

CRMS 도입을 맞아 생각해보는 디지털 시대의 기록관리

Considerations of the Record Management of the Digital Age While CRMS was Introduced

임진희(Jin-Hee Yim)

E-mail: yimjh68@seoul.go.kr

서울특별시 정보공개정책과



OPEN ACCESS

초 록

최근 중앙행정기관은 업무관리시스템을 클라우드 기반의 온나라 문서2.0으로 변경하여 사용하고 있다. 이에 맞춰 국가기록원은 클라우드 기반의 기록관리시스템을 보급하고 있다. 클라우드 컴퓨팅을 포함한 디지털 기술의 발전에 따라 기록의 보존과 활용 방안도 이에 조응하면서 효과적·효율적으로 이루어지도록 끊임없이 재설계해 나가야 한다. 전자기록 관리의 프로세스와 방법이 종이기록 관리의 단순 전자화에서 탈피하여 디지털 기술에 따른 변환(Transformation)의 관점에서 변화하길 바란다. 논리적 이관, 디지털객체의 생애주기별 저장과 중복성 제거, 공문서 서식의 기계가독형 설계, 빅데이터 분석이 용이한 포맷, 공문서 서식과 포맷의 문제, UUID와 해쉬값 기반의 진본성 인증체계 등의 도입을 제안한다.

© 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.

ABSTRACT

Recently, the central government organizations have changed their Business Management System to the cloud-based On-nara Document 2.0. According to this, the National Archives of Korea is spreading a cloud-based records management system. With the development of digital technology, including cloud computing, preservation and utilization of records must be redesigned continuously to be effective and efficient. It is needed that the process and method of the electronic records management will change from simple digitization of paper-based recording to digital technology. This article offers opinions related to the logical transfer, storage and redundancy elimination of digital objects, machine-readable format, big-data analysis, templates of official documents, and authenticity authentication system based on universally unique identifiers (UUID) and hash value.

Keywords: 기록 서식, 기록 식별체계, 클라우드 저장소, 콘텐츠 중복, 디지털 변환 record template, record identifier, cloud storage, contents redundancy, digital transformation

1. 클라우드 기반 전자기록 관리 동향

중앙정부가 G-클라우드를 활용하고 있고, 서울시와 같은 지방정부에서도 클라우드 컴퓨팅 환경을 점차 확대하여 활용하고 있다. 공공 영역에서는 행안부가 클라우드서비스 기반의 온나라 문서2.0, 클라우드 행정 포털 서비스를 개발하여 중앙행정기관에 모두 보급하였고, 지자체에 확산 보급하기 위해 노력 중이다. 이에 조응하여 국가기록원은 온나라 문서2.0에 짝하는 클라우드기록관리시스템 (CRMS, Cloud Records Management System)을 개발하여 보급하였다. 기록관리 관점에서 온나라 문서2.0에서 가장 획기적인 혁신은 결재문서 본문을 HWP 포맷에서 ODT로 전환했다는 점, PDF/A-1로도 저장을 한다는 점이다.

디지털 시대의 기록관리를 논의할 때 비평의 기준은 기록관리 도구들이 디지털 속성에 친화적이나 여부로 둔다. 즉, 아날로그적 원리를 표면적으로만 디지털화한 것은 근본적인 한계를 노정하게 된다는 점에 주목한다. 또한, 향후 미래 기술로 떠오르는 기술들과의 호환성, 연계성, 확장성을 보장하는 쪽이어야 유리하다는 점에 주목한다.

디지털 환경에 맞춰 전자기록을 관리하고자 할 때 아날로그의 전자화 수준이 아니고 디지털의 본성에 맞춰 사고할 필요가 있다. 예를 들어, 클라우드 컴퓨팅 기술이 보편화되면 그 기술을 기반으로 기록이 생산되고 관리되는 과정을 면밀히 살펴보고, 자연스럽게 전자기록이 흐르면서 이용되도록 설계하고, 기록관리에서 특별히 신경써야할 진본성과 장기보존이라는 목표에 유리하도록 검토하는게 필요하다.

클라우드의 장점 중 기록관리와 연관되는 점은 클라우드 스토리지를 활용하여 기록객체가 여러 어플리케이션에서 공유될 수 있다는 점이다. 그간 전자기록을 종이기록처럼 물리적으로 이관해온 관행을 극복할 수 있는 보다 확실한 기술적 기반이 마련된 것이라 볼 수 있다. 물리적 이관은 그 자체로 비용이 많이 들고, 무결성이 훼손될 위험에 노출되는 과정이므로 기술적 기반이 따라준다면 물리적 이관을 최소화하여 관리하는 것이 유리하다.

이 글에서는 디지털 기술의 특성을 적극 반영하여 기록관리를 효과적으로 할 수 있는 몇 가지 방법을 제안하고자 한다.

2. 이관 및 저장 방식 들여다보기

클라우드 서비스의 기본 중 하나는 인프라로서 저장소를 제공한다는 것이다. 이 때, 클라우드 저장소는 여러 어플리케이션들이 쉽게 공유할 수 있도록 설계할 수 있고 논리적인 볼륨으로 나누어 관리할 수 있다. 이러한 강점을 기록관리 분야에 적용한다면 첫째, 전자기록의 이관과정을 단순화할 수 있고 둘째, 전자기록의 생애주기에 맞춰 저장 공간을 할당할 수 있게 될 것이다.

먼저, 이관과정의 단순화이다. 이관과정이 단순해진다는 것은 소위 '물리적 이관' 대신 '논리적 이관'을 하여 얻는 이점을 의미한다. 그렇다면, '물리적 이관', '논리적 이관'이란 어떤 차이를 의미일까? 물리적 이관이란 기록물 객체를 다른 곳으로 이동시키면서 관리권한을 이양하는 방식이라면, 논리적 이관이란 기록물 객체는 이동하지 않고 관리권한을 이양하는 방식이라고 설명할 수 있겠다. 종이기록의 경우에도 전자기록의 경우에도 두 가지 방식의 이관이 가능하나 현실적으로는 종이기록물의 경우에는 처리과별 캐비닛에 흩어져 있던 기록물들을 기록관의 통합 서고로 이동하거나, 여러 기록관에 흩어져 있던 기록물들을 아카이브의 통합 수장고로 이동하는 방식의 '물리적 이관'이 일반적이라 볼 수 있다.

전자기록의 경우에도 현재 공공기관에서는 처리과에서 사용하는 업무시스템 서버에 저장된 기록을 기록관에서 사용하는 기록관리시스템 서버에 저장하기 위해 디지털 콘텐츠를 이동시키는 '물리적 이관' 방식을 취하고 있다. 여기서 주목해야 할 것은 전자기록의 경우 관리상의 기본 목표가 진본성과 이용가능성을 유지하는 것이라 했을 때, 물리적 이관을 거치게 되면 그 과정에서 진본성과 이용가능성을 훼손할 위험이 증대하므로 불리해진다는 것이다. 따라서, 논리적 이관이 가능한 상황이라면 그 방향을 취하는 것이 필요하다고 본다. 그런데, 원한다고 언제든 전자기록의 논리적 이관이 가능한게 아니다. 논리적 이관을 가능하려면 하나의 기록물저장소를 업무시스템도 기록관리시스템도 접근이 가능해야 하고, 이관된 기록물과 이관되지 않은 기록물을 구분하여 관리권한을 통제할 수 있어야 한다.

클라우드 통합저장소가 제공되고 업무시스템과 기록관리시스템이 동시에 해당 저장소에 접근할 수 있는 환경이 제공된다면, 업무시스템에서 생산한 기록이 클라우드 통합저장소에 저장된 후 이관 시점이 되었을 때 기록관리시스템에 해당 기록에 대한 관리권한만 이양함으로써 간단하게 이관을 종결할 수 있게 된다. 디지털 객체를 물리적으로 이동하는 과정은 쉽게 복사하여 붙이기(copy & paste)로 이해할 수 있다. 그런데, 이동하는 두 시스템의 특성이 다르고 상호운용성이 확보되어 있지 않은 경우에는 그리 간단치가 않다. 현재 국가기록원이 배포한 이관규격서를 보면 알 수 있다. 업무관리시스템에서 표준기록관리시스템으로 이관할 때, 전자문서시스템에서 표준기록관리시스템으로 이관할 때, 표준기록관리시스템에서 중앙기록관리시스템으로 이관할 때 등 각각의 경우 이관할 기록정보를 내보내기(export)하는 방법과 들여오기(import)하는 방법을 구체적으로 명시하고 있다. 이 규격에 맞춰 이관에 필요한 소프트웨어를 만들어 사용해야 한다. 온라인으로 기록정보를 송수신할 것인지 오프라인 매체로 기록정보를 이동할 것인지에 따라서 차이가 있을 수 있다. 내보내기를 하고 나서 빠짐없이 정확하게 기록정보가 추출되었는지 검증해야하고, 들여오기 하기 전에 임시 저장소에서 오류없이 신뢰할 만한 기록정보가 온 것인지 검증해야 한다. 물리적 이관에서는 상호간에 프로토콜을 맞추고 검증해야하는 절차가 필요하고 이에 소요되는 비용이 막대하다.

다음으로, 생애주기에 기반한 저장전략이다. 기록의 생애주기와 정보시스템의 생애주기를 연관하여 관리하는 것이 필요하다. 저장소 관점에서 적용한다면, 기록관에서는 장기보존 대상과 유한기간

보존 대상을 저장소 상에서 분리하는 것을 고려할 수 있다.

유한기간 보존 대상 디지털 객체들을 저장하는 저장소는 주기적으로 비트스트림이 신규 유입되고 삭제 폐기되는 과정을 거치게 된다. 따라서, 잦은 파일 입출력과 저장 삭제에 따른 디스크 조각(Fragmentation)을 관리하기에 용이한 스토리지를 배정하는 것이 필요하다. 장기보존 대상 디지털 객체들은 가능하면 그 저장소에서 저장매체가 노후화되어 교체할 때까지 이동없이 보존되는 것이 여러모로 유리하다. 비트스트림이 다른 저장공간으로 이동되는 시점에 무결성 훼손의 위험도가 높아지기 때문이다.

단일한 하나의 저장소를 사용하는 체계에서는 의미가 없지만 다수의 디스크를 엮어서 논리적 개념의 저장소를 만드는 클라우드 체계에서는 또 다른 접근도 가능해진다. 클라우드 저장소의 경우 하이브리드한 디스크들을 엮어서 논리적인 하나의 저장소를 제공하게 된다. 이 때, 논리적인 저장공간을 볼륨으로 나누어 정할 수 있는데, 단기간 보존하는 기록은 상대적으로 저렴한 디스크 영역에 보관하고 장기보존 대상 기록을 위해서는 장기간 보존 안정성을 제공해주는 고비용 스토리지 영역을 할당하는 방식으로 저장소를 운영할 수 있다.

이 경우, 장기보존 기록의 저장매체 리프레쉬 횟수를 줄일 수 있어 그만큼 물리적 이관에 따르는 위험성을 최소화할 수 있을 것이다. 단기간 보존하는 기록의 경우, 보존기한이 만료되는 연도가 같은 기록을 같은 저장소 영역에 저장하도록 한다면 폐기 결정 후 데이터를 삭제하는 과정이 훨씬 용이해지고 저장소를 재활용하기에도 용이해질 것이다. 아카이브의 경우에는 지속적으로 기록을 이관받아옴으로써 저장소가 계속 증설되어야 하는데 클라우드 스토리지의 자동 스케일아웃 기능을 활용하여 기록을 유연하게 저장할 수 있을 것이다.

기록관리시스템을 설계할 때, 위에서 언급한 요건들에 따라 물리적인 디스크들을 논리적인 저장소 영역으로 구분하고, 디지털 객체인 콘텐츠를 보존기간이나 공개여부, 활용도 등에 따라 적절한 위치에 저장하고, 디지털 객체의 원본과 사본들의 관계를 관리하고, 접근 요청이 들어왔을 때 특정 콘텐츠를 찾아주고, 수시로 무결성 검증을 할 수 있도록 하는 어플리케이션을 별도의 서비스로 만들 필요가 있다. MoReq2010에서 지칭한 콘텐츠서비스(Content Service)가 여기에 해당한다.

3. 공문서 서식과 포맷 들여다보기

서울시에서는 800여 종이 넘는 서식이 등록되어 사용되고 있다. 디지털 시대에 공문서의 서식은 어떻게 정의되어야 하는지 좀 더 정치하게 고민될 필요가 있다. 여기서 살펴보고자 하는 것은 공문서의 원본성 추정 시 서식이 어떤 역할을 해야 하나, 디지털 방식으로 서식을 운영한다는 것은 무엇인가, 향후 빅데이터로 문서가 활용되면서 한편으로는 진본성을 확인하기 용기하려면 어떻게 관리되어야 하나 등이다.

현재 공공영역에서 문서의 서식에 관해서는 “행정 효율과 협업 촉진에 관한 규정”과 동 규칙이 기본 원칙을 제시하고 있다. 규정에 따르면 “행정기관에서 장기간에 걸쳐 반복적으로 사용하는 문서로서 정형화할 수 있는 문서는 특별한 사유가 없으면 서식으로 정하여 사용한다.” 그리고 주요 서식은 법령이나 고시 등으로 정하도록 하고 있다. 문제는 규정에서 서식 설계의 일반 원칙을 아날로그적 관점에서 제시하고 있다는 것이다. 시행규칙 [별표 4] 서식의 설계기준을 보면 용지여백, 항목, 표, 선, 칸, 글자 등 인쇄되어 보이는 모양을 중심으로 제시되고 있다. 디지털 서식으로서의 설계 내용은 하나도 언급되어 있지 않다. 이는 종이 공문서의 서식을 외양만 디지털 환경으로 옮겨온 결과라 볼 수 있다. 서식과 관련하여 전자화된 부분은 기안기에서 공문서 본문 서식에서 몇몇 셀에 들어가는 값의 의미를 지정하여 메타데이터화하는 것이다.

서식을 디지털화하는 작업은 금융권을 중심으로 활발하게 이루어졌다. 차세대 정보시스템 구축 사업의 핵심 과제 중 하나가 금융거래의 각종 서식을 eForm으로 만드는 것이다. 은행창구에서 태블릿 단말기에 과거와 동일한 서식을 이미지로 띄워주고, 해당 칸에 전자펜으로 내용을 기입하도록 하는 방식이다. 이용자 입장에서는 종이 대신 단말기, 볼펜 대신 전자펜으로 정보를 기입하도록 하는 변화가 있는 셈이다. 그렇다면 은행의 정보시스템에서는 어떤 변화가 있을까? 각종 서식을 폼(Form)이라는 정보로 저장하고 있다가 출납원이 불러오는 폼을 태블릿에 띄워주도록 하고, 폼의 특정 셀 위치 영역에 적히는 값을 입력으로 처리하고, 저장 시에는 (1) 각 셀 별로 태그정보/값의 쌍들을 하나의 XML 문서로 저장하고 (2) 각 셀에 값이 들어간 폼의 전체 이미지를 PDF로 저장하게 된다.

하나의 트랜잭션에 관한 정보를 두 가지 버전의 포맷으로 저장하는데는 추후의 용도별로 유리한 포맷이 각기 다르기 때문이다. (1)은 정보처리용으로 (2)는 증빙용으로 보관한다. (1)에는 태블릿을 통해 이용자의 눈에 보이는 정보만이 아니라 폼의 ID, 해당 작업을 처리하는 출납원 정보, 이용자 정보, 시스템 날짜 등 기본 정보들이 함께 획득된다. 혹은 폼의 이미지를 스타일 정보로 함께 저장할 수도 있다. (2)의 경우에는 증빙을 위해 디지털서명을 추가하여 진본성 확인이 가능하도록 할 수 있다. 금융권에서 eForm을 사용하는 방식을 통해 현재 공문서 서식을 정보처리의 관점에서도 설계를 하는데 시사점을 얻을 수 있다.

공문서 서식과 관련하여 제안하고자 하는 것은 규칙의 서식 설계 원칙 부분에 정보처리 방식의 기본에 대해 추가해야 한다는 점, 서식을 구성하는 각 셀의 이름과 의미를 기계가독형으로 정의하도록 해야 한다는 점, 기관의 서식에 대해 UUID(Universally Unique Identifier)를 부여하여 관리하도록 해야 한다는 점 등이다. 또한, 저장 포맷과 관련한 제안하고자 하는 것은 데이터 재활용과 분석을 위해 서식의 이미지 부분을 데이터값과 별도로 저장하도록 하고, 해쉬값과 같은 진본성 확인을 위한 메타데이터(Fixity Metadata)를 필수적으로 생성하여 함께 저장하도록 하며, 진본사본의 생성이 가능한 상태를 유지하도록 해야 한다는 것이다.

종이기록 시절에는 서식 자체가 원본성을 추정하는데 중요한 역할을 했다면 디지털기록의 경우에는 콘텐츠 안에 내재된 UUID와 해쉬값을 통해 추정하도록 해야 할 것이다. 공문서의 저장 포맷을

HWP가 아닌 ODT, Open XML 등 빅데이터 활용에 좀 더 유리한 포맷으로 변경하는 것에 대해 고려해야 할 것이다.

4. 콘텐츠 관리체계 들여다보기

공공기관의 콘텐츠 관리가 몇 가지 문제점을 안고 있다. 중복 콘텐츠가 다량 무방비로 생산된다는 점과 콘텐츠 단위의 식별과 통제, 진본성 확인체계가 미비하다는 점이 그것이다. 여기서 콘텐츠란 공문서의 경우 본문파일, 첨부파일 등의 컴퓨터 파일을 의미한다. 이 장에서는 콘텐츠 관리체계에 대해 몇 가지 제안을 하고자 한다.

서울시의 경우, 업무관리시스템에서 문서관리카드를 만들 때 본문은 대부분 새로이 작성하지만 첨부파일에는 이전의 결재문서 본문이나 첨부파일을 재활용하여 업로드하는 경우가 많다. 결재문서 본문의 경우 결재자의 승인 정보가 빠진 문서를 첨부하는 경우가 많은데, 이를 위해 개인 사무용 PC에 한글 문서를 저장해두고 업로드하곤 한다. 접수된 문서는 첨부파일이 한 부만 서버에 저장되고 동일 문서를 접수한 여러 부서에서는 메타데이터만 생성된다. 이렇게 생산된 문서가 기록관리시스템으로 이관되면서 접수문서의 경우, 첨부파일들을 건 단위로 각각 포함하는 방식으로 처리되어 중복 콘텐츠가 급격히 많아지게 된다. 서울시 업무관리시스템 서버에는 약 30%의 파일이 중복저장되고 있고, 서울시 기록관리시스템 서버에는 약 70%의 파일이 중복저장되고 있다.

동일한 콘텐츠가 여러 문서, 여러 기록 건에 중복적으로 활용될 수 있다. 콘텐츠는 한 본문 저장되어도 여러 기록 건에서 활용되는데 문제가 없음에도 이렇게 중복파일을 만들어 관리하게 된 것은 어떤 연유일까? 현재까지 추적 인터뷰를 한 바에 따르면, 전자문서시스템에서도 업무관리시스템에서도 접수문서의 경우 당연히 중복이 많이 발생하므로 여러 접수 기록에서 동일한 콘텐츠에 대해 링크 정보를 보유하는 방식으로 콘텐츠의 중복을 없앴다고 한다. 하지만, 생산 문서의 경우에는 기안자가 직접 첨부 문서를 업로드하기 때문에 중복 여부를 알 수 없었다고 한다.

이 지점에서 콘텐츠 관리의 필요성이 제기된다. 먼저, 이 시점에 용어의 정리가 필요하다. MoReq 2010의 용어로 정리하자면, 기록 건(item), 그리고 건을 구성하는 컴포넌트(component), 그리고 컴포넌트의 실체인 콘텐츠(content) (=컴퓨터파일)에 대해 구분해 지칭해야 한다. 건의 상위레벨은 집합체(Aggregation)이다. 이 때, 건과 컴포넌트는 기록 실체에 대한 메타데이터들이다. 콘텐츠는 디지털 객체 자체를 지칭한다. 건과 컴포넌트는 1:다의 관계이고, 컴포넌트와 콘텐츠는 다:1의 관계이다.

기안자가 업로드하는 문서가 이미 서버에 존재하는 콘텐츠와 동일본일 때는 굳이 중복하여 콘텐츠를 저장할 필요없이 기존의 콘텐츠 링크 정보를 연결만 시켜주면 되는데, 이런 콘텐츠 중복 여부에 대한 확인절차를 갖지 않는다는 점이 문제인 것이다. 콘텐츠의 중복 여부를 신속하고 정확하게 확인

하기 위해서 모든 콘텐츠의 해쉬값을 메타데이터로 관리하고 있어야 하며, 새로 업로드되는 콘텐츠의 해쉬값을 도출하여 중복여부를 확인할 수 있다. 신규로 입수된 콘텐츠에 대해서는 UUID를 부여하여 관리해야 한다.

서울시도 업무관리시스템을 클라우드 기반으로 고도화할 예정인데, 서울시 뿐만 아니라 25개의 자치구 간에도 콘텐츠를 공유할 수 있도록 시스템을 운영할 예정이다. 서울시 본청 내에서 과 단위로, 본청에서 25개 자치구로 수많은 문서 수발신이 이루어지고 있어 여기서 발생하는 중복 콘텐츠의 양이 막대하기 때문이다. 통합 클라우드 저장소를 공동으로 활용하게 된다면 이런 중복을 최소화할 수 있게 된다.

최근 국가기록원이 전자기록의 장기보존과 관련한 정책을 수정 검토 중이다. 국가기록원이 전자기록 장기보존 정책, 전자기록 보존포맷 정책 등에서 밝힌 바에 따르면, 법령에서 사용 중인 문서보존포맷, 장기보존포맷이라는 용어를 각각 '보존포맷', '장기보존 패키지'라는 용어로 전환하겠다고 한다. 그간 두 개의 포맷이 의미하는 것이 무엇인지를 기록학계 내에서 설명하는 것이 어려웠던 점에 비추어 OAIS참조모형에 근거하여 용어를 바꾸는 것은 나쁘지 않은 선택이라 생각한다. 다만, 그간 국가기록원이나 학계에서 장기보존포맷, 즉 패키지에 관해 OAIS참조모형을 오독하여 적용한 부분이 있어 이에 관해 논의를 할 필요가 있다(2019.3.27. 국가기록원 연구세미나의 토론문에서 밝힘).

요약하자면, 패키지는 물리적인 단일 파일로 만드는 것을 의미하는게 아니라 논리적, 개념적인 컨테이너라는 점이며, 패키징 정보과 패키지 기술정보를 정의하는 것까지가 패키지의 설계에 포함되어야 한다는 것이다. 이 때, 패키지에 묶여질 대상들, 즉 정보객체들을 잘 살펴봐야 한다.

물리적 패키징이란 이관이나 매체수록 시 건 단위의 완전성을 기하기 위해 메타데이터와 모든 콘텐츠를 하나의 컨테이너에 묶는 과정이다. 논리적 패키징이란, 데이터베이스와 파일시스템에 메타데이터와 콘텐츠를 흩어두되 물리적 패키징을 할 때 묶여야 할 대상들을 식별을 해두고 관리하는 방식이다. 건 단위로 물리적 패키징을 할 때는 콘텐츠가 중복적으로 저장되지만, 기록관리시스템 안에서 논리적 패키징 상태로 관리할 때는 콘텐츠를 중복 저장해 놓을 필요가 전혀 없다.

기록의 처분과정에서는 처분 대상인 콘텐츠 중에서 삭제 대상인 콘텐츠를 가려내야 한다. 하나의 콘텐츠가 여러 기록 건의 컴포넌트일 수 있기 때문에 더 이상 참조하는 기록 건이 없게 될 때 삭제하도록 해야 한다.