

# XML 지향 학술정보 종합 유통 플랫폼 구현 사례 연구

조성남\*

\*한국과학기술정보연구원

e-mail:chosn@kisti.re.kr

## A Case Study on the Implementation of Academic Information Distribution Platform Based on XML

Sung-Nam Cho\*

\*Korea Institute of Science and Technology Information

### 요약

지식사회로 전환되면서 연구자들은 예전보다 더 많은 학술정보를 생산 및 활용하고 있고, 활용성 및 접근성을 강화하기 위해 학술정보의 전자화를 통해 지식정보를 공유·확산하고 있으며 학술단체에서도 이러한 논문의 수집, 평가, 발간, 출판, 서비스 등 일련의 학술정보 전주기 처리에 많은 시간과 비용을 투자하고 실정이다. 그러나 국내외적으로 논문 투고부터 출판까지 학술정보 종합적인 유통 체계가 부재하기에 각 서비스를 분리하여 별개로 활용하고 있고, 학술정보를 서비스하는 출판사마다 전주기 출판 플랫폼에 대한 정의와 서비스 체계 및 범위가 서로 다르다. 본 연구에서는 국내 학술정보 종합적인 유통 플랫폼에 대한 표준적인 모델을 제시하고 논문 투고부터 웹을 통한 XML 기반의 학술정보 서비스까지 실현한 KPubS 학술정보 유통 플랫폼에 대한 구축 사례를 제시하고자 한다.

### 1. 서론

지식사회로 전환의 가장 큰 요인은 지식을 문자로 표현한 기록물이고 기록물은 보관 및 축적, 변화, 융합되어 더 진화된 지식을 창출하는 근간이 된다. 우리나라도 세계 선도 지식사회로써 GDP대비 R&D 투자비율이 세계 상위권과 동시에 논문 생산 및 특허출원도 세계 10위권에 있다[1][2]. 학자 및 연구자는 지식 표현의 한 수단으로 논문을 통해 지식을 창출하고 이를 증명하고 있으며, 학자 및 연구자가 포함된 학술단체는 이러한 지식 생태계의 한 주체로서 논문을 수집하고 평가, 전자발간 및 출판, 데이터 구축 및 리퍼지토리 저장, 서비스 등 일련의 학술정보 전주기 프로세스를 수행하고 있다. 이에 민간 출판사가 주도가 되어 학술지 출판 지원 시스템 내지는 체계를 운영하고 있고, 출판사마다 자기 출판 플랫폼을 표준이라고 주장하고 있는 상황으로써 신생 학술단체나 영세한 학회는 민간이 운영하는 학술정보 플랫폼을 유료로 사용하기에는 비용도 비싸고 사용하기에는 너무 복잡한 구조를 가지고 있다. 또한 국내에서는 국제 수준의 학술지 출판을 One Stop으로 지원하는 학술지 전문 출판사가 부재하고 학술지 기획, 투고심사시스템, 학술지 편집, 국제적 유통, 국제 색인 등재 등이 산발적으로 수행됨에 따라 많은 우수 학술지들이 고비용과 저작권의 희생을 무릅쓰고 해외 출판사에 의존하려

는 경향이 강하다. 그리고 선도적인 학회나 출판사에서는 단순한 PDF, HTML 등의 서지 중심의 원문 서비스와는 별개로 학술정보의 유통 및 확산과 활용성 강화를 목적으로 XML 기반의 학술정보를 제공하고 있고 XML을 통한 학술정보 서비스는 점점 증가하고 있는 추세이다. 이에 국내에서도 학술정보 종합적인 유통체계 지원 및 XML 기반 웹서비스와 관련한 많은 연구가 진행되고 있다. 본 연구에서는 이와 같은 요구와 필요성에 의해 국내 유일의 XML 기반의 학술정보 종합 유통 지원 체계를 구축한 사례 조사를 통해 국내 학술정보 유통 체계와 전주기 학술출판에 대한 표준을 제시하고 학술정보의 유통 및 확산에 기여하고자 한다.

### 2. 관련 현황 조사

데이터베이스는 데이터를 객체화 및 구조화하여 저장하고 관리하는 데이터 관리 솔루션이다. 데이터베이스는 제조사에 따라 데이터 저장 및 형식이 다르기에 이종 데이터베이스 간에 데이터 교환 내지는 재활용이 어렵다. 타 기종 간의 데이터 이전은 많은 노력과 비용이 소요되는 문제점이 발생한다. 이러한 문제를 해결하기 위한 방안으로 데이터를 관계형에서 HTML Tag 확장 개념의 Stream 형식의 TEXT 포맷 XML(eXtensible Markup Language)으로 구축하는 것이다.

XML은 HTML 형식을 갖추고 있기에 웹에서 활용이 자유롭고, TEXT 형식이기에 별도의 데이터베이스가 필요가 없으며, DTD(Document Type Definition)만 제공되면 데이터베이스 종류에 관계없이 저장이 가능하다. 이러한 장점에 학술분야에서도 JATS(Journal Article Tag Suite) XML을 개발하여 활용하고 있다. 학술지 XML DTD는 미국의 국립의학도서관(National Library of Medicine)에서 제안하여 의생명 분야 학술지 데이터베이스인 PubMed Central에 적용하면서 보편화되었다. 이를 PMC XML DTD라고 하는데, 미국에서는 이것을 JATS라는 이름의 ANSI/NISO 표준으로 제정하였다[2]. 우리나라에서도 의생명분야를 중심으로 XML 원문 구축이 활발하다. 현재 KoreaMed Synapse라는 사이트에서 약 140종의 의생명분야 학술지가 XML로 구축되어 서비스되고 있다.[3] 서태설(2015) 등은 학술지 출판 모델을 해외 대표 사례를 벤치마킹하여 국내 상황에 맞게 설계하였고 [4], 투고부터 인쇄까지의 과정을 창출(Creation) 단계, DB 구축을 아카이빙(Archiving) 단계, 웹서비스를 서비스(Service) 단계로 구분하였고, 마지막에 확산(Promotion) 단계를 추가하고, [표 1]과 같이 모델을 제시하였다.[5]

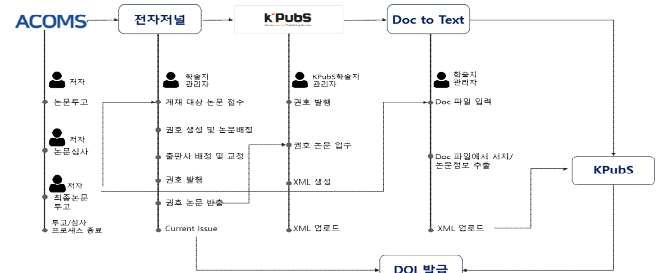
[표 1] Life-cycle Journal Publishing Model

년도	주요 수행 내용
창출 단계 Creation	- 논문투고심사시스템 연계 - 고품질의 논문 메타정보 - 고품질의 text PDF 원문 확보 - 고품질의 그림, 표 확보
아카이빙 단계 Archiving	- 아카이빙을 위한 데이터 구조화 - PDF 논문을 XML로 변환하는 기능 - DOI/DOAJ 기탁 XML 데이터 작성 기능 - 학술지 정보 연계 기능
서비스 단계 Service	- 개별 학술지의 전자저널 구축 - DOI landing page/OA repository 운영 - 서비스 색인 처리 - Pre/Post-print용 원고 제공 - 저작권 및 라이선스 정보 제공
확산 단계 Promotion	- 개별 학술지의 전자저널 구축 - DOI landing page/OA repository 운영 - 서비스 색인 처리 - Pre/Post-print용 원고 제공 - 저작권 및 라이선스 정보 제공

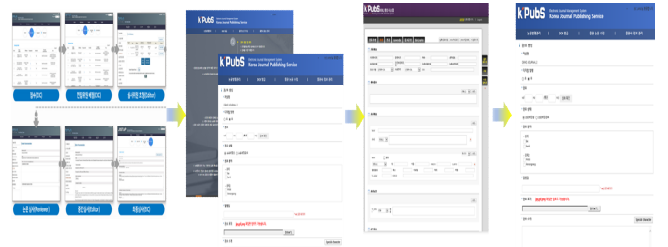
현재 학술정보 출판 시장은 독점적이고 폐쇄적이며 고비용 문제점을 개선하고자 OA(Open Access) 운동이 세계적으로 확산되고 있고 학술출판 정책 및 플랫폼도 OA를 지향하고 있는 추세이다. OA는 모든 사람이 법적·기술적 제약 없이 온라인상에서 무료로 학술정보를 이용하게 하는 국제적 운동으로써 공개하는 방식에 따라 그린 OA와 골드 OA로 나뉜다. OA의 전면 및 일괄 전환과 구독료 철폐를 위해 국제적 프로젝트인 OA2020과 SCOAP3가 구성되어 운영되고 있다.[6][7] 학술정보 종합 유통 플랫폼도 OA 정책을 반영해야만 한다.

### 3. XML 지향 학술지 유통 플랫폼 구현

서론에서도 언급했듯이 학술정보 출판 관련 시스템은 논문 투고심사시스템, 전자발간 및 출판 시스템, 아카이빙 서비스, Repository 구축, XML 구축, 검색 시스템 등 다양하고 각 단위 시스템이 단독 또는 2-3개 시스템이 조합되어 서비스되고 있다. 즉 생산에서 구축, 출판, 서비스까지의 전체 프로세스를 OneStop으로 처리할 수 있는 시스템이 부재함에 따라 이를 해결하기 위해 앞에서 언급한 서태설 등(2015)의 모델을 참조하여 전주기 학술정보 플랫폼을 구현하였다. [그림 1]은 시스템 간 업무 워크플로우를 표현하였으며, 시스템 구성은 [그림 2]와 같이 Peer Review를 담당하는 ACOMS 시스템, 전자발간을 담당하는 ACOMS 내 전자발간시스템, XML 생성은 KPubS XML 생성 시스템, 그리고 웹서비스는 XML 기반의 KPubS 시스템 등으로 구성되어 있다.



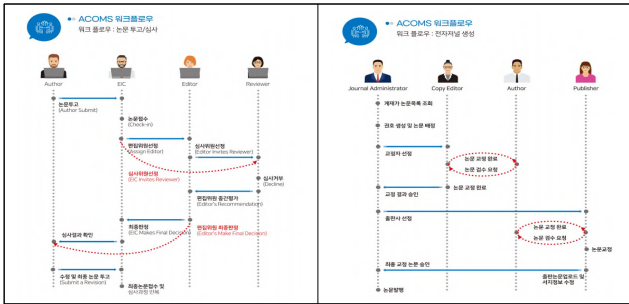
[그림 1] 학술정보 유통 플랫폼 내 서브 시스템 간 업무 흐름도



[그림 2] 학술정보 유통 플랫폼 구성도

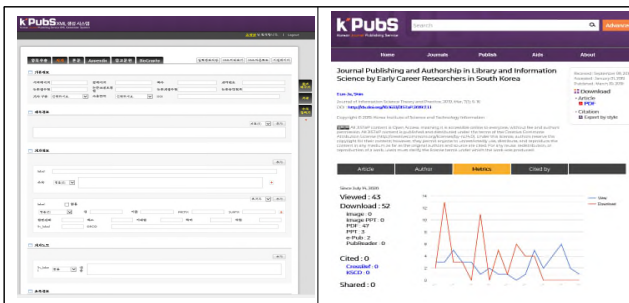
학술정보 유통 플랫폼의 첫 번째 구성 요소 ACOMS는 학회 정보화 및 학회 업무 지원과 학술정보 공동 활용 및 확산을 목적으로 KISTI가 학회에 보급하는 논문 투고심사관리 시스템이다. 초기에 Clarivate 사의 ScholarOne을 벤치마킹하여 개발하였고 현재는 4.5 버전에 이르고 있으며, 논문투고에서부터 에디터 배정, 심사위원 초청, 논문 심사, 중간 판정 그리고 최종 판정에 이르기까지 온라인으로 논문 투고심사 업무 수행이 가능하다. 두 번째 요소인 전자발간시스템은 ACOMS에서 최종 게재 승인된 논문 정보를 입력 값으로 받아 권호를 생성하고 해당 논문을 권호에 배정한다. 배정된 논문은 교정을 거친 뒤 자동적으로 전자 발간이 되고 ACOMS

내 저널 홈페이지의 Current Issue 메뉴에 최신 정보로 서비스가 된다. 전자저널의 장점은 게재 승인된 논문이 학술지로 발간되기까지의 시간이 많이 소요됨에 따라 최신성이 떨어지는 단점이 있으나 전자저널은 시간에 구애없이 즉시 서비스가 되므로 연구의 최신성이 뛰어나다. [그림 3]은 ACOMS 및 전자발간시스템 업무 흐름도를 보여주고 있다.



[그림 3] ACOMS 및 전자발간시스템 업무 흐름도

세 번째 요소인 XML 생성 시스템은 출판된 PDF를 JATS XML로 자동 변환해주는 시스템으로써 아직은 그림 및 표, 참고문헌 등은 일부 수작업이 필요한 단점은 있으나 전문 XML 업체에 의뢰하는 경우 논문 1편당 5만원~10만원 정도의 지출 비용을 절약할 수 있는 장점이 있다. XML 변환에 대한 수작업을 최소화하기 위해 DOC 파일형식의 원문에 대하여 XML으로 자동 변환해주는 doc2xml 기능을 포함하였다. 국제 저널 대부분 DOC 파일형식을 사용하나 우리나라는 여전히 Hwp 문서를 주로 사용하기에 Hwp 원문에 대한 XML 자동 변환 기능 개발이 필요하다. 외부에서 구축된 XML 문서도 시스템 내 재활용이 가능하다. 마지막 네 번째 요소는 구축된 XML 원문을 웹에서 서비스하는 KPubS 시스템이다. 원문 자체가 XML로 구축되었기에 변환 및 표현이 자유롭고 타 시스템 이전이 용이하다. [그림 4]는 구현된 XML 생성 시스템 및 KPubS 웹서비스 화면을 캡처한 이미지이다.



[그림 4] XML 생성 시스템 및 KPubS 웹서비스 화면

지금까지 논문을 생산하는 투고심사 시스템에서부터 전자발간시스템, XML 구축 및 변환 시스템, XML 기반 학술정보 웹서비스까지 가능한 KISTI의 학술정보 종합 유통 플랫폼

구축 사례를 살펴보았다. 학술정보의 유통 및 학술 출판 플랫폼에 대한 구성 및 범위, 역할 등은 연구자마다 다를 수 있다.

본 연구에서는 하나의 투고된 논문이 투고심사시스템을 통해 학술정보로 생성되고, 생성된 논문은 자동으로 전자 발간되며, 전자 발간된 논문은 XML 데이터로 자동 변환되어 웹을 통해 XML 형태로 서비스되는 일련의 과정을 시스템으로 구현한 사례를 제시하였다.

#### 4. 결론

지식정보화 사회로의 전환과 전자정보의 공유 및 확산으로 학술정보의 활용성은 강화되고 있으나 학술정보의 근간인 논문의 수집, 평가, 발간, 출판, 서비스 등 일련의 학술정보 전주기 유통에 많은 시간과 비용이 소요되고 있다. 이를 해결하기 위한 방안으로 학술정보의 종합적인 유통 플랫폼 구축을 통해 논문 투고심사부터 전자발간, 출판, XML 생성 및 구축, XML 전문 웹서비스까지 업무를 윈스탑으로 수행할 수 있는 기반을 마련하고 학술 업무의 효율적 지원과 학술정보 유통 및 확산에 기여가 예상된다. 또한 학술단체에서 본 플랫폼을 이용할 시 논문 투고심사시스템 및 전자발간시스템에 대한 연간 사용료, XML 구축비용과 XML 원문 웹호스팅에 대한 연간 사용료 등 수백에서 수천만원을 절약하는 효과가 있다.

향후 연구로는 학술정보의 화두인 OA에 맞는 OA 학술정보 전주기 출판 모델을 개발하고자 하며, 국내에서 가장 많이 사용하고 있는 아래한글 hwp 문서를 XML 문서로 자동변환 기능을 연구하고자 한다.

“본 연구는 (2020년도) 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행한 것입니다.”

#### 참고문헌

- [1] OECD, "Main Science and Technology Indicators", OECD Science, Technology and R&D Statistics, 2012.
- [2] WIPO, "2010 World Intellectual Property Indicator", 2010.
- [3] <http://synapse.koreamed.org>.
- [4] 국내정보실, "학술지국제출판가이드", 한국과학기술정보연구원, 2014.
- [5] Tae-Sul Seo and Seon-Heui Choi, "Data Models for Visualization Service of Scholarly Journal and Article Information", ICC 2014, Jeju, June 2014.
- [6] 이덕환, "논문 무료로 공개하라", 동아일보, 9월, 2016년.
- [7] 신은자, "오픈엑세스 확산을 위한 APC 지원 정책에 관한 연구", 정보관리학회지, 제 31권 3호, pp. 249-27, 2014년.