 수집하는 것이 중요하다고 밝히고 있다. 이 규정에서 메타데이터는 수집할 정보의 성격과 목적에 따라 관리(administrative), 설명(descriptive), 기술(technical), 전송(transfer) 메타데이터로 구분되며, 이에 해당하는 요소는 <표 2>와 같이 총 34개이다.

<표 2> 디지털화 기록 관리를 위한 메타데이터 요소(미국 NARA)

구분	레이블	요구 필드	설명
관리 (ad- mini- stra- tive)	식별자:파일명 (Identifier: File Name)	-	확장자를 포함한 컴퓨터 파일의 전체 이름
	식별자:레코드 ID (Identifier: Record ID)	-	기관 또는 기록관리시스템에서 할당된 고유 식별자 (§ 1236.20(b)(1)은 기관이 각 레코드에 고유 식별자를 할당하도록 요구)
	식별자:처분일정 건 (Identifier: Disposition Schedule Item)	-	레코드가 속한 처분 일정 항목에 지정된 번호
	관계:부분자원 (Relation: Has Part)	-	완전한 레코드를 형성하기 위해 물리적 또는 논리적으로 필요한 관련 기록. 여러 미디어 유형에 대한 기록 포함하는 혼합 미디어 파일은 이 요소를 사용하여 모든 구성요소 식별
	관계:전체자원 (Relation: Is Part Of)	-	설명된 레코드가 물리적 또는 논리적으로 포함된 관련 레코드 또는 파일, 혼합 미디어 파일의 구성요소인 기록은 이 요소를 사용하여 상태를 표시
설명 (des- crip- tive)	제목(Title)	-	원본 레코드에 지정된 이름, 이름이 없는 경우 <식별자: 레코드 ID>를 제목으로 사용
	설명(Description)	-	레코드에 대한 요약물 포함한 레코드 내용에 대한 설명
	생산자(Creator)	-	원본 레코드 작성에 주된 책임이 있는 대리인(개인, 기관, 기타 조직 등)
	일자:생산일자 (Date: Creation Date)	-	원본 레코드가 연방 레코드의 정의를 충족하는 날짜 또는 날짜 범위
	자원유형(Source Type) 자원 치수 (Source Dimensions)	-	디지털 스캔 이미지를 만들기 위해 스캔된 원본 레코드의 매체 원본 레코드의 치수(측정 단위 포함)

	접근제한 (Access Restrictions)	접근제한상태 (Access Restriction Status)	레코드에 대한 액세스 제한 여부
		특정접근제한 (Specific Access Restriction)	국가 보안 고려사항(예: CNSI, CUI), 기증자 제한, 법원 명령 및 기타 법적 또는 규제 조항(개인정보보호법 및 정보자유법(FOIA) 면제)을 기반으로 한 기록의 특정 접근 제한
	이용제한 (Use Restrictions)	이용제한상태 (Use Restriction Status)	레코드에 대한 이용 제한 여부
		특정사용제한 (Specific Use Restriction)	개인정보보호법 및 정보자유법(FOIA) 면제를 포함한 저작권, 상표, 서비스 표시, 기증자 또는 법적 조항에 기초한 레코드의 사용 제한 유형
	권한: 권한보유자 (Rights: Rights Holder)	-	레코드에 관한 지적재산권을 소유하거나 관리하는 개인 또는 단체
기술 (technical)	파일 크기(File Size)	-	이미지 파일의 크기(바이트)
	포맷명과 버전 (Format Name and Version)	-	파일 형식의 형식 이름 또는 설명
	이미지 너비(Image Width)	-	디지털 이미지의 너비(픽셀) (예: 수평 또는 X 치수)
	이미지 높이(Image Height)	-	디지털 영상의 높이(픽셀) (예: 수직 또는 Y 치수)
	색공간(Color Space)	-	사용된 국제 컬러 컨소시엄(ICC) 프로파일의 잘 정의된 이름
	생성일시(Date and Time Created)	-	디지털 이미지가 생성된 날짜 또는 시간
	사용 스캐너와 모델 (Scanner Make and Model)	-	이미지를 만드는 데 사용되는 스캐너의 제조업체 및 모델
	스캔 소프트웨어와 버전 (Scanning Software Name and Version)	-	스캐너가 이미지를 만드는 데 사용하는 소프트웨어의 이름 및 버전
	사용 디지털카메라와 모델 (Digital Camera Make and Model)	-	이미지를 만드는 데 사용되는 디지털 카메라의 제조업체 및 모델
	픽셀당 색상 구성 (Samples Per Pixel)	-	픽셀당 색상 구성 요소 수
	메시지 다이제스트 알고리즘 (Message Digest Algorithm)	-	디지털 개체 또는 비트스트림에 대한 메시지 다이제스트를 구성하는 데 사용되는 특정 알고리즘
	메시지 다이제스트(체크섬) (Message Digest (checksum))	-	메시지 요약 알고리즘의 출력
	전송 (transfer)	전송명(Transfer Title)	-
일자 (Dates)		시작일자 (Inclusive Start Date)	사용자가 NARA로 전송하려는 레코드 그룹, 컬렉션, 시리즈 또는 세트가 작성자에 의해 생성, 유지 또는 누적된 시작 날짜
		마감일자 (Inclusive End Date)	작성자가 NARA로 전송하려는 레코드 그룹, 컬렉션, 시리즈 또는 세트를 생성, 유지 또는 누적한 마지막 날짜
생성조직 (Creating Organization)		-	(주요) 작업 중인 경우 컬렉션, 시리즈 또는 세트의 생성, 누적 또는 유지 관리를 담당하는 조직의 이름
기록물군 번호 (Record Group Number)		상위 기록물군 번호 (Parent Record Group Number)	레코드 그룹에 할당된 고유 번호
일반적인 기록유형 (General Records Type)		-	전송 중인 레코드 세트, 시리즈 또는 컬렉션의 일반적인 형식(예: 건축 및 엔지니어링 도면, 아티팩트, 데이터 파일, 지도 및 차트, 움직이는 이미지, 사진 및 기타 그래픽 자료, 녹음, 텍스트 레코드 또는 웹 페이지)
접근제한	접근제한상태	NARA로 전송할 레코드 세트, 컬렉션 또는 시리즈에 대한	

(Access Restrictions)	(Access Restriction Status)	엑세스 제한 여부
	특정접근제한 (Specific Access Restriction)	국가 보안 고려사항(예: CNSI, CUI), 기증자 제한, 법원 명령 및 기타 법적 또는 규제 조항(개인정보보호법 및 정보자유법(FOI) 면제)을 기반으로 한 일련의 기록에 대한 특정 접근 제한
이용제한 (Use Restrictions)	이용제한상태 (Use Restriction Status)	NARA로 전송할 레코드 세트, 컬렉션 또는 시리즈에 대한 사용 제한 여부
	특정사용제한 (Specific Use Restriction)	개인정보보호법 및 정보자유법(FOIA) 면제를 포함한 저작권, 상표, 서비스 표시, 기증자 또는 법적 조항에 따라 레코드 세트, 컬렉션 또는 시리즈에 대한 사용 제한 유형
기록물 일정 번호 (Record Schedule Number)	여러 기록물 일정 번호 (Records Schedule Number)	전송 중인 컬렉션, 시리즈 또는 세트의 모든 레코드에 적용되는 레코드 스케줄에 할당된 NARA 번호

NARA의 디지털화 기록을 위한 메타데이터 특징은 다음과 같다. 첫째, 같은 요소를 사용하더라도 상세구분을 활용하여 적용대상을 구분하고 있다. 관리 메타데이터에서는 디지털화 기록의 식별자를 ‘식별자:파일명(Identifier: File Name)’으로 관리하여 원천 기록의 식별자인 ‘식별자:레코드 ID(Identifier:Record ID)’와 구분하여 사용한다. 동일한 ‘식별자’라는 요소를 사용하지만 상세구분을 활용하여 적용대상을 명확하게 정하고 있는 것이다.

둘째, 디지털화 기록의 내용, 생산맥락 등에 대한 정보는 원천 기록과 동일하다. 설명 메타데이터는 원천 기록에 대한 정보로 구성된 반면 기술 메타데이터는 디지털 파일 사이즈 등 디지털화 기록 관리에 필요한 정보로 구성되어 있다. 이를 통해 디지털화 기록이 원천 기록과 형태만 다를 뿐 내용, 맥락 등은 원천 기록과 다르지 않은 사본이라는 전제 하에 메타데이터를 설계했음을 확인할 수 있다.

마지막으로, 디지털화 프로세스 전반에 걸친 기록화를 위해 프로세스를 대상으로 한 메타데이터 요소가 반영되어 있다. 디지털화한 결과물을 NARA로 전송하는 과정에서 생성하여야 하는 정보를 전송 메타데이터로 정해놓아 디지털화 프로세스 전반에 대한 기록화를 가능하도록 구현하고 있다.

3.3 영국 TNA의 「디지털화 문서」

「디지털화 문서」(2016)는 영국 TNA에서 소장한 컬렉션 중 아날로그 기록을 디지털화할 때 필요한 표준과 요구사항에 대한 규정으로서 이미지 규격을 포함한 초기 이미지 스캔에서부터 이미지 전송까지의 과정을 포함하고 있다. 이 규정에서는 파기할 원본을 대신하여 영구보존을 위해 디지털화한 기록을 디지털화 기록(digitised records)으로, 원본 기록은 그대로 보존한 상태에서 디지털화한 기록을 디지털 대체물(digital surrogates)로 구분한다. 이 연구에서 대상으로 하는 디지털화 기록은 후자의 의미에 가깝다.

「디지털화 문서」에서 메타데이터는 임베디드(embedded)과 외부(external)으로 구분하며, 외부 메타데이터는 다시 기술(technical) 메타데이터와 전사(transcription) 메타데이터로 구분한다. 마지막으로 전사 메타데이터는 전사할 기록에 따라 내용이 다양하게 나올 수 있으나 이에 대해서는 메타데이터 요소가 아니라 일반적인 원칙으로 대신하고 있다. 따라서 전사 메타데이터는 분석에서 제외하였다.

먼저, 임베디드 메타데이터는 이미지 파일 자체에 포함되는 최소한의 메타데이터이다. 임베디드 메타데이터는 기록의 장기 보존 시 식별이 용이하도록 할 목적에서 설계되었다. 임베디드 메타데이터를 구성하는 요소는 왕실 저작권 혹은 신뢰기반 제3자 저작권 등 저작권 문구, UUID(Universally unique identifier, RFC4122), URI(Uniform resource identifier)이며, 이 요소는 구조적인 XML로 작성되어 최종적으로 JPEG 2000 규격의 이미지에 탑재된다.

다음, 기술 메타데이터는 각 이미지별로 디지털 이미지 생성 및 전자 텍스트 생산과 관련하여 수집되는 데이터로, 디지털화 기록의 장기적인 보존을 위한 역할을 수행한다. 기술 메타데이터의 범위는 하드웨어, 소프트웨어와 이미지를 만드는 디지털화 과정을 모두 포함한다. 이 데이터는 특정 스캐너가 낮은 품질의 이미지를 만들거나 소프트웨어 패키지에서 버그가 발생하는 등 품질 보증 과정에서 처리할 수 없는 시스템 측면에서의 문제를 확인할 수 있도록 하는 역할도 담당한다. 기술 메타데이터는 UTF-8 인코딩을 거쳐 RFC 4180 규격에 따라 CSV 텍스트 파일로 생성된다.

「디지털화 문서」에서는 <표 3>과 같이 디지털화 기록과 디지털 대체물에 대한 기술 메타데이터를 별도로 정하고 있다. 이 가운데 종이문서의 디지털화와 디지털 대체물에 공통적으로 적용되는 요소는 29개이며, 디지털화 기록을 대상으로 추가로 적용되는 요소가 13개, 디지털 대체물을 대상으로 추가로 적용되는 요소가 1개 있다. 종이문서가 아니라 마이크로 폼 혹은 사진자료를 대상으로 한 디지털화 기록이라면 1~4개의 요소를 추가할 수 있어 마이크로 폼 자료의 디지털화 기록이라면 기술 메타데이터는 최대 46개까지 확장될 수 있다.

<표 3> 디지털화 기록을 위한 기술 메타데이터 요소(영국 TNA)

기술(technical) 메타데이터	종이문서 등	(추가)마이크로폼 및 사진자료	(추가)마이크로폼
공통 (디지털화 문서, 디지털 대체물)	총 29개 요소 배치코드(Batch code), 부처(Department), 원(Division), 시리즈(Series), 하위시리즈(Sub series), 하위-하위시리즈(Sub sub series), 첩(Piece), 건(Item), 설명(Description), 순서(Ordinal), 파일 UUID(File UUID), 파일 경로(File path), 파일 체크섬(File checksum), 자원 URI(Resource URI), 스캔연산자(Scan Operator), 스캔 ID(Scan ID), 스캔 위치(Scan location), 이미지 해상도(Image Resolution), 이미지 너비(Image width), 이미지 높이(Image height), 이미지 톤 해상도(Image tonal resolution), 이미지 포맷(Image format), 이미지 색공간(Image colour space), 이미지 분할(Image split), 이미지분할 순서(Image split ordinal), 이미지 분할 UUID(Image split other UUID), 이미지 절단(Image crop), 이미지 측정렬(Image de-skew), 주기(Comments)	총 3개 요소 이미지 반전 (image inversion), 이미지 반전 연산자 (image inversion operator), 이미지 반전 타임스탬프 (image inversion timestamp)	총 1개 요소 피시 참조번호 (fiche reference)
디지털화 기록 추가 요소	총 13개 요소 스캔 원포맷(Scan native format), 스캔 타임스탬프(Scan timestamp), 이미지 분할 연산자(Image split operator), 이미지 분할 타임스탬프(Image split timestamp), 이미지 절단 연산자(Image crop operator), 이미지 절단 타임스탬프(Image crop timestamp), 이미지 측정렬 연산자(Image de-skew operator), 이미지 측정렬 타임스탬프(Image de-skew timestamp), 프로세스 위치(Process location), Jp2 생성 타임스탬프(Jp2 creation timestamp), UUID 타임스탬프(UUID time stamp), 추출 타임스탬프(Embed timestamp), QA코드(QA Code)	-	-
디지털 대체물 추가 요소	총 1개 요소 이미지 압축(Image compression)	-	-

이와 별도로 기술 환경(technical environments)과 관련한 메타데이터 요소 8개도 획득하도록 설계되어 있다. 이 역시 마이크로 폼 혹은 사진자료를 대상으로 한 디지털화 프로젝트에서는 1개 요소를 추가한다. 기술 환경 메타데이터에 해당하는 전체 요소는 <표 4>와 같다. 「디지털화 문서」에서는 이들 요소에 대해 요소별 필드명, 데이터 포맷, 설명, 옵션 및 예시, 사유, 일관성 검사 여부 등으로 구성된 메타데이터 스키마를 제공하고 있다. 디지털 대체물의 경우에는 기술 환경 메타데이터를 적용하지 않는다.

〈표 4〉 디지털화 기록을 위한 기술 환경 메타데이터 요소(영국 TNA)

디지털화 기록에 대한 기술(technical) 환경 메타데이터 요소	(추가)마이크로폼 및 사진자료
총 8개 요소 배치코드(Batch code), 회사명(Company name), 이미지 축정렬 소프트웨어(Image de-skew software), 이미지 분할 소프트웨어(Image split software), 이미지 절단 소프트웨어(Image crop software), Jp2 생성 소프트웨어(Jp2 creation software), UUID 소프트웨어(UUID software), 임베드 소프트웨어(Embed software)	총 1개 요소 이미지 반전 소프트웨어(Image inversion software)

TNA 문서에서 발견되는 메타데이터 특징은 다음과 같다. 첫째, 저작권 문구 외 UUID, URI 등 디지털화 자원을 구분할 수 있는 여러 식별자를 활용한다는 점이다. 앞서 살펴본 미국 NARA의 메타데이터 규정에서 ‘식별자(Identifier)’ 요소를 상세구분을 이용하여 구분한 것과 달리, 영국은 ‘배치코드(Batch code)’와 함께 UUID, URI 요소를 별도로 활용한다. UUID 등은 단독으로 사용될 수도 있지만 ‘파일 UUID(File UUID)’와 같이 다른 요소의 상세구분으로도 활용된다. 실제 사례에서도 기본 식별자 외 UUID, URI 등 자동 생성 기반 식별자를 추가로 활용하는 것을 확인할 수 있다.

둘째, 디지털 자원의 특성에 따라 메타데이터 요소를 달리 구성하여 적용하였다. 메타데이터 작성 대상을 원본대체사본과 같은 성격의 디지털화 기록과 디지털 대체물을 구분하여 메타데이터 요소를 달리 적용하고 있다. 또한 종이문서가 아닌 마이크로 폼이나 사진자료에 대한 디지털화에서 적용하는 메타데이터 요소를 추가로 제시하고 있다.

마지막으로, 디지털화 기록의 영구적인 유지보수 및 관리를 위해 자동 추출 기반의 기술 메타데이터 요소를 적용하고 있다. 디지털화 기록 관리를 위한 기술 메타데이터 요소가 기술 환경에 대한 요소를 포함하여 최대 54개 정도로 많고, 대체로 자원 자체에서 자동으로 추출할 수 있는 요소로 구성되어 있다.

4. 디지털화 기록 관리를 위한 메타데이터 요소 설계

4.1 설계 시 고려사항

디지털화 기록의 특성과 관련 메타데이터 표준을 분석하여 도출한 디지털화 기록의 메타데이터 설계 방향은 다음과 같다.

4.1.1 원천 기록과 디지털화 기록의 관계 정립

먼저, 국내 표준에서 확인할 수 있었던 가장 큰 문제점은 원천 기록과 디지털화 기록 간의 관계를 제대로 정립하지 않은 채 기록관리 메타데이터 표준이 마련되어 있었다는 점이다. 「기록관리 메타데이터 표준(v2.2)」 중 상위요소 ‘유형’의 하위요소인 사본유형의 작성방법을 보면, 원본대체사본은 원천 기록과 구분하여 별도로 관리하도록 되어 있으나, 일반 사본은 원천 기록과 구분하여 작성하여야 한다는 내용 없이 다르게 표기할 수 있는 기술요소 몇 개만을 제시하는데 그치고 있다. 그 외 다른 요소에서도 디지털화 기록을 대상으로 한 작성방법 및 요령은 확인할 수 없다.

또한 현재는 비록 폐지된 표준이지만, 「NAK 26:2013(v1.0) 기록물 디지털화 지침-제1부: 종이 및 사진·필름」의 <부속서 B>에서도 ‘생산자’ 요소에 이미지 생산자를, ‘생산일시’ 요소에 이미지 생산일자를 입력하도록 한 반면, ‘생산이력’은 원천 기록의 것을 그대로 활용하도록 설계되어 있었다. 이를 적용하여 메타데이터를 작성하게 되면 디지털화 이미지의 생산자와 생산일자가 원천 기록의 생산정보와 같은 방식으로 작성되어 혼란을 야기할 가능성이 농후하다. 이 또한 원천 기록과 디지털화 기록의 관계가 제대로 정립되지 않은 채 기록관리 메타데이터 요소를 설계하였다는 데 문제의 원인이 있다.

이러한 이유에서, 디지털화 기록의 메타데이터 설계를 위해서는 먼저 디지털화 기록을 원천 기록과 구분되는 새로운 기록으로 간주할 것인지 혹은 원천 기록의 연장으로 만들어진 사본으로 볼 것인지 정하여야 한다. 3장에서 살펴본 미국, 영국 사례에서는 디지털화 기록의 내용, 맥락 등에 대한 정보를 원천 기록의 것을 그대로 받아오도록 하면서 디지털화 기록을 원천 기록의 사본으로 간주하고 있었다. 특히 영국은 디지털화 기록과 디지털 대체물로 구분하고 대상에 맞는 메타데이터 요소를 규정하고 있음을 확인할 수 있었다. 따라서 국내에서도 디지털화 기록을 원천 기록의 대체물로 보고, 이러한 관계 속에서 원천 기록으로부터 메타데이터를 승계받을 요소와 디지털화 기록을 위해 별도로 작성할 요소를 구분하는 것이 타당하다고 판단된다.

4.1.2 자원유형별 특성에 따른 자동추출 메타데이터 요소 확대

디지털화 기록의 큰 특징 중 하나는 종이, 아날로그 음성 파일부터 디지털 매체에 수록한 파일까지 모두 디지털화 대상이 될 수 있으며, 원천 기록의 자원유형이 하나가 아닌 둘 이상인 혼합매체일 가능성도 있다는 것이다. 그렇기 때문에 디지털화 기록 메타데이터는 여러 자원유형을 표현할 수 있도록 설계되어야 한다. 국내 디지털화 관련 표준에서도 디지털화에 사용된 하드웨어와 소프트웨어의 모델, 제조사나 압축 시 압축방법 등을 상세하게 기록으로 남기도록 규정하고 있다.

그러나 「기록관리 메타데이터 표준(v2.2)」에는 자원 특성을 반영할 수 있는 요소로 최상위 요소인 ‘포맷’이 있거나 하위요소가 ‘포맷명’, ‘포맷버전’, ‘생성 애플리케이션명’, ‘생성 애플리케이션버전’ 정도 밖에 없어 관련 정보를 충분히 기술하는 데에 한계가 있다. 또한 ‘크기’의 하위요소도 ‘용량’과 ‘단위’만 있어 디지털화 기록은 물론 원천 기록의 자원 특성까지 세세하게 설명하기에는 무리가 있다. 이처럼 자원유형에 대한 정보의 부족은 기록의 장기적인 관리와 식별을 어렵게 만드는 요인이 된다는 점에서 문제가 될 수 있다.

미국과 영국의 사례에서는 디지털화 기록 메타데이터에 기술(technical) 메타데이터가 큰 비중을 차지하고 있음을 알 수 있었다. 기술 메타데이터에는 파일의 크기, 포맷명, 버전, 이미지의 너비에서부터 픽셀당 색상 구성까지 상세하게 생성할 수 있도록 관련 요소가 구성되어 있었다. 특히 영국은 각 요소별로 UUID, URI와 같은 식별자를 활용한 상세구분을 활용하여 기술 정보를 폭넓게 수집할 수 있도록 설계되어 있었다. 대체로 종이문서의 디지털화 프로세스를 대상으로 하지만 마이크로 폼이나 사진을 디지털화하는 경우에는 해당 유형의 특성을 표현할 수 있도록 추가 요소가 정의되어 있었다. 또한 메타데이터 요소의 수가 상당히 많은 데 비해, 대부분의 요소가 디지털화 프로세스에서 별도의 작업 없이 자원이나 시스템을 통해 해당 값을 자동으로 추출할 수 있는 것으로 구성되어 있다는 점도 발견할 수 있었다. 하지만 「기록관리 메타데이터 표준(v2.2)」에서는 이러한 정보를 기록할 수 있는 기술요소가 부족하여 디지털화 기록을 관리하는데 제한이 있다. 따라서 디지털화 기록의 영구적인 관리를 위해서 자원유형별로 각종 기술정보가 체계적으로 수집되도록 메타데이터가 설계되어야 할 것이다.

4.1.3 업무과정 및 프로젝트에 대한 메타데이터 작성

마지막으로 「기록관리 메타데이터 표준(v2.2)」에서 확인할 수 있었던 문제점은 디지털화 프로세스를 설명하는 요소가 부족하다는 것이었다. 요소의 대부분이 자원 자체에 대한 정보로 구성되어 있었으며 디지털화 프로세스에 대한 요소는 일부에 불과하였다. 관련 표준이나 지침에서 작업자 등 디지털화 프로세스와 연관된 행위 주체에 관한 사항을 기록화하도록 지시하고 있으나, 이와 관련된 기록관리 메타데이터 요소는 ‘관리이력’과 ‘보존이력’의 세부요소인 ‘행위자’ 뿐이었다. 이 역시 세부요소로 ‘기관명’, ‘기관코드’, ‘부서명’, ‘부서코드’, ‘직위(직급)명’, ‘직위(직급)코드’ 등으로 정해져 있어 공공기관이 아닌 사업체 등 다른 행위주체를 기록화하기에 적절하지 않았다.

디지털화 프로세스에 대한 정보는 디지털화 기록의 생성뿐만 아니라 이어질 관리에서도 활용할 수 있는 중요한 정보이며, 반드시 프로세스의 결과물로 생성된 디지털화 기록과 연결되어 확인할 수 있어야 한다. 미국은 디지털화 기록을 NARA로 전송할 경우 수집하여야 하는 메타데이터를 따로 정해두고 디지털화 프로세스의 마지막 단계에서

해당 서버의 업로드 정보까지 획득하고 있었다. 영국은 회사명, 사용 소프트웨어 등 디지털화 기록 외 디지털화 기록을 둘러싼 기술 환경 메타데이터를 수집하도록 정하고 있다는 것을 알 수 있었다. 따라서 디지털화 기록의 메타데이터는 기록 자체뿐만 아니라 디지털화 프로세스와 관련된 다양한 정보를 관리할 수 있도록 설계되어야 할 것이다.

4.2 설계원칙 및 모형

지금까지 살펴본 설계방향을 고려하여 디지털화 기록 메타데이터 설계원칙을 도출한 결과는 다음과 같다.

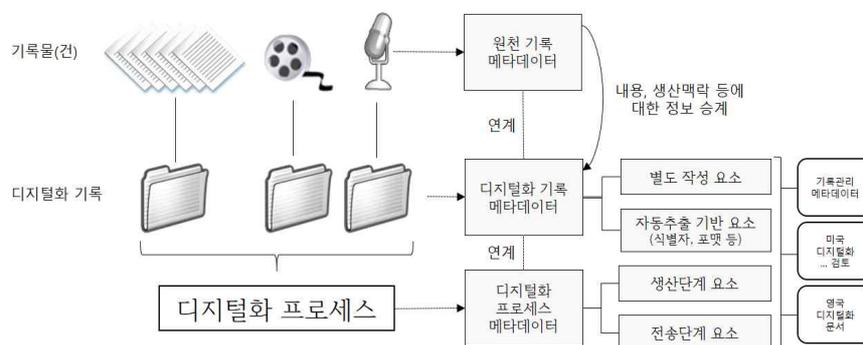
- 1) 디지털화 기록의 메타데이터는 건(item)별로 생성한다.
- 2) 디지털화 기록은 원천 기록의 사본으로 내용, 생산맥락 등에 대한 정보는 원천 기록과 동일하게 처리한다.
- 3) 디지털화 기록의 특성을 기록하고 관리하기 위해 별도의 메타데이터 요소를 개발한다.
- 4) 디지털 자원에 대한 기술적인 특성은 최대한 자동 추출 기반의 메타데이터 요소로 구성한다.
- 5) 기본식별자 외 자동 생성 기반의 식별자를 추가로 사용한다(예시 : UUID, URI).
- 6) 원천 기록이 혼합매체파일일 가능성을 고려하여 기술 메타데이터는 중복으로 기술할 수 있도록 한다.
- 7) 기록 자체에 대한 메타데이터 외 디지털화 프로세스(업무과정, 프로젝트)에 대한 메타데이터를 생성한다.
- 8) 관계요소를 매개로 원천 기록과 디지털화 기록, 디지털화 기록과 디지털화 프로세스의 관계를 유지한다.

디지털화 기록 관리를 위한 메타데이터 설계 모형은 아래의 <그림 3>과 같다.

우선 디지털화 기록 관리를 위한 메타데이터는 기존 시스템과의 연계 및 다른 메타데이터 표준과의 상호운용성을 고려하여 국내 표준인 「기록관리 메타데이터 표준(v2.2)」을 기반으로 한다. 「기록관리 메타데이터 표준(v2.2)」의 요소 중 내용, 생산맥락 등은 원천 기록 메타데이터로부터 승계받고, 자원유형이나 관리·보존이력 등 일부 요소는 디지털화 기록을 대상으로 별도로 생성하도록 한다. 하지만 현재의 「기록관리 메타데이터 표준(v2.2)」으로는 원천 기록과 디지털화 기록이 가질 수 있는 자원의 여러 유형을 표현할 수 없다는 한계가 있다. 따라서 필요한 요소를 추가적으로 개발하되, 새로운 요소의 개발은 메타데이터의 상호운용성 확보를 위해 미국과 영국 사례를 참조한다.

다음으로 업무과정, 프로젝트 등 디지털화 프로세스에 대한 메타데이터를 생성하도록 한다. 이 역시 「기록관리 메타데이터 표준(v2.2)」만으로는 충분히 기술될 수 없으므로 미국과 영국 사례로부터 관련 요소를 추가하고, 디지털화 기록의 각 시점에 활용할 요소를 구분하여 설계한다.

마지막으로 원천 기록, 디지털화 기록과 디지털화 프로세스의 세 개체가 메타데이터를 통해 서로 연계될 수 있도록 식별자, 관계 요소 등을 활용한다.



<그림 3> 디지털화 기록 메타데이터 설계 모형

4.3 메타데이터 설계(안)

<그림 3>의 디지털화 기록 메타데이터 설계 모형에 따라 디지털화 기록을 대상으로 하는 메타데이터 요소와 디지털화 프로세스 관련 메타데이터 요소를 제안하면 다음과 같다.

4.3.1 디지털화 기록 대상 메타데이터

앞서 3장에서 「기록관리 메타데이터 표준(v2.2)」을 기준으로 원천 기록과 동일하게 적용되는 메타데이터를 제외하고, 디지털화 기록에 적용해야 할 메타데이터를 추출한 결과는 모두 59개로 나타났다. 이들 요소를 디지털화 프로세스 중 자동으로 획득 가능한 요소와 메타데이터 작성자가 직접 입력해야 할 요소로 구분한 결과는 다음 <표 5>와 같다.

<표 5> 디지털화 기록 대상 메타데이터 요소

디지털화 프로세스 중 자동으로 획득(20)		메타데이터 작성자 입력(39)	
상위요소	하위요소 (세부요소는 ()안에 개수만 표시)	상위요소	하위요소 (세부요소는 ()안에 개수만 표시)
기록식별자	기본식별자, 시스템식별자(2), 보조식별자(2)	전자기록물 여부	-
포맷	포맷명, 포맷버전, 생성 애플리케이션명, 생성 애플리케이션버전	유형	사본 유형
매체	-	위치	소장처, 소장위치
크기	용량, 단위	관리이력	관리유형, 관리설명, 관리일시, 관리행위자(8), 변경요소(2)
무결성체크	무결성체크법, 무결성체크값	보존이력	보존처리 유형, 보존처리설명, 보존처리일시, 보존행위자(8)
		관계	관계유형, 관계대상식별자, 관계설명

<표 5>를 기준으로 하되, 디지털화 프로세스에서 자동으로 획득 가능한 요소는 더 추가될 필요가 있다. 먼저, ‘기록식별자’는 UUID와 URI 등을 활용하여 다른 메타데이터와 연계될 수 있는 형태로 자동 생성 식별자가 추가되어야 한다. ‘포맷’, ‘매체’, ‘크기’ 요소는 앞서 언급한 바와 같이 디지털화 기록의 특성을 반영하기에는 기재할 수 있는 정보가 많이 부족하기 때문에 기록관리를 위한 정보를 추가적으로 획득할 필요가 있다. 또한 ‘무결성체크’ 요소도 자원의 진본성, 신뢰성 등을 위해 시스템 혹은 사람에 의해 주기적으로 진행하되 이력이 남도록 해야 한다. 이러한 이유로, 디지털화 프로세스에서 자동으로 획득 가능한 메타데이터 요소 중 이 연구에서 제안하는 요소 확장(안)은 다음 <표 6>과 같다. 확장(안)에서 미국이나 영국 사례로부터 참조한 요소는 미국과 영국의 해당 표준명이나 각 표준에 정의된 요소명을 표기하는 대신에, 참조 사례인 미국의 ‘NARA’, 영국의 ‘TNA’로 표시하였다.

<표 6> 디지털화 기록 대상 자동추출 메타데이터 요소(안)

기술요소(43)		설명	참조 사례
상위요소	하위요소		
자원 URI(Resource URI)	-	예측 가능한 패턴을 가진 고유 식별자	TNA
파일 크기(File Size)	-	이미지 파일의 크기(바이트)	NARA
포맷명(Format Name)	-	파일 형식의 형식 이름 또는 설명	NARA
포맷버전(Format Version)	-	파일 형식의 버전	NARA
파일 UUID(File UUID)	-	모든 이미지에 포함된 이미지에 대한 범용 고유 식별자	TNA
파일 경로(File path)	-	배치를 포함하는 파일 시스템의 루트에 상대적인 파일 위치	TNA
이미지 너비(Image Width)	-	디지털 이미지의 너비(픽셀)(예: 수평 또는 X 치수)	NARA,TNA
이미지 높이(Image Height)	-	디지털 이미지의 높이(픽셀) (예: 수직 또는 Y 치수)	NARA,TNA

생성일시 (Date and Time Created)	-	디지털 이미지가 생성된 날짜 또는 시간	NARA
사용 스캐너와 모델 (Scanner Make and Model)	스캐너 연산자 (Scan operator)	이미지를 만드는 데 사용되는 스캐너의 제조업체 및 모델	NARA
	스캔 ID(Scan ID)	이미지를 생성한 스캐너를 사용하는 특정 운영자를 나타내는 코드. 공급업체가 디코딩할 수 있는 익명화된 코드	TNA
	스캔 ID(Scan ID)	이미지를 생성하는 데 사용되는 스캔 장치의 개별 식별자	TNA
	스캐너 위치(Scan location)	스캐너의 물리적 위치	TNA
스캔 소프트웨어와 버전 (Scanning Software Name and Version)	스캐너가 이미지를 만드는 데 사용하는 소프트웨어의 이름 및 버전	스캐너가 이미지를 만드는 데 사용하는 소프트웨어의 이름 및 버전	NARA
	스캔 소프트웨어 연산자 (Scanning Software operator)	이미지를 생성한 스캔 소프트웨어를 사용하는 특정 운영자를 나타내는 코드. 공급업체가 디코딩할 수 있는 익명화된 코드	TNA
	스캔 소프트웨어 ID (Scanning Software ID)	이미지를 생성하는 데 사용되는 스캔 소프트웨어의 개별 식별자	TNA
	스캔 소프트웨어 위치 (Scanning Software location)	스캔 소프트웨어의 물리적 위치	TNA
디지털 카메라와 모델 (Digital Camera Make and Model)	디지털 카메라 연산자 (Digital Camera operator)	이미지를 만드는 데 사용되는 디지털 카메라의 제조업체 및 모델	NARA
	디지털 카메라 연산자 (Digital Camera operator)	디지털카메라를 사용하는 특정 운영자를 나타내는 코드. 공급업체가 디코딩할 수 있는 익명화된 코드	TNA
	디지털 카메라 ID (Digital Camera ID)	이미지를 생성하는 데 사용되는 디지털 카메라의 개별 식별자	TNA
	디지털 카메라 위치 (Digital Camera location)	디지털 카메라의 물리적 위치	TNA
픽셀당 색상 구성(해상도) (Samples Per Pixel, Image resolution)	-	픽셀당 색상 구성 요소 수	NARA,TNA
이미지 색공간 (Image color space)	-	색공간, 좌표계에서 나타낼 수 있는 색상의 범위	TNA
이미지 색조해상도 (Image tonal resolution)	-	이미지의 색조해상도	TNA
이미지 압축 (Image compression)	-	이미지를 압축하는 데 사용되는 손실 이미지 압축 비율	TNA
이미지 분할(Image split)	-	이미지 분할의 결과인지 여부	TNA
이미지 분할 순서 (Image split ordinal)	-	합성물의 경우, 이미지나 영상의 순서	TNA
이미지 분할 소프트웨어 (Image split software)	-	이미지 분할에 사용한 소프트웨어	TNA (기술환경)
이미지 분할 UUID (Image split other UUID)	-	이미지가 분할된 경우, 이 이미지와 동일한 원래 이미지에서 분할된 다른 이미지의 UUID가 포함(분할의 일부로 이미지가 세 개 이상인 경우 이 필드에는 쉽표로 구분된 여러 UUID가 포함)	TNA
이미지 절단(Image crop)	-	이미지가 잘렸는지 여부와 수행된 자르기 유형인지 여부	TNA
이미지 절단 소프트웨어 (Image crop software)	-	이미지 자르기에 사용한 소프트웨어	TNA (기술환경)
이미지 측정렬 (Image de-skew)	-	이미지가 측정렬되었는지 여부	TNA
이미지 측정렬 (Image de-skew software)	-	이미지 측정렬에 사용한 소프트웨어	TNA (기술환경)

이미지 반전 (image inversion)	-	(마이크로폼 또는 사진 네거티브 프로젝트) 이미지가 음에서 양로 반전되었는지, 자동프로세스에 의해 수행되었는지 또는 수동으로 수행되었는지 여부를 지정	TNA
이미지 반전 연산자 (image inversion operator)	-	(마이크로폼 또는 사진 네거티브 프로젝트) 수동반전을 위해 이미지를 생성한 반전소프트웨어를 사용하는 특정 작업자, 공급업체가 디코딩할 수 있는 익명화된 코드	TNA
이미지 반전 타임스탬프 (image inversion timestamp)	-	(마이크로폼 또는 사진 네거티브 프로젝트) 이미지가 반전된 날짜 및 시간	TNA
이미지 반전 소프트웨어 (Image inversion software)	-	(마이크로폼 또는 사진 네거티브 프로젝트) 이미지 반전에 사용한 소프트웨어	TNA (기술환경)
피시 참조번호 (fiche reference)	-	(마이크로폼) 정보 처리용 마이크로카드나 필름류 참조번호	TNA
주기(Comments)	-	작업자의 코멘트	TNA
파일 체크섬(File checksum)	-	SHA256 표준을 준수하는 이미지 파일의 체크섬	TNA
메시지 다이제스트 알고리즘 (Message Digest Algorithm)	-	디지털 개체 또는 비트스트림에 대한 메시지 다이제스트를 구성하는 데 사용되는 특정 알고리즘	NARA
메시지 다이제스트(체크섬) (Message Digest (checksum))	-	메시지 요약 알고리즘의 출력	NARA

<표 6>에서 제안한 메타데이터 요소를 활용하여 가상의 메타데이터를 생성한 사례는 <표 7>과 같다. 사례에 모든 요소를 포함하기가 어려워 부득이 각 요소의 입력값과 방법을 이해하는데 필요한 최소한의 요소를 적용한 사례를 제시하였다. <표 7>에서 보듯이 디지털화 기록의 메타데이터 중 '기록물명'과 같이 원천 기록의 메타데이터로부터 획득되는 요소가 있고, '전자기록물 여부' 등과 같이 디지털화 기록을 대상으로 별도로 작성하여야 하는 요소가 있다. 그리고 '관계대상식별자'와 이 연구에서 추가로 제시한 요소인 '파일 UUID'는 원천 기록이나 관련 메타데이터와의 연결을 유지하는 역할을 한다. 마지막으로 '사용 스캐너와 모델'과 같은 요소는 기관이 미리 정한 지침에 따라 값이 자동으로 입력될 수 있다.

<표 7> 디지털화 기록 대상 메타데이터 생성 사례

구분	상위요소	하위요소	요소값	요소값에 대한 설명	입력방법
「기록 관리 메타 데이터 표준 (v2.2)」	4. 기록물명	4.1 제목	공무원 교육제도 개선	원천 기록의 제목	획득
	7. 전자기록물 여부	-	전자기록물	전자기록물 여부	작성
	8. 유형	8.2 사본 유형	디지털화 기록	사본 유형(원본대체 등)	작성
	22. 관계	22.1 관계 유형	원본	현재 기술하는 기록물이 사본(대체사본 또는 보존사본)이며 원본이 존재하는 경우, 입력	작성
		22.2 관계대상식별자	00000000-0000-0000-0000-000000000000	자동 부여 혹은 직접 입력 (파일 UUID 등)	작성
22.3 관계설명		디지털화 기록, 사본	원천 기록과의 관계 설명	작성	
추가 요소	파일 UUID	-	00000000-0000-0000-0000-000000000000	관계식별자의 값으로 연계	자동추출
	사용 스캐너와 모델	스캐너 연산자	ABG001	미리 정한 지침에 따라 부여	자동추출
		스캔 ID	002A	미리 정한 지침에 따라 부여	자동추출
		스캐너 위치	○○시 ○○구 ○○로, 000호	미리 정한 지침에 따라 부여	자동추출

4.3.2 디지털화 프로세스 대상 메타데이터

현행 「기록관리 메타데이터 표준(v2.2)」에서 디지털화 프로세스 정보를 표현할 수 있는 요소는 ‘관리이력’과 ‘보존이력’ 정도에 불과하다. 하지만 ‘관리이력’과 ‘보존이력’의 하위요소인 ‘유형’, ‘설명’, ‘일시’, ‘행위자’, ‘변경요소’만으로는 디지털화 프로세스의 전반적인 사항을 반영하기는 어렵다. 또한 ‘행위자’ 요소의 세부요소로 ‘기관명’, ‘기관코드’, ‘부서명’, ‘부서코드’, ‘개인명’, ‘개인코드’, ‘직위(직급)명’ 및 ‘코드’로 구성되어 있어 공공기관이 아닌 다른 주체에 대한 정보를 입력하기에 적절하지 않은 상태이다.

이와 관련하여, 영국 TNA의 「디지털화 문서」(2016)에서는 기술환경 메타데이터를 별도로 작성하여 디지털화를 진행하는 사업체와 주요 사용 소프트웨어를 기록화할 수 있는 메타데이터를 적용하고 있었다. 미국 NARA의 규정에는 디지털화 기록의 생산 프로세스에 대한 메타데이터가 마련되어 있지는 않았지만 NARA로 전송하는 과정에서 생성하여야 할 메타데이터를 별도로 정하고 있었다. 이러한 두 국가의 사례를 혼합하여 이 연구에서는 다음 <표 8>과 같이 디지털화 프로세스 대상 메타데이터 요소(안)를 생산단계와 전송단계로 나누어 제안하고자 한다.

<표 8> 디지털화 프로세스 대상 메타데이터 요소(안)

구분	기술요소(14)		설명	참조 사례
	상위요소	하위요소		
생산	생성 조직(회사명) (Creating Organization, Company name)	-	프로세스를 진행한 사업자명, 조직 등 (주요) 작업 중인 경우 컬렉션, 시리즈 또는 세트의 생성, 누적 또는 유지 관리를 담당하는 조직의 이름	NARA
	UUID 소프트웨어 (UUID software)	-	UUID 생산에 사용한 소프트웨어	TNA
	추출 소프트웨어 (Embed software)	-	디지털화 프로세스에서 사용한 추출 소프트웨어	TNA
전송	전송명(Transfer Title)	-	보존서버 및 장치로 전송할 기록 컬렉션, 세트 또는 시리즈에 할당된 이름	NARA
	날짜(Dates)	시작일자 (Inclusive Start Date)	사용자가 보존서버 및 장치로 전송하려는 기록 그룹, 컬렉션, 시리즈 또는 세트가 작성자에 의해 생성, 유지 또는 누적된 시작 날짜	NARA
		마감일자 (Inclusive End Date)	작성자가 보존서버 및 장치로 전송하려는 기록 그룹, 컬렉션, 시리즈 또는 세트를 생성, 유지 또는 누적한 마지막 날짜	NARA
	전송 조직(회사명) (Creating Organization, Company name)	-	전송시 컬렉션, 시리즈 또는 세트의 생성, 누적 또는 유지 관리를 담당하는 조직의 이름	NARA
	기록물 군 번호 (Record Group Number)	-	기록 그룹에 할당된 고유 번호	NARA
	일반적인 기록 유형 (General Records Type)	-	전송 중인 기록 세트, 시리즈 또는 컬렉션의 일반적인 형식	NARA
	접근제한 (Access Restrictions)	접근제한상태 (Access Restriction Status)	보존서버 및 장치로 전송할 기록 세트, 컬렉션 또는 시리즈에 대한 액세스 제한 여부	NARA
		특정접근제한 (Specific Access Restriction)	국가 보안 고려사항, 기증자 제한, 법원 명령 및 기타 법적 또는 규제 조항(개인정보보호법 등)을 기반으로 한 일련의 기록에 대한 특정 접근 제한	NARA
	이용제한 (Use Restrictions)	이용제한상태 (Use Restriction Status)	보존서버 및 장치로 전송할 기록 세트, 컬렉션 또는 시리즈에 대한 사용 제한 여부	NARA
		특정사용제한 (Specific Use Restriction)	개인정보보호법 면제를 포함한 저작권, 상표, 서비스 표시, 기증자 또는 법적 조항에 따라 기록 세트, 컬렉션 또는 시리즈에 대한 사용 제한 유형	NARA
기록물 일정 번호 (Record Schedule Number)	기록물 일정 번호 (Records Schedule Number)	전송 중인 컬렉션, 시리즈 또는 세트의 모든 기록에 적용되는 기록 스케줄에 할당된 보존서버 및 장치 번호	NARA	

생산단계 메타데이터는 앞서 디지털화 기록을 대상으로 한 메타데이터에 포함된 ‘이미지 분할 소프트웨어’ 등 4개의 요소를 제외한 3개의 요소로 구성하였다. 사업체 및 조직명, 프로세스 전반에 적용하는 소프트웨어에 대한 정보는 기존 기록관리 메타데이터의 ‘관리이력’, ‘보존이력’ 요소와 결합하여 생산단계 프로세스를 설명해주는 역할을 수행한다. 전송단계 메타데이터는 디지털화 기록을 보관하는 보존서버 및 장치로의 전송 시 자동으로 생성되거나 획득될 수 있는 정보로 전송 시 할당된 이름, 일시, 주체, 접속 및 사용조건 등이 포함된다.

디지털화 프로세스 대상 메타데이터를 활용하여 가상의 메타데이터를 생성한 사례는 아래의 <표 8>과 같다. 디지털화 기록 대상 메타데이터와 같이 모든 요소를 반영하기 어려워 이해에 필요한 최소한의 요소만 포함하였다. 각 요소는 디지털화 프로세스에서의 행위주체나 관계자, 사용한 도구 등에 대한 정보를 포함한다. 그리고 ‘기록물 군 번호’ 등 식별자 요소를 활용하여 원천 기록이나 디지털화 기록과의 관계정보를 파악할 수 있다.

<표 9> 디지털화 프로세스 대상 메타데이터 생성 사례

구분	상위요소	하위요소	요소값	요소값에 대한 설명	입력방법
생산	생성조직(회사명)	-	○○○○사	디지털화를 진행한 주체	작성
	UUID 소프트웨어	-	Oracle Java JDK 1.6	UUID 언어 및 소프트웨어	작성
	추출 소프트웨어	-	Luratech Lurawave 11a	사용 소프트웨어	작성
전송	전송명	-	2022년도 제1차 디지털화 사업	사업명 등	작성
	날짜	시작일자	2022-01-01	전송시작일자	자동추출
		마감일자	2022-12-31	전송마감일자	자동추출
	전송조직(회사명)	-	□□□□사	파일을 전송한 주체	작성
	기록물 군 번호	-	NA10	원천 기록 혹은 디지털화 기록물 군에 부여된 번호	작성

5. 결론

공공기관을 비롯하여 기록을 생산, 관리하는 많은 기관에서는 비전자기록의 활용 및 보존을 위해서 디지털화 사업을 추진해오고 있다. 디지털화 사업은 비전자 형태의 여러 기록을 디지털화하여 보다 많은 사람이 동시에 해당 기록을 사용할 수 있도록 하며, 시간이 지남에 따라 점차 훼손되어 가는 기록을 안전하게 보존할 수 있도록 하는 방안이다. 나아가 전자기록 중심의 기록관리 체계에 맞춰 여러 유형의 기록을 통합적으로 관리하기 위한 노력이기도 하다.

하지만 개별 기관에서 크고 작은 디지털화 사업을 추진하면서 기록의 체계적인 관리를 위해 메타데이터 생성이 제대로 이루어지고 있는지는 확인하기가 어렵다. 2018년 이전까지는 「NAK 26:2013(v1.0) 기록물 디지털화 지침-제1부: 종이 및 사진·필름」의 <부속서 B>에서 제시한 디지털화 기록 메타데이터를 활용할 수 있었으나, 실질적으로 메타데이터에 생성하는데 턱없이 부족하다는 것을 이 연구를 통해 확인할 수 있었다. 하지만 그마저도 「NAK 26:2018(v2.0) 기록물 디지털화 기준」으로 개정되면서 메타데이터 관련 부속서가 삭제되어 디지털화 프로세스에서 메타데이터 작성의 의무는 없어지게 된 상황이다.

이 연구는 디지털화 기록의 특성에 대한 분석과 해외 사례와의 비교를 통해 디지털화 기록의 관리 및 품질 유지를 위해 필요한 메타데이터 요소(안)을 제안하는데 목적을 두었다. 이를 위해 이 연구에서는 먼저 디지털화 기록과 디지털화 기록 메타데이터에 대한 개념을 정립하고, 「NAK 8:2021(v2.2) 기록관리 메타데이터 표준」에서 디지털화 기록에 적용 가능한 요소만 별도로 추출하는 과정을 수행하였다. 「기록관리 메타데이터 표준(v2.2)」은 현재 디지털화 기록에 대한 메타데이터로 활용할 수 있는 국내 유일의 표준이지만, 디지털화 기록에 대한 메타데이터를 별도로 제시하고 있지 않아 부득이 분석을 통해 추출하는 과정을 거쳐야 했기 때문이다. 다음으로, 이 결과를 해외 사례와

비교하여 「기록관리 메타데이터 표준(v2.2)」의 문제점 및 메타데이터 요소(안) 설계 시 고려할 점을 도출하였다. 이 과정에서 비교 대상으로 삼은 해외 사례는 미국 NARA의 「연방기록관리 : 영구기록의 디지털화와 레코드 스케줄 검토」(2020)와 영국 TNA의 「디지털화 문서」(2016년 최종 개정)였다.

이 연구에서 발견된 주요 내용은 다음과 같다. 첫째, 디지털화 기록은 자원의 형태만 다른 원천 기록의 사본으로 간주되므로 원천 기록의 정보를 승계하면서 원천 기록-디지털화 기록-디지털화 프로세스에 대한 메타데이터가 각각 생산되고 상호 연계될 필요가 있다. 그러나 현행 「기록관리 메타데이터 표준(v2.2)」에서는 원천 기록과 디지털화 기록의 관계를 제대로 정립하지 않았으며, 기록의 다양한 유형을 설명하는데 필요한 요소도 많이 부족하였다. 둘째, 현행 「기록관리 메타데이터 표준(v2.2)」은 자원 중심의 기술요소로만 구성되어 있어 디지털화 프로세스에 대한 정보를 수용할 수 없는 구조였다. 그에 비해 미국과 영국은 디지털화 기록의 정의를 원천 기록의 사본으로 간주하고 내용과 맥락 정보는 원천 기록의 메타데이터를 그대로 활용하되 각종 기술 정보를 자동으로 추출하여 기록의 영구적인 보존을 가능케 하였다. 또한 전송과정이나 기술환경을 대상으로 한 메타데이터 작성으로 디지털화 프로세스에 대한 전반적인 정보를 수집, 생성할 수 있도록 메타데이터가 설계되어 있었다.

이러한 분석 결과를 토대로 하여 이 연구에서는 디지털화 기록의 메타데이터 설계 방향, 원칙과 모델을 제안하였다. 또한 이에 근거하여 메타데이터(안)을 도출하여 제안하되, 디지털화 기록 자체에 관한 메타데이터와 디지털화 프로세스에 관한 메타데이터로 구분하여 각각의 요소를 제안하였다. 디지털화 기록 대상 메타데이터는 「기록관리 메타데이터 표준(v2.2)」을 활용하되 디지털화 기록을 중심으로 별도로 생성하여야 할 요소와 원천 기록으로부터 정보를 획득할 요소를 구분하였고, 해외 사례를 참조하여 디지털화 기록에서 자동추출하여 활용할 메타데이터 요소를 추가적으로 제안하였다. 디지털화 프로세스 대상 메타데이터 역시 해외 사례를 참조하여 생산단계와 전송단계로 구분하여 메타데이터 요소(안)을 제안하였다.

이 연구는 디지털화 기록에 대한 메타데이터 설계의 필요성과 그와 관련된 메타데이터 요소 도출 과정까지를 범위로 하고 있다. 향후 표준을 개발하는 과정에서 세부 요소에 대한 전문가 검증 및 다양한 유형의 디지털화 기록을 대상으로 하는 적용 테스트 등을 거쳐야 할 것이지만, 원천 기록과 디지털화 기록의 관계를 정립하여 이를 메타데이터 요소(안)에 반영하였고, 선행연구에서 그 필요성을 언급해 왔던 디지털화 기록에 대한 메타데이터를 구체적으로 도출해보았다는 점에서 연구의 의의가 있다. 여러 기관에서 이때까지 진행하였고, 앞으로도 계속 이어질 많은 디지털화 사업이 보다 성공적으로 추진되기 위해서는 디지털화 기록 관리에 대한 보다 폭넓고 실리적인 논의가 필요하다. 이 연구가 디지털화 기록의 메타데이터에 대한 학계와 현장의 관심을 불러일으키는 촉매제가 되기를 기대한다.

참고문헌

- 기록관리 메타데이터 표준(v2.2). NAK/S 8:2021(v2.2).
기록관리시스템 데이터연계 기술규격-제2부 중앙연구기록관리시스템과의 연계(v1.2). NAK 29-2:2016(v1.2).
기록물 디지털화 기준(v2.0). NAK 26:2018(v2.0).
기록물 디지털화 지침-제1부: 종이 및 사진·필름(v1.0). NAK 26:2013(v1.0).
기록물 목록 작성 및 디지털화 작업 지침(v1.2). NAK 23:2017(v1.2).
김도환 (2021). 공공기록물 디지털화 사업 만족도 평가와 개선방안 연구 : 경상남도 기초지방자치단체 디지털화 사업을 중심으로 석사학위논문, 동의대학교 대학원 문헌정보·사학과.
녹음·동영상 기록물 디지털화 지침(v1.0). NAK/A 17:2014(v1.0).
류재현 (2017). 공공기관 중요기록물의 디지털화 프로세스 개선 방안 연구. 석사학위논문, 한신대학교 기록관리학(협)전공.
문헌정보 - 기록관리 - 제1부: 개념과 원칙. KS X ISO 15489-1:2016.

문헌정보 - 기록관리과정 - 기록메타데이터 - 제1부 : 원칙. KS X ISO 23081-1:2017.
 문헌정보-기록의 디지털화 이행 지침. KS X ISO TR 13028:2010.
 박은경 (2005). 디지털이제이전의 디지털 이미지와 메타데이터 표준화작업. 한국기록관리학회지, 5(2), 139-152.
<https://doi.org/10.14404/JKSARM.2005.5.2.139>
 안지현 (2014). 공공기록물 디지털화에 관한 연구 : 지방자치단체 구기록물 DB구축 사업을 중심으로. 석사학위논문, 중부대학교 대학원 기록물관리학과.
 이선경 (2008). 디지털화 기록의 품질 확보 방안. 석사학위논문, 한국외국어대학교 대학원 정보·기록관리학과.
 이에경 (2007). 우리나라 행정기록물의 디지털화에 관한 연구 : 행정기관 기록물 DB구축 사업을 중심으로. 석사학위논문, 한국외국어대학교 대학원 정보·기록관리학과.
 이유정 (2016). 지방자치단체 비전자기록물의 디지털화에 관한 연구 : 대구광역시 8개 기초자치단체를 중심으로. 석사학위논문, 대구가톨릭대학교 대학원 기록관리학과.
 임나영, 남영준 (2019). 기록의 디지털화 기준에 관한 연구. 한국비블리아학회지, 30(3), 5-30.
<https://doi.org/10.14699/kbiblia.2019.30.3.005>
 한희정, 김태영, 김용 (2016). 다중개체모형을 적용한 무형문화유산 메타데이터 요소 개발에 관한 연구. 한국문헌정보학회지, 50(3), 329-359. <https://doi.org/10.4275/KSLIS.2016.50.3.329>
 Federal Records Management: Digitizing Permanent Records and Reviewing Records Schedules (RIN 3095-AB99)
 Government Records Service, Province of British Columbia (2020). Digitizing Government Information Guide. BRITISH COLUMBIA. Available :
<https://www2.gov.bc.ca/assets/gov/british-columbians-our-governments/services-policies-for-government/information-management-technology/records-management/guides/digitizing.pdf>
 NARA (2022). Digitization at the National Archives. The U.S. National Archives and Records Administration. Available: <https://www.archives.gov/digitization>
 NARA (2022). Guidelines for Digitizing Archival Materials for Electronic Access. The U.S. National Archives and Records Administration. Available:
<https://www.archives.gov/preservation/technical/guidelines.html>
 TNA (2016). Digitisation at The National Archives. THE NATIONAL ARCHIVES. Available:
<https://www.nationalarchives.gov.uk/about/commercial-opportunities/digitisation-services/>

• 국문 참고자료의 영어 표기
 (English translation / romanization of references originally written in Korean)

An, Ji-Hyeon (2014). A Study on digitization of public records : focused on records DB construction business. Master's thesis. Dept. of Archives Management Graduate School of Joongbu University.
 Criteria for Digitization of Records(v2.0). NAK 26:2018(v2.0).
 Guideline for Digitization of Records. - Part 1: Papers, Photographs and Films.(v1.0). NAK 26:2013(v1.0).
 Guideline for Records Cataloging and Digitization Project(v1.2). NAK 23:2017(v1.2).
 Guidelines for Digitization of Audio Recording Records and Video Records(v1.0). 2014(v1.0).
 Han, Hui-Jeong, Kim, Tae-Young & Kim, Yong (2016). A Study on the Development of Metadata Schema for Intangible Cultural Heritage Based on Multiple Entity Model. Journal of Korean Society for Library and Information Science, 50(3), 329-359. <https://doi.org/10.4275/KSLIS.2016.50.3.329>
 Information and documentation - Implementation guidelines for digitisation of records. KS X ISO TR 13028:2010.
 Information and documentation - Records management - Part 1: Concepts and principles. KS X ISO 15489-1:2016.

- Information and documentation – Records management processes – Metadata for records – Part 1: Principles. KS X ISO 23081-1:2017.
- Kim, Dohwan (2021). A Study on Improvement Actions based on Satisfaction Evaluation of Public Records Digitization Project- Focusing on the Digitization Project of Basic Unit of Local Government in Gyeongsangnam-do. Master's thesis, Dept. of Library and Information Science & History, Graduate School, Dong Eui University.
- Lee, Sun-Kyung (2008). Developing strategies for quality of digitized record. Master's thesis, Department of Information and Records Management, The Graduate School of Hankook University of Foreign Studies.
- Lee, Ye Kyoung (2007). A Study on digitization of the administrative records - focused on administrative organization records DB construction business.-. Master's thesis, Department of Information and Records Management, The Graduate School of Hankook University of Foreign Studies.
- Lee, You Jeong (2016). A Study on Digitization of the Nonelectronic Records of Local Government : Focusing on 8 Local Self Government of Daegu. Master's thesis, Department of Interdisciplinary Program in Archival Studies The Graduate School, Catholic University of Daegu.
- Metadata Standard for Records and Archives Management (Version 2.2). NAK/S 8:2021(v2.2)
- Lim, Nayoung & Nam, Youngjoon (2019). A study on the criteria for digitization of records. Journal of the Korean BIBLIA Society for library and Information Science, 30(3), 5-30.
<https://doi.org/10.14699/kbiblia.2019.30.3.005>
- Park, Eun G (2005). Imaging and Metadata Standards in Digitization: Practical Strategies. Journal of Korean Society of Archives and Records Management, 5(2), 139-152.
<https://doi.org/10.14404/JKSARM.2005.5.2.139>
- Ryu, Jae Hyun (2017). A Study on the Digitization Process and Improvement Plan of Important Records in Public Institutions. Master's thesis, Hanshin University.
- Technical Specification for the Data Interchange of the Records Management SystemPart 2: With the Archives Management System(v1.2). NAK 29-2:2016(v1.2).