



「 리포지터리 사례분석 및 시사점 도출 」

이상기 · 정영미



## ■ 목 차 ■

### 요 약

1. 서론 .....	1
2. 리포지터리 개념 및 성공요인 .....	3
3. 해외 리포지터리 현황 및 주요사례 조사 .....	9
4. 사례분석 및 시사점 도출 .....	30
5. 결론 .....	38
<참고문헌> .....	39



## 《 요 약 》

- 새로운 학술정보 유통 패러다임인 오픈액세스 운동이 활성화되면서, 이를 실천하기 위한 모델로 기관에서 생산된 학술정보를 체계적으로 아카이빙하여 장기보존·확산하는 리포지터리 구축이 급격하게 증가하고 있음.
  - 리포지터리 유형에는 기관차원에서 학술정보를 수집하여 공유하는 기관 리포지터리, 여러 기관이 공동으로 특정 주제에 관련된 정보를 집중적으로 구축하는 주제 리포지터리, 국가 차원의 아카이빙센터인 국가리포지터리가 있음.
  
- 최근 리포지터리 구축이 전 세계로 확산되고 있지만, 국내의 경우 리포지터리에 대한 인식이 매우 부족하고 정부지원도 미흡함.
  - 국내에는 5개 기관에서 리포지터리를 구축·운영하고 있으며, 공신력 있는 국제기구에 등록된 리포지터리는 KAIST KOASAS(KAIST Open Access Self-Archiving System)외 2개임.
  
- 본 리포트는 국내의 리포지터리 운영 및 활성화 방안을 모색하기 위해, 해외 주요 리포지터리를 조사·분석하였으며 조사결과 및 시사점은 다음과 같음.
  - 리포지터리 운영 및 지원정책
    - 리포지터리의 실질적인 운영주체는 대학이나 연구소가 대부분이며, 운영주체와 관리주체가 분리되어 있는 곳이 많음.
    - 주제 리포지터리의 경우 커뮤니티에 자율성을 부여하기 위해 운영주체에 대한 권한을 최소화하고 있으며, 기관 리포지터리는 소속 구성원들의 참여 활성화를 위해 제출 의무화 등 보다 강력한 운영정책을 추구함.

#### - 콘텐츠 수집 및 관리정책

- 연구자의 직접 제출뿐만 아니라 다양한 방법으로 콘텐츠를 수집함.
- 연구자가 직접 제출하는 경우 자격요건에 특별한 제한은 없으며 그 방법이나 절차가 매우 단순함.
- 콘텐츠에 대한 품질관리 역시 규제를 최소화하고 언어 제한도 없음.
- 콘텐츠 포맷은 리포지터리마다 다양하지만 PDF 파일을 선호함.
- 효율적으로 학술정보를 관리하기 위해 고유 식별자 또는 URL을 부여함.
- 동료심사 과정이 없으며 다만 최소한의 품질관리를 위해 등록된 회원만 콘텐츠를 제출할 수 있음.
- 제출된 콘텐츠는 제출자가 요청하지 않는 한 운영주체나 다른 연구자들이 삭제하거나 수정할 수 없음.

#### - 콘텐츠 배포 및 이용 활성화 정책

- 광범위한 콘텐츠 배포 및 이용 활성화를 위해 이용자에게 특별한 제한을 두지 않음.
- 신속하게 정보를 제공하고 지속적으로 관심을 유도하기 위해 RSS, 알리미 서비스, 이용통계 서비스, 인용정보 분석서비스, 다양한 검색 및 인터페이스를 제공함.
- 홈페이지와 하위 페이지를 자국어와 영어로 제작하여 병행 서비스하고 있음.

#### - 장기보존 정책

- 대부분의 리포지터리에서 장기보존을 정책적, 기술적으로 지원함.
- 명확하고 강력한 장기보존 정책은 연구자들에게 신뢰를 주기 때문에 구축된 콘텐츠를 운영주체가 지속적으로 유지 및 장기보존한다고 명시하고 있음.

#### - 저작권 정책

- 유통시 저작권 문제가 없는 출판 전 논문이나 출판중인 논문 위주 수집
- 저작권이 이미 출판사에 있는 경우 출판사에서 승인을 얻은 후 해당 콘텐츠를 아카이빙하고 있음.

## 1. 서론

- 새로운 학술정보 유통 패러다임인 오픈액세스에 대한 관심이 고조되고 있으며, 학술정보의 생산과 유통에 변화의 바람이 불고 있음.
  - 전통적으로 학술정보 생산은 출판사와 배급자가 담당하고, 도서관과 정보 서비스 기관들이 정보유통을 담당하는 분화된 구조였음.
  - 상업 출판사에 의한 저작권 독점 및 전자저널의 급격한 가격상승은 학술정보 유통 위기를 초래하였으며 새로운 변화를 요구함.
  
- 전 세계적으로 오픈액세스 기반 리포지터리 구축이 매우 활발하며, 특히 2006년 이후 기관 리포지터리 구축이 급격하게 증가하고 있음.
  - 많은 대학이나 연구소에서 학술정보를 수집·장기보존·확산하기 위해 리포지터리를 구축하고 있으며 향후 이런 움직임은 더욱 가속화될 것임.
    - 전 세계 리포지터리는 1,650개로 매일 평균 1개의 신규 리포지터리가 등록되고 있음(ROAR, <http://roar.eprints.org>, 2010.8).
  - 해외 주요 리포지터리들은 정부의 지원과 연구자들의 참여로 다량의 학술정보를 구축하여 공유하고 있으며, 활발히 이용됨에 따라 그 규모가 더욱 확대되고 있음.
    - 전 세계적으로 리포지터리를 통해 2,300백만건 이상의 학술정보를 수집하여 공유하고 있음(OCLC, <http://www.oclc.org/oaister>, 2010.8).
  
- 국내 리포지터리의 경우 초기단계로 리포지터리에 대한 인식도 부족하고 정부지원도 미흡함.
  - 국내에 총 5개 기관(KAIST, 서울대, 인하대, 지질자원연구원, 대외경제정책연구원)에서 리포지터리를 구축·운영하고 있으며, 공식적인 기구에 등록된 리포지터리는 KAIST의 KOASAS(KAIST Open Access Self Archiving System), 서울대의 S-Space, 대외경제정책연구원의 Korea Institute

for International Economic Policy임.

- 전 세계 리포지터리 웹 순위(Ranking Web of World Repositories, <http://repositories.webometrics.info>)의 상위 400위에 국내 기관은 한 곳도 포함되지 않음. 그러나 일본은 25개 리포지터리가 포함됨.
- 본 리포트는 국내 리포지터리 운영 및 활성화 방안을 모색하기 위해 해외 주요 리포지터리를 중심으로 운영·지원·수집·관리·배포·이용활성화·장기보존·저작권 정책을 집중적으로 조사 분석하여 시사점을 도출함.

## 2. 리포지터리 개요 및 성공요인

### 1) 리포지터리 정의

- 리포지터리란 용어는 물건을 넣어들 수 있는 용기나 방이라는 뜻의 라틴어인 repositorium에서 유래하였으며, 물건들이 모여지는 장소를 의미함(위키백과사전).
- 리포지터리는 오픈엑세스를 실현하는 양대 산맥 중 하나로, 오픈엑세스 기반 디지털 학술정보를 수집·관리·배포·장기보존하는 개방적 디지털 도서관임.
  - Lynch(2003)는 리포지터리를 대학이나 기관에서 구성원들이 생산한 디지털 콘텐츠를 관리·편목·접근·배포하고 장기보존하는 것으로 정의함.
  - SPARC 보고서에서 Crow(2002)는 리포지터리를 하나 또는 다수의 대학이나 기관에서 생산된 학술정보를 수집하여 장기보존하는 디지털 서고라고 정의함.
  - 위키 백과사전은 대학이나 연구기관의 구성원들이 생산한 학술정보를 공공의 시각에서 수집하고, 축적하여 학내외의 연구자들에게 무료로 제공하는 인터넷상의 전자서고 또는 개방형으로 운영되는 디지털 객체(digital objects)로 정의함.

### 2) 리포지터리 특징

- 리포지터리는 기관에서 생산된 학술정보를 체계적으로 아카이빙하여 오픈 액세스 기반의 표준화된 방식으로 콘텐츠를 상호공유하기 위한 솔루션으로 기존 디지털도서관나 지식관리시스템과 다음과 같은 차이가 있음.
  - 리포지터리에 등록된 학술정보는 누구나 무료로 자유롭게 이용하고 유통할 수 있는 디지털 형태의 오픈엑세스 콘텐츠이어야 함.
  - 등록된 모든 콘텐츠는 국제표준 라이선스인 CCL(Creative Commons License)로 저작권 사항과 이용범위를 명확하게 표기하고 있음.

- 메타데이터를 수집(harvest)하여 공유할 수 있도록 개방형 표준 프로토콜인 OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)를 지원함.
- 기관 리포지터리를 외부에 적극 홍보하고 구축된 콘텐츠를 글로벌 확산하기 위해 공신력 있는 국제기구(OpenDOAR, ROAR 등)에 등록하여 학술정보를 공유함.
- 기관에서 생산된 학술정보의 수집에서 확산까지 전주기 프로세스를 국제표준에 입각하여 개발한 토털 솔루션임.
- 디지털도서관 등의 메타데이터는 전문적인 교육을 받은 사서가 작성하지만 기관 리포지터리의 메타데이터는 대부분 셀프 아카이빙에 의해 메타데이터 교육을 받지 않은 연구자가 입력함.

※ CCL(Creative Commons License)은 자신의 창작물에 대하여 일정한 조건하에 모든 이의 자유이용을 허락하는 내용의 라이선스(License)임. 즉, 저작권이 원칙적으로 다른 이의 이용을 금지하되 개별적인 계약에 의해 이용을 허락하는 형태지만, CCL은 이와 달리 원칙적으로 모든 이의 자유이용을 허용하되 몇 가지 이용방법 및 조건을 부가하는 개방적인 이용허락임(CCK, <http://www.creativecommons.or.kr/>, 2010.9).

### 3) 리포지터리 개발동향

- 리포지터리를 구축하는 기관들이 급격하게 증가하고 있으며 초창기 미국, 영국 중심에서 독일, 호주, 일본 등 전 세계로 확산되고 있음.
  - 리포지터리를 등록·확산하는 국제기구인 OpenDOAR 기준 현재 1,650개의(2010.8) 리포지터리가 구축되어 약 2,300백만건의 학술정보를 공유하고 있음. 특히, 2006년 이후 급격하게 리포지터리를 구축하기 시작하였으며 최근에는 하루 평균 1개 신규 리포지터리가 등록되고 있음.
  - 국가별로는 미국 370개(23%), 영국 169개(11%), 독일 139개(9%), 일본 77개(5%), 호주 63개(4%)순임.

- 리포지터리에 등록된 콘텐츠 유형은 매우 다양하며 대부분 학술정보임.
  - 학술논문 986개(62%), 학위논문 801개(50%), 보고서 650개(40%), 콘퍼런스 559개(35%), 전자책 500개(31%), 멀티미디어 콘텐츠 382개(24%), 교육용 콘텐츠 246개(15%), 데이터 셋 74개(4%). 특허 31개(1%), 소프트웨어 31개(1%), 기타 274개(17%) 등
- 리포지터리 구축 솔루션에는 오픈소스 무료 솔루션과 상업적인 유료 솔루션이 있으며 무료 솔루션을 훨씬 선호함.
  - 무료 솔루션인 미국 MIT와 HP가 공동개발한 Dspace(513개, 32%)와 영국 사우스햄튼 대학에서 개발한 EPrints(266개, 17%)를 가장 많이 사용함.
  - 유료 솔루션으로 상업출판사인 ProQuest에서 개발한 Digital Commons 73개(5%)와 디지털도서관 구축 업체인 ExLibris의 DigiTool이 있음.
  - 무료 솔루션은 오픈소스로 구축비용이 매우 저렴하고 프로그램 소스를 전부 공개하기 때문에 커스터마이징하기 용이한 이점이 있으며, 유료 솔루션은 개발업체들의 자사 제품과 호환성이 뛰어난 장점이 있음.

#### 4) 리포지터리 성공요인 및 장애요인

- 황혜경(2009)은 조직적, 정책적, 기술적 측면에서 리포지터리 성공요인을 다음과 같이 제시함.
  - 조직적 요인: 기관장의 적극적 의지
  - 정책적 요인: 핵심콘텐츠 선정, 품질통제 수준의 최소화, 저작권 보호, 적극적 홍보를 통한 옹호집단 구성, 장기보존 약속, 인센티브, 의무조항
  - 기술적 요인: 제출방식의 편이성, 상호운용성 지원
- 스페인 국가연구위원회(CSIC) 소속 사이버메트릭스연구실에서는 전 세계 리포지터리 순위를 제공하는 Ranking Web of World Repositories를 운영하고 있음. 이 연구실에서는 리포지터리를 성공적으로 운영하기 위한 십계명을 <표 1>과 같이 제시함.

〈표 1〉 성공적인 리포지터리 운영을 위한 십계명

구분	내용
URL 명명	<ul style="list-style-type: none"> <li>- 기관 도메인명을 변경하지 말고 유지하라. 이것은 리포지터리 가시성(visibility)에 영향을 미친다.</li> <li>- 도메인명으로 잘 알려진 두문자어를 사용하라.</li> </ul>
콘텐츠 생성	<ul style="list-style-type: none"> <li>- 많은 연구자 그룹을 확보하라. 가장 좋은 방법은 미래의 저자들이 될 연구자 또는 대학원생을 허용하라.</li> <li>- 분산시스템은 다양한 수준에서 운영될 수 있다. (중앙조직, 도서관이나 데이터센터, 개인이나 팀)</li> <li>- 외부 콘텐츠에 대한 호스팅은 가시성을 증가시킬 수 있다.</li> </ul>
콘텐츠 변환	<ul style="list-style-type: none"> <li>- 비 전자포맷의 중요 콘텐츠들은 웹페이지로 쉽게 변환될 수 있다.</li> </ul>
인터링킹	<ul style="list-style-type: none"> <li>- 사이즈가 부족하거나 고품질의 콘텐츠가 아니라면 다른 사이트로부터 링크가 거의 없을 것이다.</li> <li>- 다른 사이트의 링크들을 평가하고 분류하라. 그러면 다른 사이트(파트너, 포털, 유사한 웹디렉토리, 동료나 파트너의 개인적인 페이지)에서도 링크할 것이다.</li> </ul>
언어(특히 영어)	<ul style="list-style-type: none"> <li>- 메인페이지 뿐만 아니라 세부영역까지 영어로 표현해야 한다.</li> </ul>
미디어 파일	<ul style="list-style-type: none"> <li>- html이 웹페이지 표준 포맷이지만 때로는 PDF나 doc와 같은 파일포맷이 더 좋다.</li> <li>- PostScript가 다양한 주제영역에서 인기 있는 포맷이지만 열어보기가 어렵기 때문에 대안적으로 PDF 포맷이 필요하다.</li> </ul>
검색엔진의 친화적 설계	<ul style="list-style-type: none"> <li>- Flash, Java, JavaScript는 웹로봇의 접근을 방해한다. 또한 깊은 디렉터리나 복잡한 인터링킹도 사용하지 않는 것이 좋다.</li> <li>- 검색엔진은 데이터베이스나 매우 동적인 페이지를 검색할 수 없다.</li> </ul>
인기도와 통계	<ul style="list-style-type: none"> <li>- 방문자 수가 중요하다(방문동기, 출신, 분포를 모니터링 하는 것이 중요).</li> <li>- 로그분석이 인구통계학적 분석과 지리적 데이터를 제공할 것이다.</li> <li>- 검색엔진을 통해 접근했다면 사용한 검색어나 구, 참조정보를 보여주는 옵션이 있어야 한다.</li> </ul>
장기보존과 지속성	<ul style="list-style-type: none"> <li>- 사이트에서 오래된 콘텐츠들의 복사본을 유지하는 것은 필수적이다.</li> <li>- 때때로 재설계하거나 갱신하면서 손실될 수 있고 이렇게 손실된 것은 복구하기 어렵다.</li> </ul>
표준	<ul style="list-style-type: none"> <li>- 의미 있는 타이틀과 기술적인 메타태그는 페이지의 가시성을 높여준다.</li> <li>- 더블린 코어(Dublin Core)와 같은 표준은 저자정보, 키워드, 웹사이트의 다른 데이터를 사용할 수 있게 한다.</li> </ul>

- Johnson(2002)은 기관 리포지터리 운영주체와 환경이 다양하지만 기관 리포지터리 향상을 위해 다음과 같이 네 가지 요소를 제시함.
- 기관 정의: 기관 리포지터리의 목적과 범위가 명확해야 함.
  - 콘텐츠 범위: 품질관리를 위해 학술적인 가치가 있는 정보들로 제한하여 구축해야 함.
  - 수집 및 장기보존: 수집정책 및 제출기준을 명시해야 함.
  - 상호운용성: OAI 국제표준을 준수하여 타 시스템과 상호운용성을 지원해야 함.
- Davis와 Connolly(2007)는 코넬대학에서 리포지터리를 사용하는 이유로 장기보존성, 적시성, 등록 이점을 들고 있으며 리포지터리를 사용하지 않는 이유로 훈련의 필요성, 저작권, 품질, 표절, 정확성을 언급함.
- 리포지터리를 사용하는 이유
    - 장기보존성: 기관은 장기적 사용을 위해 문서를 마이그레이션하여 장기 보존할 의무가 있음.
    - 에이전시나 출판사에게 양도하는 정책: 연구자는 자동적으로 저작물이 제출되길 희망함.
    - 적시성: 출판전 보거나 온라인 보기를 통해 논문이 공식적으로 출판되기까지 지연되는 시간을 줄일 수 있음.
    - 등록 이점: arXiv에서 보듯이 논문이 공식적으로 출판되기 전에 등록할 수 있으므로 연구에 대한 선취권을 획득할 수 있음.
  - 리포지터리를 사용하지 않는 이유
    - 훈련의 필요성: 추가적인 가치 없이 새로운 시스템을 사용하거나 이를 위해 필요한 기술을 익힐 필요가 없음.
    - 저작권 문제: 출판사가 리포지터리에 아카이빙을 허용 할 지 불확실함.
    - 질적 보장: 학술적 가치가 높은 자료들은 여전히 출판사의 저널에서 입수해야 함.
    - 표절에 대한 두려움: 인문학 분야의 연구자들은 적시성이 그다지 중요하지 않고, 오히려 연구주제를 초기에 공개함으로써 다른 연구자들에 게

그 연구영역을 빼앗길 위험이 있음.

- 정확성 문제: 리포지터리 저작물은 동료심사 등을 거치지 않기 때문에 정확성 문제가 있을 수 있으며, 평판이 떨어지는 자료가 포함될 수 있음.

□ Westrienen과 Lynch(2005)는 리포지터리를 사용하지 않는 이유를 다음과 같이 정리함.

- 참여할 필요성을 납득할 수 없거나 지적재산권 문제에 대한 혼란과 불확실성
- 리포지터리 콘텐츠가 어떻게 사용될 것인지에 대한 신뢰도가 낮거나 오픈 액세스로 제공되는 콘텐츠는 수준이 낮다는 인식
- 논문 제출 의무화에 대한 정책 결여

### 3. 해외 리포지터리 현황 및 주요사례 조사

#### 1) 해외 리포지터리 현황

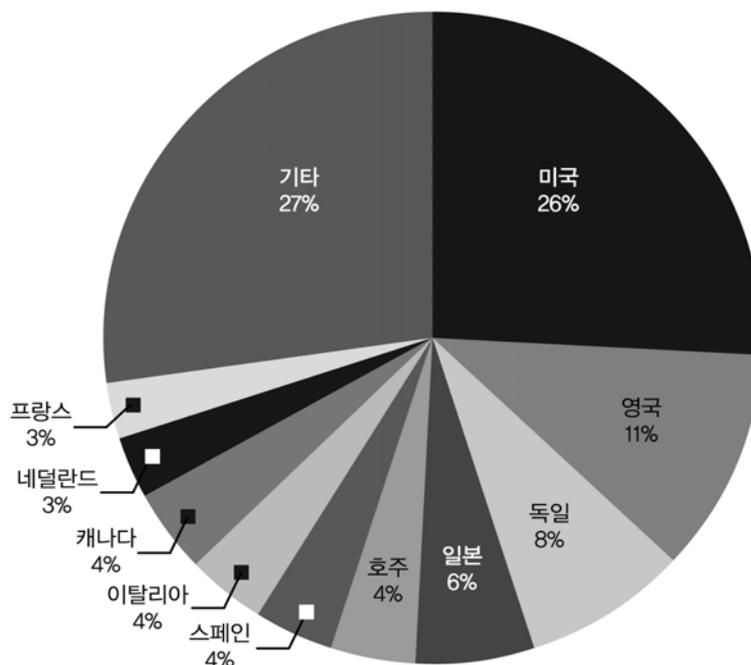
- 해외 리포지터리 현황을 분석하기 위해 2010년 1월 스페인 국가연구위원회 (CSIC) 소속 사이버메트릭스연구실에서 발표한 세계 리포지터리 웹 순위 (Ranking Web of World Repositories)를 이용함. <그림 1>은 Ranking Web of World Repositories에서 제공한 상위 400개 리포지터리이며, 미국에서 운영하는 리포지터리들이 최상위에 랭크됨.

<그림 1> Ranking Web of World Repositories의 Top 1~18

WORLD RANK	REPOSITORY	COUNTRY	POSITION			
			SIZE	VISIBILITY	RICH FILES	SCHOLAR
1	Arxiv.org e-Print Archive	USA	4	2	2	3
2	CiteSeerX	USA	1	1	411	2
3	Scientific and Technical Information Network	USA	7	7	1	4
4	Social Science Research Network	USA	6	3	51	6
5	Research Papers in Economics	USA	2	6	140	5
6	Smithsonian/NASA Astrophysics Data System*	USA	3	4	539	1
7	Hal CNRS	France	18	9	5	8
8	Institut National de Recherche en Informatique et en Automatique Archive Ouverte	France	19	11	26	14
9	CERN Document Server	Switzerland	11	16	40	11
10	MIT Dspace	USA	14	34	7	9
11	École Polytechnique Federale de Lausanne Infoscience	Switzerland	8	14	49	95
12	Agecon Search Research in Agricultural and Applied Economics	USA	21	45	4	17
13	University of Michigan Deep Blue	USA	27	28	13	20
14	Hyper Article en Ligne Sciences de l'Homme et de la Société	France	83	12	122	30
15	Igitur Archive Universiteit Utrecht	Netherlands	26	60	3	28
16	Munich Personal Repec Archive	Germany	47	30	14	22
17	Universität Stuttgartelektronische Hochschulschriften	Germany	35	13	37	241
18	University of Southampton: Department of Electronics and Computer Science	UK	63	20	32	67

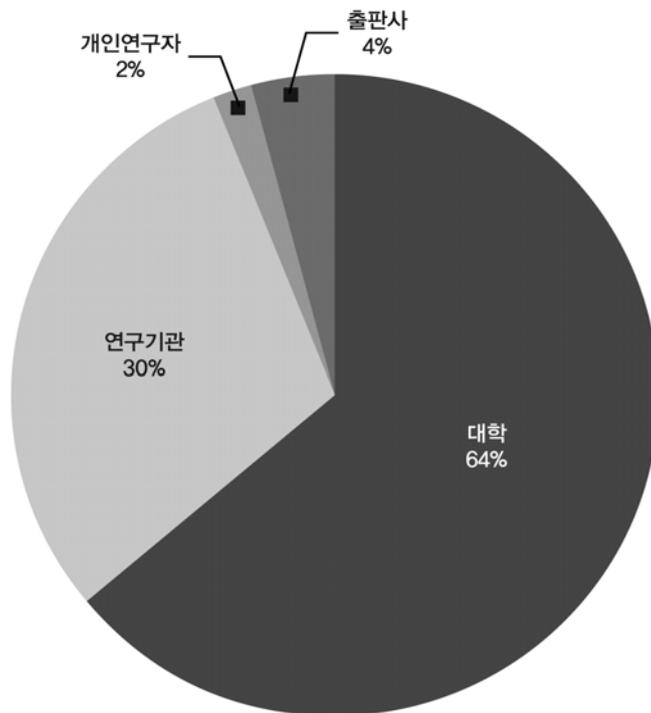
- 스페인 사이버메트릭스연구실에서는 리포지터리를 크기, 가시성, 리치 파일, 구글 스칼라 문헌수 측면에서 평가하여 매년 순위를 발표하고 있으며 세부 내용 및 가중치는 다음과 같음(2010.4).
- 크기(size, S): 세계적으로 가장 큰 4개 검색엔진(Google, Yahoo, Live Search, Exalead)에 나타난 페이지 수
  - 가시성(visibility, V): Yahoo와 Exalead로부터 확보한 사이트의 외부링크 총 수
  - 리치파일(rich files, R): Google과 Yahoo로부터 수집한 PDF 파일 수
  - 구글 스칼라 문헌수(scholar, Sc): Google Scholar를 이용하여 2001년에서 2008년 사이에 출판된 최근 논문과 총 논문수의 정규화된 평균을 계산
  - 평가지표에 대한 가중치
    - 크기(20%) + 가시성(50%) + 리치파일(15%) + 구글 스칼라(15%)
- 상위 400개 리포지터리의 국가별 분포는 미국, 영국, 독일, 일본 순임. 미국이 101개(26%)로 가장 많고 그 다음 영국 45개(11%), 독일 32개(8%), 일본 25개(6%) 순임. 대부분 북미와 유럽 영어권 국가들로 아시아권은 일본(25개), 중국(2개)이 포함되며 우리나라는 하나도 없음.

〈그림 2〉 Top 400 리포지터리 국가별 분포



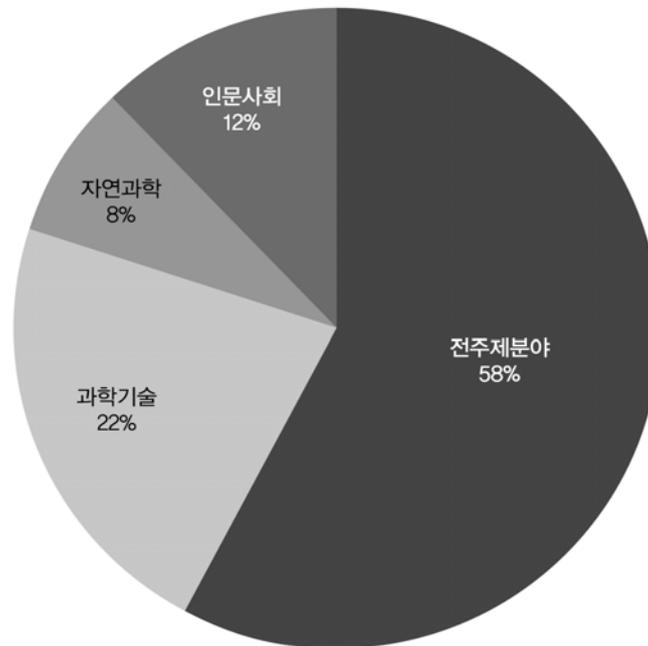
- 상위 50위내 리포지터리 운영주체는 대학과 연구소로, 구축된 콘텐츠는 특정주제 분야보다는 전 주제 분야가 대부분임. 상위 리포지터리들은 언어장벽을 고려하여 자국어뿐만 아니라 영어를 병행 서비스하고 있음.
- 운영주체별 분포현황은 <그림 3>과 같이 대학이 32개(64%)로 가장 많고, 연구소 15개(30%), 출판사 2개(4%), 개인연구자 1개(2%)순임.

<그림 3> Top 50 리포지터리 운영주체별 분포



- 주제별 분포를 살펴보면 <그림 4>와 같이 전 주제 분야를 포괄하는 리포지터리가 58%로 가장 많고 과학기술분야 22%, 인문사회분야 12%, 의학, 농학, 자연과학 분야가 8%를 차지하는 것으로 나타남. 대학에서 운영하는 리포지터리는 대부분 전 주제 분야를 다루고 있음.

〈그림 4〉 Top 50 리포지터리 주제별 분포



- 상위 50위내 리포지터리는 홈페이지와 세부 페이지까지 모두 자국어와 영어를 병행 서비스함으로써 언어장벽 문제를 해결하고 있음.

## 2) 주요사례 조사

- 해외 리포지터리를 심층조사 분석하여 시사점을 도출하기 위해 상위 10위 내에 있는 Arxiv.org e-Print Archive 등 7개 주요 리포지터리를 선정함.
- 심층조사 분석한 7개 주요 리포지터리의 개발배경, 운영정책, 콘텐츠 수집 및 관리정책, 서비스 특징 등을 요약하면 다음과 같음.

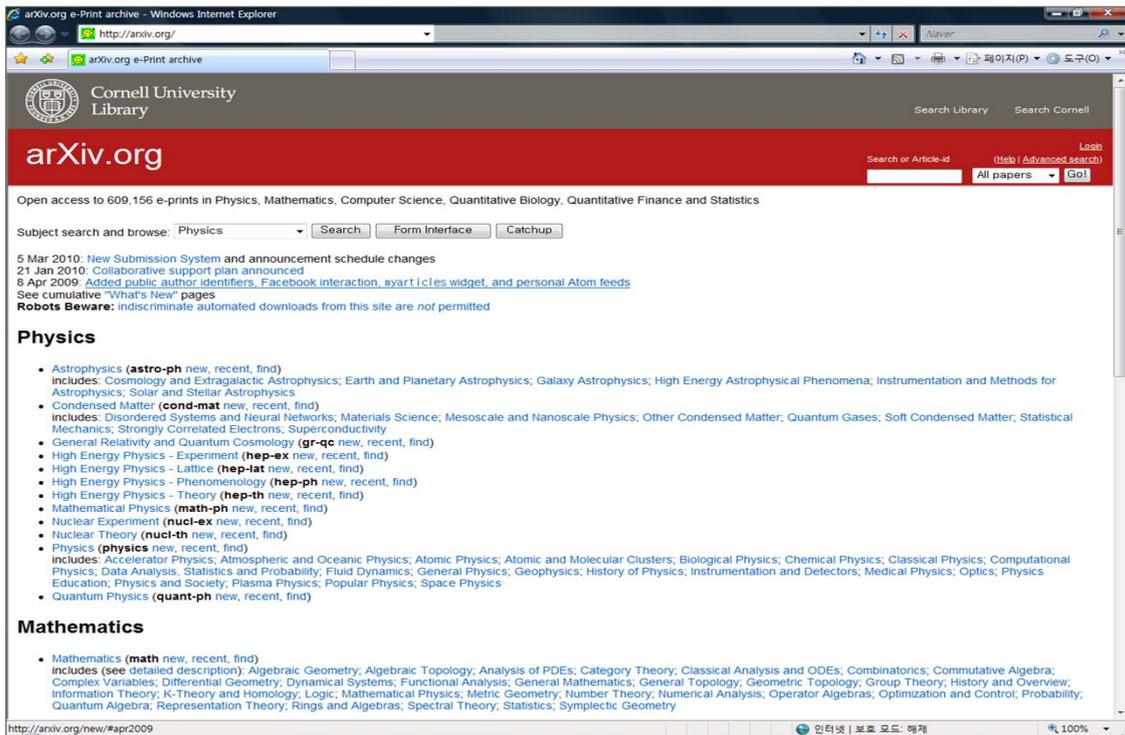
### 가. Arxiv.org e-Print Archive

#### ① 개요

- 코넬대학도서관(Cornell University Library)에서 운영하고 있는 arXiv는 학술지 발표 이전의 논문을 공유하기 시작한 전자서고이자 배포서버로 1991년 8월부터 서비스를 시작함. 국제적인 주제 리포지터리로 물리학, 수학, 컴퓨터공학, 미생형과학, 계량생물학, 통계학 분야의 출판전 논문(preprint)을 아카이빙하고 있음.

- 전통적으로 물리학 분야 연구자들은 논문을 투고하기 이전에 연구결과물을 동료들과 돌려보는 전통이 있었으며, 연구성과에 대한 우선권 확보 및 과학 기술정보의 생명인 속보성을 위해 출판전 논문(preprint)을 arXiv에 등록하기 시작함.

〈그림 5〉 arXiv 리포지터리



## ② 콘텐츠 수집 및 관리정책

- arXiv는 최소한의 품질을 관리하기 위해 제출된 논문에 대한 인증절차가 있으며, 인증은 논문을 심사하는 것은 아니고 해당 분야에 적합한지 개략적으로 검토하는 것임.
- 등록된 연구자에 한해 다양한 포맷으로 학술논문을 제출할 수 있음.
- 장기보존을 위해 제출한 파일은 포맷 등 문제가 없는지 확인하는 과정을 거침.
- 모든 제출물의 영구적인 접근을 보장하며 다음과 같이 업로드 파일유형을

제한함.

- 지원하는 텍스트 파일포맷은 (La)Tex, AMS(La)Tex, PDFLaTeX, DOCX (Word 2007), PDF, PostScript, HTML with JPEG/PNG/GIF이며, 그림은 PostScript, JPEG, GIF, PNG, PDF임.

- 편리함과 이용 활성화를 위해 등록절차 없이 누구나 이용할 수 있지만, 학술논문을 제출하기 위해서는 반드시 등록해야함.
- 제출한 논문은 원칙적으로 삭제하지 않고, 신속한 유통을 통해 전통적인 출판시스템의 단점인 출판시간 지연을 해소함.

### ③ 논문제출 절차

#### 회원 승인

- 논문을 제출하기 위해 회원등록 시 일정한 승인절차를 거치는데 이는 커뮤니티에 활동 중인 멤버인지 확인하기 위해 다른 회원에게 승인을 받는 것임.

#### 제출물 심사

- 제출 시 간단한 심사과정을 거쳐 공개함. 제출물에 대한 심사는 다음 제출물의 마감시간 전(매일 미국 동부시간 기준 16:00)까지 실시함. 제출물 승인 후 식별자를 부여하여 메타데이터 복사본과 제출물의 링크가 포함된 이메일을 제출자에게 보냄. 또한 공저자에게 저작권을 청구할 수 있는 임시식별자와 비밀번호가 포함된 안내 메일을 발송함.

#### 제출물 변경

- 제출 후 에러 수정이나 개정판을 제출할 필요가 있으면 제출한 날 마감시간 전에 처리해야함.

#### 제출물 버전

- 마감시간 후 변경한 것은 새로운 버전으로 취급함.
- 제출한 논문의 버전을 관리하기 위해 별도의 레코드를 유지함.

- 이미 제출한 모든 버전은 삭제하지 않는 것이 원칙이지만, 삭제하길 원하면 취소통지를 보내야함.
- 취소는 제출물을 삭제하는 것이 아니라 이전 제출물에 대한 문제점을 설명하는 것을 허락하는 것임.

#### 제출물 사용

- 제출물은 공식적으로 심리에 소요되는 시간을 포함하여 1일 후에 서비스함(금요일이나 주말에 제출한 경우는 예외).

#### 제출물 중재

- 모든 제출물은 승인과정을 거쳐 삭제되거나 몇몇 경우는 재분류됨.

### ④ 서비스

- arXiv에서는 트랙 백(track back), 참조데이터, CiteBase 인용정보, 주제별 브라우징 등 다양한 서비스를 제공함.

- 연구자들이 자신의 논문을 arXiv에 제출하고 이용자들은 등록된 논문을 무료로 이용함. 관심있는 주제 분야에 신규 등록된 논문을 RSS와 이메일 알리미 서비스를 통해 받아 볼 수 있음.

- 연구를 위해 데이터 셋을 구축하거나 로컬 카피를 위해 다량의 논문을 빠르게 다운로드하는 기능은 지원하지 않음.

## 나. CiteSeerX

### ① 개요

- 사이트시어엑스(CiteSeerX)는 미국과학재단의 지원을 받아 펜실베이니아 주립대학교 소속 정보과학기술대학(IST)이 운영하는 오픈엑세스 리포지터리로 핵심 주제 분야는 컴퓨터 및 정보학임.

- CiteSeerX는 1997년 NEC 연구원, 프린스턴, 뉴저지 등에 소속된 연구자들이 공동으로 개발한 CiteSeer의 차세대 버전임.

- CiteSeerX는 과학과 학술적인 지식접근을 위한 기능성, 유용성, 이용가능성, 편의성, 유효성, 적시성을 제고하고, 학술정보의 유통 활성화 및 비용 장벽을 제거하기 위해 개발함.

〈그림 6〉 CiteSeerX 홈페이지



## ② 콘텐츠 수집 및 관리정책

- 연구자들이 학술논문을 등록하는 방식보다는 주로 웹 크롤러를 통해 학술 논문을 수집함.
- 연구자들은 CiteSeerX에 부합하는 주제 분야의 콘텐츠라면 어떤 콘텐츠든지 제출할 수 있으며, 제출 전 동료에게 검토 받는 것을 권장함.
- 파일포맷은 PDF 파일을 권장하며 PDF, PostScript, 압축파일(zip·gz·z) 등 다양한 파일을 지원함.
- 검색 서비스를 이용하기 위해 회원으로 등록할 필요는 없음. CiteSeerX에 논문을 제출하거나 정보를 주고받을 때, 질문하거나 의견을 제시하고 정보

갱신을 요구할 때 등록이 필요함.

### ③ 서비스

CiteSeerX에서는 인용정보뿐만 아니라 참조링킹, 인용통계 등 콘텐츠 간 연계서비스를 제공하고 있음.

- CiteSeerX는 정보검색과 평가를 목적으로 인용색인을 자동으로 생성함.
  - 전통적인 인용색인과 비교했을 때 자동인용색인(Autonomous Citation Indexing, ACI)은 비용·편의성·유효성 측면에서 월등함.
- 인용된 논문과 인용통계 정보(Citation statistics)를 제공하며 참조링킹(Reference linking)을 통해 많은 온라인 출판사와 연계함.
- 관심 있는 논문에 쉽고 빠르게 접근할 수 있도록 인용상황 서비스(Citation context) 기능을 제공하며, 회원 프로파일에 등록된 논문의 새로운 인용 정보 통지 서비스인 주지 및 트래킹(Awareness and tracking) 서비스를 제공함.
- 현재 약 1,614,142건의 논문정보와 이들과 링크된 31,317,127건의 인용정보를 서비스하고 있음(2010.6).

전문을 색인(Full-text indexing)하여 불리언검색, 구문검색, 그리고 인접검색을 지원함.

일반적인 요약 대신 질의어가 논문기사에서 어떻게 사용되는지 문맥을 알 수 있는 질의기반 요약 기능(Query-sensitive summaries)을 제공함.

RSS, 소셜북마킹, 소셜네트워크, 개인별 검색환경 설정 등 개인화서비스(Personal Content Portal)를 제공함.

## 다. DTIC Online

### ① 개요

DTIC는 미 국방부(Department of Defense, DoD)에서 운영 중이며, 연구개발을 지원하기 위해 콘텐츠를 수집하고 데이터베이스를 구축하여 검색 서

비스를 제공하는 국가정보센터임.

- DTIC(Defense Technical Information Center) online은 전 세계 리포지터리 중에서 Google과 Yahoo에 PDF 파일을 가장 많이 제공하였으며, 지난 65년간 국방 분야 정보서비스 중심기관 역할을 수행하고 있음.
  - 웨보메트릭스의 리포지터리 순위 3위에 있는 Scientific and Technical Information Network는 DTIC online으로 더 잘 알려짐.
- 국방 커뮤니티들을 지원하기 위해 정보를 수집, 조직, 검색, 배포하고 있으며, 국방 관련 정보서비스, 데이터베이스, 시스템, 네트워크를 중점적으로 운영하고 있음.
- DTIC 협력기관으로는 미 국방부와 정부부처, 대학, 과학기술 활동에 참여하는 비영리 조직, 외국정부 등이 있음.

〈그림 7〉 DTIC 홈페이지



## ② 콘텐츠 수집 및 관리정책

- 오프라인과 온라인 두 가지 방식으로 콘텐츠를 제출할 수 있음.
- MS Word, Excel, Powerpoint, WordPerfect, ASCII 텍스트, PDF 등 다양한 포맷을 지원함.
- DTIC는 회원등록 시 자격규정을 두어 DTIC의 웹서비스를 이용하거나 접근할 수 있는 수준을 정하는데 활용하고 있음.
- 수집 콘텐츠는 생물학, 의학, 환경공학, 행동과학 그리고 사회과학을 포함한 국방 관련 모든 분야임.
  - 미 국방부의 지침이나 지시, 예산정보, 회의나 심포지엄, 특허와 특허제품 그리고 국방 커뮤니티와 연관된 다양한 콘텐츠 포함.
  - DTIC 기술보고서
    - 미 국방부에서 지원한 연구개발시험평가(RDT&E) 결과보고서와 과학기술 문서
  - 연구과제 요약
    - 현재 진행 중인 미 국방부 R&D과제를 설명하는 기술적이며 관리적인 데이터
  - 독자연구개발(Independent Research and Development, IR&D)
    - 국방과 관련된 독자적인 연구개발 결과물을 방위산업체에서 자발적으로 제공

## ③ 서비스

- DTIC는 세 개의 웹사이트(DTIC Online, DTIC Online Access Controlled, DTIC Online Classified)를 운영하고 있으며, 각 사이트마다 등록 수준이 상이하고, 포함하고 있는 정보범위와 접근수준도 다름.
  - DTIC Online
    - 일반인을 위한 것으로 별도의 등록절차가 필요 없음.

- DTIC Online Access Controlled
    - 등록된 회원만 이용할 수 있고, 미 국방부 내 연구 활동에 관한 정보를 제공하는 것에 초점을 맞춤.
    - 미 국방부 사이트, 협력기관으로부터 수집한 콘텐츠, 기술보고서, 예산정보, 학술지나 회의자료, 전자책, 진행 중인 연구과제, 표준과 지침 등
  - DTIC Online Classified
    - 이용을 위해 추가적인 등록이 필요하고, SIPRNET(Secret Internet Protocol Router Network)상에서만 무료로 이용 가능함.
    - 2백만건 이상의 DTIC 기술보고서를 이용할 수 있으며, DTIC Online Access Controlled에서는 제공하지 않는 30만건 이상의 연구요약 정보 전문(Full Text) 서비스
- 최근 정보보안과 무결성을 보장하기 위해 시스템 보안, 연구자 인증, 접근 통제와 서명시스템을 고도화함.

#### 라. Social Science Research Network

##### ① 개요

- SSRN(Social Science Research Network)은 사회과학분야 학술정보를 웹 상에서 빠르게 배포하기 위한 연구네트워크로, 약 140개 국가의 400,000명의 연구자들이 등록한 연구논문과 초록을 SSRN eLibrary를 통해 무료로 제공함.
- SSRN에는 수백 종에 달하는 학술지, 출판사, 그리고 출판사 협력기관들이 참여하여 SSRN의 eLibrary를 통해 논문을 서비스하고 있으며, SSEP(Social Science Electronic Publishing Inc.)에서 운영함.

##### ② 콘텐츠 수집 및 관리정책

- 논문제출 시 어떤 제한도 없으며 논문을 제출해야만 이용할 수 있는 것도 아님. 연구자들은 최신 논문뿐만 아니라 오래된 논문도 제출할 수 있음.
- 제출자가 요청하면 SSRN eLibrary에 등록된 논문을 즉시 삭제하고, 연구

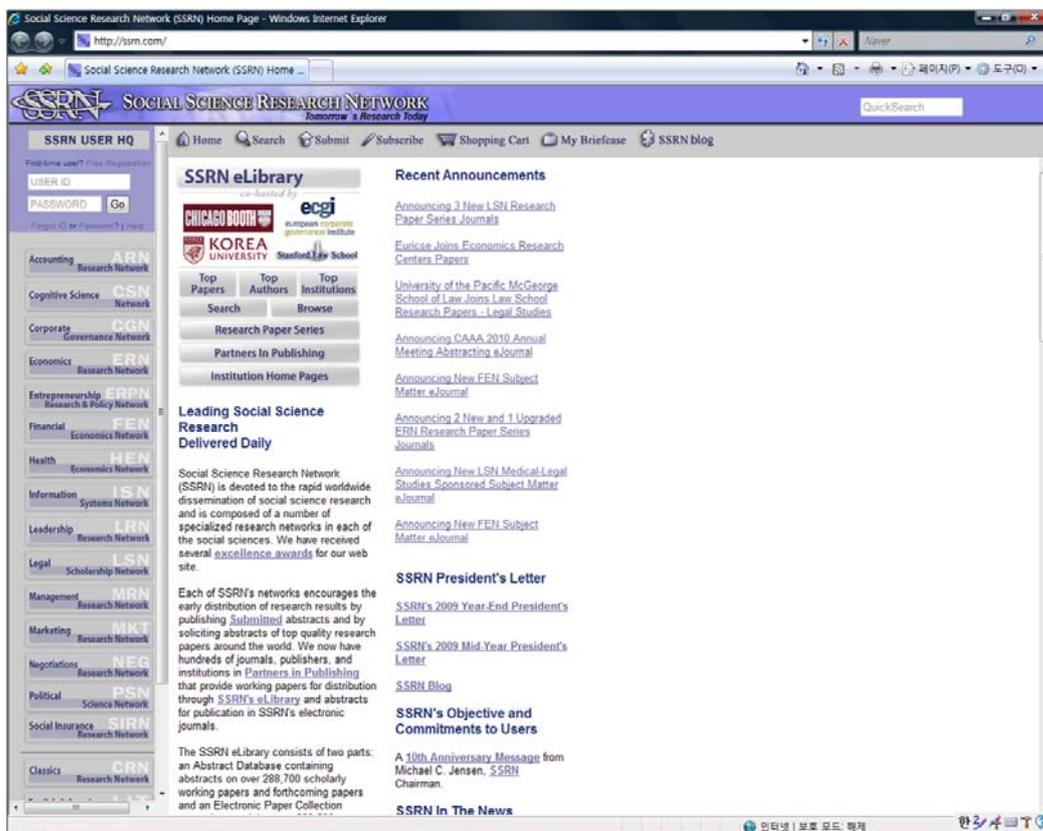
자는 로그인후 자신이 제출한 논문을 언제든지 수정할 수 있음.

□ SSRN은 연구자가 초록과 논문을 직접 제출하고, 자신의 논문 구독자를 관리함.

- 연구자는 초록과 함께 연구자 이름, 연락처를 명기하고 모든 공저자도 SSRN에 가입해야 함.
- 연구자와 저작권자가 다른 경우 저작권자에게 받은 위임장, 이메일, 주소, 전화번호, 팩스번호, PDF로 된 논문을 함께 제출해야 함.

□ 회원 가입에 제한을 두진 않지만, 18세 미만은 논문을 제출할 때 사전에 보호자나 관리자의 승인이 필요함.

〈그림 8〉 SSRN 홈페이지



□ 권한 제한

- 어떠한 경우라도 공지 없이 이용이나 서비스를 제한할 수 없음.

- 웹사이트에 저장된 어떤 콘텐츠도 이동하거나 수정하는 권한이 제한되며, 공지 없이 다운로드 카운트, 초록보기, 이용통계를 조정할 수 없음.
- 독단적으로 서비스를 거절하거나 중지하고, 콘텐츠를 삭제하거나 편집하고, 제출된 논문을 취소할 수 없음.

권리 유지

- 연구자는 SSRN에 콘텐츠를 제공하더라도 권리는 완전하게 유지함.

출판사 저작권

- 일부 논문의 경우 출판사가 저작권을 가지고 있음. 따라서 이런 콘텐츠를 이용할 때는 요금을 지불해야 함.

SSRN 권리

- SSRN은 연구자나 출판사가 제출한 콘텐츠를 제외한 모든 콘텐츠에 대한 권리를 소유함.

개인적인 용도로만 사용 가능

- SSRN 웹사이트에서 제공하는 서비스와 정보는 상업적인 목적이 아닌 오직 개인적인 용도로만 이용할 수 있음.
- 사이트의 어떤 콘텐츠도 SSRN의 허락 없이 상업적이든 비상업적이든 재포맷, 재저장, 또는 재배포할 수 없음.

③ 서비스

SSRN eLibrary에서는 금융, 회계, 법률, 경제 그리고 경영분야의 논문들을 구축하여 서비스함.

- 초록 데이터베이스(Abstract Database)
  - 약 288,700건의 논문 서비스
  - 제목, 저자 연락처, 등록정보, 초록, 이용이력, 원문정보 수록
- 전자논문 컬렉션(Electronic Paper Collection)
  - 239,500건 이상의 PDF 논문 서비스
- 연구자 홈페이지(Author Home Page)
  - 초록 데이터베이스에 각 연구자 홈페이지 링크

- 연구자와 연구자 간 직접적인 커뮤니케이션을 지원하기 위해 모든 논문에 연구자 이메일 정보를 제공함.

#### 마. Smithsonian/NASA Astrophysics Data System

##### ① 개요

- ADS(SAO/NASA Astrophysics Data System)는 NASA의 동의하에 스미소니언 천체물리관측소(Smithsonian Astrophysical Observatory, SAO)가 운영하는 천문학과 물리학 연구자들을 위한 디지털도서관임.
- ADS는 8,400,000건 이상의 레코드를 보유한 세계 데이터베이스를 구축·운영하고 있음.
  - 천문학과 천체물리학(Astronomy and Astrophysics), 물리학, arXiv e-prints 미리 서버

〈그림 9〉 ADS 홈페이지



## ② 콘텐츠 수집 및 관리정책

- 초록 데이터베이스는 NASA의 STI 그룹(Scientific and Technical Information group), 학술지 출판사인 SIMBAD, NED 등 다양한 정보원에서 콘텐츠를 수집하여 구축함.
  - NASA의 STI 그룹으로부터 1975년부터 1995년 중반까지 콘텐츠를 제공 받았지만, 그 이후에는 대부분 출판사로부터 직접 받음. 2007년부터는 CrossRef에서도 수집하고 있음.
  - 매주 초록을 갱신하고, 출판 전에 출판사로부터 레코드를 받아 시스템에 저장함.
  
- 1997년부터 천문학 데이터베이스에 참고문헌 데이터를 구축함.
  - ADS는 ISI(Institute for Scientific Information, 현재 Thomson Reuters)사의 인용 데이터베이스에서 1980년부터 1998년까지 천문학 분야 학술지에 수록된 인용정보를 구입하여 초기 데이터를 구축함.
  - 1999년부터는 학술지 출판사에서 제공받은 원문에서 참고문헌 데이터를 추출하기 시작하였으며, 2007년 10월부터는 참고문헌을 CrossRef과 연계함.
  - 2009년 7월 1,600백만건 이상의 참고문헌 데이터를 추출함.
  
- 콘텐츠 처리원칙
  - 출판전 논문(preprint): 고유 식별자를 부여하여 인용 가능함.
  - 출판중인 논문: 대체로 심사나 검토과정을 마친 상태로 출판된 것과 대동소이하기 때문에 데이터베이스에 포함하고 있음.
  - 제출된 논문(submitted paper): 권/호나 페이지가 없고 출판된 최종논문과 상이하므로 데이터베이스에 포함하지 않음.
  
- 문헌을 식별하기 위해 고유 식별자를 부여함.
  - 고유 식별자는 19개의 digit으로 구성하고, 포맷은 SIMBAD와 NED 프로젝트에 의해 채택된 것을 사용함.
  - 포맷은 YYYYJJJJVVVVMPPPPA 형태로, YYYY는 출판년, JJJJ는 저널의 표준 축약명, VVVV는 볼륨번호, M은 콘텐츠 상태와 유형, PPPP는 페이지 수, A는 제1저자 성의 첫 글자임.

- 데이터베이스 업데이트는 일주일에 한 번 이루어지며, 모든 미러 사이트도 동시에 갱신함.

### ③ 서비스

- 편리하고 다양한 검색기능 제공
  - Basic Query Form
    - 기본 검색을 제공하며 하나의 필드에 검색 키워드를 입력하고 검색버튼을 클릭하는 방식임.
  - Abstract Query Form
    - ADS 초록 데이터베이스를 검색할 수 있음.
  - ADS Topic Search
    - 문헌을 검색하기 위해 6개의 다른 모드를 제공함.
  - Searching by Reference Information
    - 발행년·저널명·권/호·페이지·저자명을 이용하여 참고문헌을 검색할 수 있음.
  - Current/Unread Journals
    - 선택한 학술지의 신착논문을 보여주거나 선택한 학술지 중 아직 읽지 않은 논문을 제공함.
  - Searching the ADS by Email
    - 이메일을 통해 질의하고 결과를 이메일로 받을 수 있음.
- ADS는 전자논문, 데이터 목록, 기록물을 포함하여 풍부한 외부 콘텐츠로 링크를 제공함.
- ADS에 포함되어 있는 모든 초록과 논문의 저작권을 출판사로부터 양도받아 무료로 서비스함.
- 이용자들은 arXiv e-prints에 등록된 출판전 논문을 이메일로 받아 볼 수 있으며, RSS를 통해서도 최신정보를 실시간으로 확인할 수 있음.

## 바. RePEc(Research Papers in Economics)

### ① 개요

- RePEc(Research Papers in Economics)은 경제학 분야 학술정보 유통을 촉진하기 위해 70개국 수백 명 연구자들이 참여하는 지식협력체임.

〈그림 10〉 RePEc 홈페이지



- 연구보고서, 논문, 소프트웨어 등 분산데이터베이스로 구성되어 있으며 RePEc의 모든 콘텐츠를 무료로 이용할 수 있음.

### ② 콘텐츠 수집 및 관리정책

- 경제학 분야 학술정보는 어떤 언어로 되어있든지 제출할 수 있으며, 저작권 소유자의 동의하에 이전에 출판한 논문도 제출할 수 있음.

- 제출된 논문은 편집자의 검토와 승인을 거쳐 등록함.

- RePEc 논문제출 규정

- 논문의 경우 학술적이어야 하며 경제학 분야거나 경제학과 관련된 수학적 방법, 역사, 사회, 행동과학, 통계 관련 주제 분야도 포함함.
  - 다른 학술정보를 요약한 것은 허용하지 않지만 리뷰 논문은 가능함.
  - 논문의 원문을 다운로드할 수 있어야 하며, 원문이 없는 논문을 제출할 수 없음.
  - RePEc 관리자들은 기술적인 부문과 가이드라인에 따라 논문을 검토함.
- 연구자들은 기관 리포지토리를 통해 RePEc으로 콘텐츠를 제출할 수 있으며, 소속기관에 리포지토리가 없는 경우 MPRA(Munich Personal RePEc Archive)에 제출하면 자동으로 RePEc에 반영됨.
  - 학술정보를 제출하기 위해서는 RePEc Author Service를 통해 회원을 등록해야 됨. 등록된 학술정보와 연구자는 항상 링크되어 있으며, 연구자는 제출한 학술정보의 다운로드 회수나 인용정보를 확인할 수 있음.
  - 상업적으로 이용되는 콘텐츠를 제외하고 누구나 RePEc에 있는 콘텐츠를 무료로 이용할 수 있으며, 저작권이 있는 콘텐츠의 경우 이를 준수해야 함.
  - RePEc은 개인저장소는 지원하지 않으며 기관 리포지토리 구축을 지원함.
  - RePEc 저장소를 구축한 기관에서 콘텐츠를 등록하면, RePEc에서는 정기적으로 기관들의 서버를 방문하여 추가, 수정, 삭제된 콘텐츠를 수집하여 RePEc 중앙서버에 반영함.
  - 소속기관에 RePEc 리포지토리가 없는 경우 MPRA(Munich Personal RePEc Archive)를 통해 개별적으로 RePEc에 학술정보를 등록할 수 있음.
- ③ 서비스
- 학술정보를 RePEc의 EconPapers, Ideas, Economists Online and other services의 데이터베이스를 통해 서비스함.

- RePEc은 논문의 원문을 구축하지 않고 원문으로의 링크정보만 제공함. 약 90만건 구축되어 있으며 이중 80만건 이상을 온라인으로 이용할 수 있음.
  - 연구보고서 35만건, 논문 55만건, 단행본 19,000건, 연구자와 출판사 리스트 24,000건, 기관 연락처 11,700건, 그리고 소프트웨어 1,900건
- 저자에게 매달 이용현황 및 인용정보를 메일링 서비스함.

## 사. MIT DSpace

### ① 개요

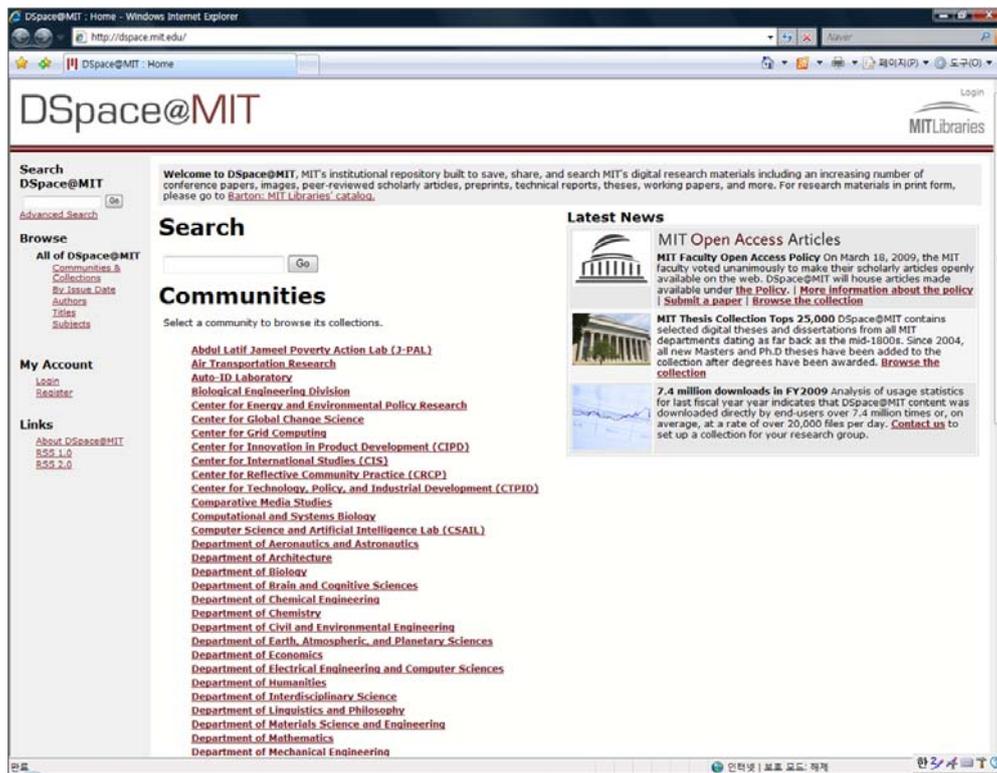
- MIT(Massachusetts Institute of Technology) 교수회는 연구 및 학술적인 결과물을 광범위하게 배포하기 위한 방안으로 DSpace@MIT를 구축함.
- MIT 교수들은 자신들의 학술논문에 대한 권리를 학교에 양도함으로써 저술한 학술논문을 무료로 사용할 수 있게 하고 있음.

### ② 콘텐츠 수집 및 관리정책

- 초기에는 교수들의 논문만 수집하다가, 연구원들의 논문 중에서 유용성을 인정받은 것과 MIT 교수가 공저자인 논문까지 확대함. 하지만 학생들이 작성한 논문은 수집하지 않음.
- 콘텐츠 수록범위는 논문, 출판전 논문, 기술보고서, 토의문서, 전자논문, 각종 연구데이터, 이미지/오디오/비디오, 기타 도서관 장서임.
- 출판사가 저작권 문제로 리포지터리에 논문 아카이빙을 허용하지 않는 경우 다음과 같은 해결방안을 제시함.
  - MIT의 비독점적 라이선스를 출판사가 받아들이도록 설득
  - 리포지터리에 논문을 아카이빙할 수 있는 다른 출판사를 찾도록 권고
  - 학술출판 라이선싱 고문(Scholarly Publishing & Licensing Consultant)이나 총괄 고문 변호사(Office of General Counsel)에게 자문
  - 논문에 대한 저작권 포기각서를 얻기 위해 노력

- 오픈액세스 의무화 정책 후 MIT 교수들은 학술논문 저작권을 MIT에 이양하고 있음.

〈그림 11〉 MIT DSpace 홈페이지



### ③ 서비스

- 회원자격에 따라 핵심 서비스(Core Service)와 프리미엄 서비스(Premium Service)로 이원화
- 핵심 서비스(Core Service): 누구나 자유롭게 이용할 수 있는 서비스로, 셀프아카이빙, 검색 및 브라우징, 콘텐츠 장기보존, 커뮤니티 지원, 질의 응답, 시스템 모니터링, 백업 등을 제공함.
  - 프리미엄 서비스(Premium Service): 커뮤니티 회원 중 자관 환경에 맞게 커스터마이징을 요청하는 기관들에게 제공하는 유료서비스로, 전자문서 변환서비스, 메타데이터 제공서비스, 추가 저장용량 제공, 이용통계 제공 서비스 등이 있음.

## 4. 사례분석 및 시사점 도출

### 1) 운영 및 지원정책

□ 리포지터리의 실질적인 운영주체는 대학이나 연구소가 대부분임. 대학이 32개(64%)로 가장 많고, 연구소 15개(30%), 출판사 2개(4%), 개인연구자 1개(2%)순임.

□ 주제 리포지터리의 경우 실질적인 운영주체와 관리주체가 분리되어 있는 곳이 많음. 이와 같은 경우 실질적인 운영주체는 주제별 커뮤니티로 논문 제출 및 심사 등을 주도하고, 관리주체는 대학 또는 연구소로 사이트 관리 등 최소한의 역할만 수행함.

- arXiv는 코넬대학도서관에서 사이트를 관리하고 있지만 실질적인 운영주체는 물리학·수학·컴퓨터공학 등 학회 중심의 전 세계 차원의 연구자 중심 커뮤니티임.

• arXiv에서는 규제를 최소화하고 사이트를 유지관리하기 위한 의무를 강화함.

• 코넬대학 도서관은 arXiv를 유지하고 지속적인 운영체제를 구축하기 위해 운영자금 확보 등 새로운 비즈니스 모델을 발굴하기 위해 노력하고 있음.

- SSRN은 운영주체가 웹사이트·정책·이용조건 등을 임의로 변경할 수 없으며, 제출된 어떤 콘텐츠도 삭제·수정·이동할 수 없도록 운영주체의 권한을 제한하고 있음.

- RePEc은 연구자가 논문을 중앙에 직접 제출하거나 각 기관에 구축된 RePEc 저장소를 통해 제출할 수 있는 분산체제임. RePEc에서는 각 기관에서 공통적으로 필요한 리포지터리 운영기술 등을 제공하고, 실질적인 운영은 각 개별기관이 담당함.

□ 이에 반해 기관 리포지터리는 운영 및 관리주체가 동일하고 보다 강력한

권한을 가지고 있음.

- MIT DSpace는 논문 제출을 의무화하기 위해 교수회의 지원을 받는 매우 강력한 권한이 있는 운영주체임.

- 학회와 같이 자발적인 커뮤니티가 형성될 수 있는 범기관적 주제 리포지터리는 활성화를 위해 운영주체 권한을 최소화하는 것이 필요하지만, 기관 리포지터리 같은 경우 소속 구성원들의 참여와 활성화를 위해 제출 의무화 등 보다 강력한 운영주체가 필요함.

## 2) 수집정책

- 정보수집 방법에는 연구자가 직접제출하는 방식과 로봇 등을 이용한 자동 수집 방식이 있음. 대부분의 리포지터리는 보다 많은 콘텐츠를 확보하기 위해 연구자가 직접제출하는 방식과 다양한 수집방식을 병행하고 있음.

### - 직접제출

- arXiv는 연구자들의 직접제출 방식만으로 운영되는 리포지터리이며, MIT DSpace는 소속 대학 교수와 연구원들에게 논문 제출을 의무화하여 수집함.

### - 직접제출과 로봇 수집 등 병행

- CiterSeerX는 연구자 직접제출뿐만 아니라 웹상에 존재하는 유용한 학술 논문을 로봇으로 자동 수집함.
- DTIC도 미 국방부 관련 기관들이 구축한 데이터베이스를 일차적으로 수집하고 직접제출을 통해 확장하는 혼합형임.
- SSRN은 연구자 직접 제출뿐만 아니라 출판사를 통해 초록을 확보하고 있음.
- ADS는 연구자가 직접제출하는 방식과 NASA와 arXiv로부터 콘텐츠 수집을 병행함.
- RePEc은 기본적으로 각 기관 리포지터리에 구축된 콘텐츠를 수집하고, 기관 리포지터리가 없는 기관에 소속된 연구자의 경우 중앙에 직접 제출할 수 있도록 정책을 마련함.

- RePEc은 다량의 원문을 직접 구축하지 않고 원문을 이용할 수 있는 링크 정보를 제공함으로써 이용증가와 가시성을 높이고 있음.

### 3) 관리정책

#### 가. 품질관리

- 리포지터리에 논문을 제출하는 것은 대부분 개방성과 자율성을 목적으로 하기 때문에 출판사나 학회에서 실시하는 동료심사나 검토과정이 없음. 다만 최소한의 품질관리를 위해 등록된 연구자만 콘텐츠를 제출할 수 있으며, 등록된 콘텐츠에 질의 응답할 수 있도록 저자정보를 제공함.
- 대부분은 리포지터리를 활성화하기 위해 논문 제출을 적극 권장하고 있으며 어떤 제한도 두지 않고 있음. 이에 반해 최소한의 품질을 관리하는 곳도 있음.
  - MIT DSpace는 콘텐츠를 제출할 수 있는 사람을 교수와 연구원으로 한정하여 콘텐츠의 품질을 관리하며, DTIC는 승인받은 사람만 제출할 수 있음.
  - RePEc은 연구자가 직접 제출하는 경우 가이드라인을 명시하고 이에 부합하지 않는 논문은 관리자가 판단하여 거절함.
  - arXiv에서도 최소한의 인증절차를 두어 제출한 논문이 주제범위를 벗어난 경우 거절함.
  - 하지만 대부분 깊이 있는 심사보다는 학술적인지, 리포지터리 주제범위에 부합하는지만 판단함.

#### 나. 콘텐츠 포맷

- 대다수 리포지터리는 제출할 수 있는 문서포맷을 제한하고 있음. 이는 콘텐츠 장기보존뿐만 아니라 이용 편리성을 제고하기 위해서임.
  - arXiv는 텍스트 제출포맷은 (La)Tex, AMS(La)Tex, PDFLaTex, DOCX, PDF, PostScript, HTML만 허용하고, 그림은 PostScript, PNG, PDF로 제한함.

- CiteSeerX는 PDF, PostScript, zip, gzip을 허용하지만 가장 권장하는 포맷은 PDF임. 따라서 파일포맷이 맞지 않은 경우 대개 PDF로 변환하여 저장함.
- SSRN은 PDF만 허용하고, ADS는 PDF와 PostScript 파일만 가능함.
- 이에 반해 가장 많은 콘텐츠를 보유하고 있는 DTIC는 다양한 파일포맷을 허용함.
- PDF 형태를 선호하는 이유는 논문에 텍스트뿐만 아니라 이미지나 멀티미디어를 포함할 수 있으며, 전 세계 이용자들이 무료로 뷰어 프로그램을 이용하여 편리하게 원문을 볼 수 있기 때문임.

#### 다. 고유 식별자 부여

- 제출된 논문을 관리하고 접근을 개선하기 위해 고유 식별자를 부여함.
  - arXiv는 승인된 각 논문에 고유 ID를 부여하고, 보다 편리하게 메타데이터를 제출할 수 있도록 관련 시스템을 제공함.
  - SSRN에서도 제출한 논문에 고유 URL를 부여하고 있음.

#### 라. 사용언어

- 모든 리포지터리들이 홈페이지와 하위 페이지를 영어로 제작하여 병행 서비스하고 있음.
  - 영어가 모국어인 경우 모국어와 함께 영어 버전을 별도로 제공함.
- 대부분 논문 제출시 언어에 제한을 두지 않으며, RePEc은 정책에 언어제한이 없음을 명시하고 있음.

#### 마. 콘텐츠 유지

- 리포지터리에 제출된 콘텐츠는 제출자가 요청하지 않는 한 운영주체나 다른 연구자들이 삭제하거나 수정할 수 없음.
  - arXiv는 제출자가 요청하더라도 이전 판을 삭제하지 않으며, 초판, 수정판, 최종판이 제출되더라도 각각 별도의 레코드로 관리함.

#### 바. 이용조건

- 대부분의 리포지터리는 오픈액세스 기반 무료 이용과 확산을 목적으로 구축한 시스템이기 때문에 회원등록이나 로그인 없이 모든 콘텐츠를 무료로 이용할 수 있음. 단, DTIC와 같이 국가적 보안이 요구되는 경우만 제한함.

### 4) 배포 및 이용활성화 정책

#### 가. RSS 및 알리미 서비스

- 대부분의 리포지터리에서는 최신정보, 뉴스정보, 이용통계, 인용상황 등을 신속하게 제공하기 위해 RSS와 알리미 서비스를 제공함.
  - arXiv는 회원 프로파일에 등록된 관심 분야와 관련 있는 새로운 논문이 입수되면 RSS와 이메일 알리미 서비스로 제공함.
  - CiteSeerX도 알리미 서비스로 관심 있는 논문의 인용상황을 제공함.
  - SSRN는 RSS를 이용하여 새로운 논문이나 뉴스 등을 신속하게 전달함.
  - RePEc은 알리미 서비스로 이용자가 제출한 논문의 이용통계와 인용정보 까지 제공함.
- 이와 같은 서비스는 최신정보 제공뿐만 아니라 콘텐츠에 대한 지속적인 관심을 유발시켜 리포지터리를 활성화하는데 도움이 됨.
  - RePEc과 같이 제출한 논문의 이용통계와 인용상황을 제공하면 리포지터리에 대한 관심을 유발시켜 지속적으로 새로운 논문을 확보할 수 있음.

#### 나. 이용통계 정보

- 대다수 리포지터리에서 이용통계 정보를 제공하고 있으며, RePEc은 방문하지 않는 회원에게는 매달 알리미 서비스로 이용통계 정보를 제공함.
  - 리포지터리와 같이 자발적인 참여가 운영의 큰 부분을 차지하는 경우, 높은 이용자수는 많은 이용으로 이어지고 더 많은 논문 제출을 유도할 수 있는 요인임.

#### 다. 인용정보 분석서비스

- CiteSeerX, RePEc, ADS에서는 인용정보 분석서비스를 제공하고 있으며,

인용정보는 기관 순위를 높이는데 매우 중요한 요소임.

- <표 2>와 같이 CiteSeerX, RePEc, Smithsonian/NASA Astrophysics Data System 리포지터리에 구축된 콘텐츠 보유량은 다른 리포지터리와 비교하여 적은 편이지만, 논문의 인용 및 참고문헌 정보와 링크를 함께 제공함으로써 리포지터리 크기(Google, Yahoo, Live Search, Exalead에서 발견할 수 있는 페이지의 수)는 1, 2, 3위로 가장 크고, 리포지터리 순위에 가장 큰 영향력을 미치는 가시성도 매우 높게 나타남.

<표 2> 인용정보 제공으로 높은 순위에 오른 리포지터리들

종합 순위	리포지터리	크기 순위	가시성 순위	리치파일 순위	구글스칼라 문헌수
2	CiteSeerX	1	1	411	2
5	RePEc(Research Papers in Economics)	2	6	140	5
6	Smithsonian/NASA Astrophysics Data System	3	4	539	1

- 이와 같은 사례는 대량의 콘텐츠 수집이 어렵거나 부족한 경우, 다른 사이트와 연계체제를 구축함으로써 크기와 가시성을 향상시킬 수 있는 매우 성공적인 사례임.

#### 마. 다양한 검색 인터페이스

- 대다수 리포지터리에서는 다양한 검색기능을 제공함.
- arXiv는 저장소와 논문번호를 아는 경우와 그렇지 않은 경우를 구분하여 두 가지 인터페이스를 제공함.
  - DTIC는 세 가지 검색 인터페이스인 DTIC Online, DTIC Online Access Controlled, DTIC Online Classified를 제공함.
  - ADS는 Basic Query Form외 총 7가지 검색 인터페이스를 제공함.
  - MIT DSpace는 회원에 따라 핵심 서비스(Core Service)와 프리미엄 서비스(Premium Service)로 구분함.

## 5) 장기보존 정책

- 리포지터리의 중요한 역할 중 하나는 디지털 콘텐츠의 장기보존임. 제출된 논문을 장기적으로 보존함으로써 일정기간 경과하여 저작권 문제가 해결되면 콘텐츠를 무료로 이용할 수 있음.
  - arXiv는 제출된 논문에 대한 장기보존 정책을 다른 리포지터리보다 강력하게 명시하고 있음.
    - 등록된 논문은 제출자의 요구가 있더라도 삭제하거나 수정하지 못함.
  - SSRN도 리포지터리에 저장된 어떤 콘텐츠도 이동하거나 수정할 수 있는 권한이 운영주체에게 없음.
    - SSRN 단독으로 서비스를 중지하거나 거절할 수 없도록 명시함.
  - 제출한 논문의 유지, 관리, 장기보존, 그리고 지속적인 서비스 약속은 더 많은 연구자 참여를 유도할 수 있음.

## 6) 저작권 정책

- 리포지터리는 오픈액세스 기반의 학술정보 장기보존 및 확산이 주 목적이므로 리포지터리에 구축된 콘텐츠는 누구나 무료로 이용 가능함.
- arXiv, SSRN eLibrary, RePEc은 저작권이 저자에게 있는 출판전 논문이나 출판중인 논문을 주로 수집하고, 저작권이 출판사로 이관된 논문을 리포지터리에 아카이빙하는 경우 출판사의 저작권 위임서나 승인서를 함께 제출 받고 있음.
- MIT DSpace는 소속 교수들과 연구원들에게 출판사의 발행일과 같은 날에 논문을 제출하도록 의무화함. 이는 연구 중인 논문의 경우 품질이 낮을 수 있으므로 최종 논문을 제출받아 보다 양질의 논문을 수집하고 동시에 저작권 문제를 해결하기 위한 전략임.
  - 출판사가 아카이빙을 허락하지 않는 경우 다른 출판사로 변경을 요구하는 보다 강력한 가이드라인을 제시하고 있음.

## 7) 리포지터리간 정책비교

- 대다수 리포지터리는 콘텐츠를 제출하거나 이용하는데 별다른 제한이 없으며, 미 국방 관련 콘텐츠를 서비스하는 DTIC의 경우 보안상의 문제로 일부 이용을 제한함.
- 콘텐츠 확충 및 수집 활성화를 위해 직접제출뿐만 아니라 로봇, 기관 간 협력 등 다양한 수집방식을 사용함.

〈표 3〉 리포지터리간 주요정책 비교

구분		arXiv	Cite	DTIC	SSRN	ADS	RePEc	MIT	
운영 및 지원정책	운영권한 최소화	○	○		○	○	○		
	운영권한 강화			○				○	
수집정책	직접제출	○		○					
	직접제출 + 로봇		○		○	○	○	○	
관리정책	최소 품질관리	○	×	○	×	×	○	○	
	콘텐츠 포맷 제한	○	○	×	○	○	○	○	
	고유식별자 부여	○	○	○	○	○	○	○	
	영어 + 자국어	○	○	○	○	○	○	○	
배포 및 이용활성화 정책	이용제한 없음	○	○	×	○	○	○	○	
	메일링 서비스	알리미	○	○	○	○	○	○	○
		인용정보		○			○	○	
		이용통계						○	
장기보존 정책	명문화		○		○				
저작권 정책	출판전 논문 수집		○		○	○	○		

- 정보 접근성 제고 및 장기보존을 위해 콘텐츠 포맷을 제한하고 고유식별자를 부여함. 아울러 자국어와 세계 공용어인 영어를 병행 서비스함.
- 최신정보 알리미서비스, 논문에 대한 인용정보 및 이용통계를 연구자에게 제공하여 지속적으로 리포지터리에 관심을 갖게 함으로써 더 많은 콘텐츠를 제출하도록 유도함.
- 저작권 문제를 해소하기 위해 저작권이 저자에게 있는 출판전 논문기사 위주로 수집하고 있으며, 리포지터리 주요역할 중 하나인 콘텐츠 장기보존을 정책적, 기술적으로 지원함.

## 5. 결론

- 학술정보 유통의 새로운 패러다임인 오픈액세스 기반 학술정보 유통을 실현하기 위해서는 리포지터리 역할이 매우 중요함.
- 국내외 많은 대학이나 연구소에서 리포지터리를 활발하게 구축·운영하고 있으며 향후 급격하게 증가할 전망이다. 국내 리포지터리는 아직 초기단계로 리포지터리 활성화를 위해 명확한 운영 가이드라인과 정책을 마련할 필요가 있음.
- 본 연구를 통해 도출된 시사점을 국내 리포지터리에 일괄적으로 적용하기에는 다소 무리가 있으며, 기관의 특성과 리포지터리의 역할 등을 감안할 필요가 있음.
- 리포지터리의 근본적인 목적이 연구자의 자발적인 참여와 이를 통한 개방적인 정보공유 환경을 조성하는 것이므로 너무 많은 규제나 제한은 오히려 장애요인이 될 수 있음.
- 향후 출판사 등과의 저작권 문제는 성공적인 리포지터리 운영의 큰 도전 과제임.

## 〈참고문헌〉

- 황혜경, 김혜선, 최선희. 2004. “오픈엑세스기반 지식정보저장소 구축에 관한 연구.” 한국비블리아, 15(1): 91-116.
- 황혜경, 이지연. 2009. “오픈엑세스기반 리포지터리 성공에 미치는 요인 분석.” 정보관리학회지, 26(4): 35-56.
- Crow, Raym. 2002. The Case for Institutional Repositories: A SPARC Position Paper. [cited 2010.4.23].  
<[http://scholarship.utm.edu/20/1/SPARC\\_102.pdf](http://scholarship.utm.edu/20/1/SPARC_102.pdf)>
- Davis, Philip M., Connolly, Matthew J. L. 2007. “Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University’s Installation of DSpace.” D-Lib Magazine, 13(3/4). [cited 2010. 4. 20].  
<<http://www.dlib.org/dlib/march07/davis/03davis.html>>
- Johnson, Richard K. 2002. “Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication.” D-Lib Magazine, 8(11). [cited 2010.4.20].  
<<http://www.dlib.org/dlib/november02/johnson/11johnson.html>>
- Lynch, C. A. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. [cited 2010.6.9].  
<<http://www.arl.org/newsltr/226/ir.html>>
- Westrienen, Gerald van, Lynch, Clifford A. 2005. Academic Institutional Repositories. D-Lib Magazine, 11(9). [cited 2010.4.20].  
<<http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>>
- arXiv.org. [cited 2010.4.15]. <<http://arxiv.org>>
- CiteSeerX. [cited 2010.4.20]. <<http://citeseerx.ist.psu.edu/>>
- DSpace@MIT. [cited 2010.4.20]. <<http://dspace.mit.edu/>>
- DTIC Online. [cited 2010.4.20]. <<http://www.dtic.mil/dtic/>>
- OCLC. [cited 2010.8.10]. <<http://www.oclc.org/oaister>>
- Ranking Web of World Repositories. [cited 2010.4.15].  
<<http://repositories.webometrics.info>>

RePEc. [cited 2010.4.20]. <<http://repec.org/>>

ROAR. [cited 2010.8.10]. <<http://roar.eprints.org>>

Social Science Research Network. [cited 2010.4.20]. <<http://ssrn.com/>>

The SAO/NASA Astrophysics Data System. [cited 2010.4.20].

<<http://adsabs.harvard.edu/>>

WIKIPEDIA. [cited 2010.6.15].

<[http://en.wikipedia.org/wiki/Repository\\_\(publishing\)](http://en.wikipedia.org/wiki/Repository_(publishing))>

creative commons korea, [cited 2010.9.24].

<<http://www.creativecommons.or.kr/info/about>>

◀ 저 자 ▶

이 상 기 | • KISTI 정보서비스실 선임연구원  
• sklee@kisti.re.kr

정 영 미 | • 동의대 문헌정보학과 조교수  
• yomjung@deu.ac.kr

KISTI 지식리포트 제12호

## 리포지터리 사례분석 및 시사점 도출

인 쇄 2010년 10월 5일

발 행 2010년 10월 7일

펴낸곳



펴낸이 박영서

편집장 최희운 편집간사 노경란

주 소 대전시 유성구 과학로 335

전화 042-869-1234, 팩스 042-869-1091

서울시 동대문구 회기로 66

전화 02-3299-6114

등 록 1991. 2. 12, 제5-258호

ISBN 978-89-6211-532-1 93500

인쇄처 승림디엔씨