

서울대답게 서울대처럼
LikeSNU 2023 ALA 국제혁신상 수상 기념
대학도서관 혁신 컨퍼런스

학술대회 발표자료집

2023. 7. 6.(목)



Contents

- 01 LikeSNU 탄생기: 어떻게 이력이 데이터가 되는가 07
서울대학교 중앙도서관장 장덕진
- 02 대학의 지식 빅데이터와 활용 방향 21
(주)사이람 배수진 차장
- 03 서울대학교 데이터거버넌스 전략과 실행 31
서울대학교 정보화본부 이종표 담당관
- 04 서울대학교 중앙도서관 빅데이터 특징과 활용방안 45
서울대학교 대학혁신센터 서은경 선임연구관
- 05 데이터와 데이터 이용문헌 간 링크의 기대 효과 59
서울대학교 한국사회과학자료원 구혜란 책임연구원

서울대답게 서울대처럼

LikeSNU 2023 ALA 국제혁신상 수상 기념

대학도서관 혁신 컨퍼런스

학술대회 발표자료집

01

LikeSNU 탄생기: 어떻게 이력이 데이터가 되는가

서울대학교 중앙도서관장
장덕진







Like SNU

도서관 데이터의 중요성과 활용

Google Scholar

전세계 아무나 검색

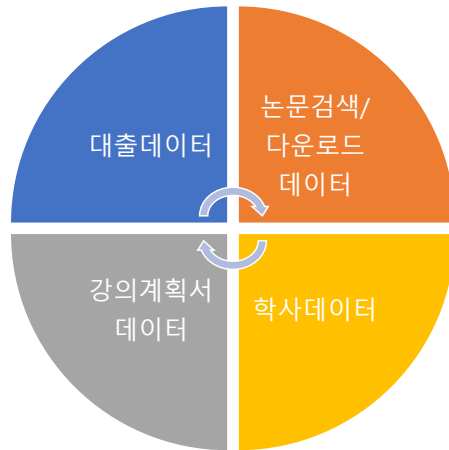
사용자 정보 적고
 세션 달라지면 동일
 인물 인식 불가

중앙도서관

최소 서울대 학부생
 수준 검색

상세한 사용자 정보.
 세션 달라져도 모두
 인식

- 중요성에도 불구하고 대출빈도 이외에는 한번도 분석된 적 없음
- 대출 데이터는 9년치 약 250만 건
- 논문 다운로드 데이터는 모두 Ex Libris로 넘어가며, 추후 외국 출판사는 우리의 대출 데이터를 분석해서 유료 구독을 유도할 것으로 예상
- 유료 구독 하더라도 분석결과만 받아보는 서비스로는 혁신에 활용할 수 없음
- 서울대에서 만들어지고 유통되는 전체 지식의 구조와 흐름을 알아야 혁신을 효율적으로 만들 수 있음
- 서울대 도서관 사서 100명, 하버드대 500명. 글로벌 경쟁을 위해서는 혁신적 방법이 필요



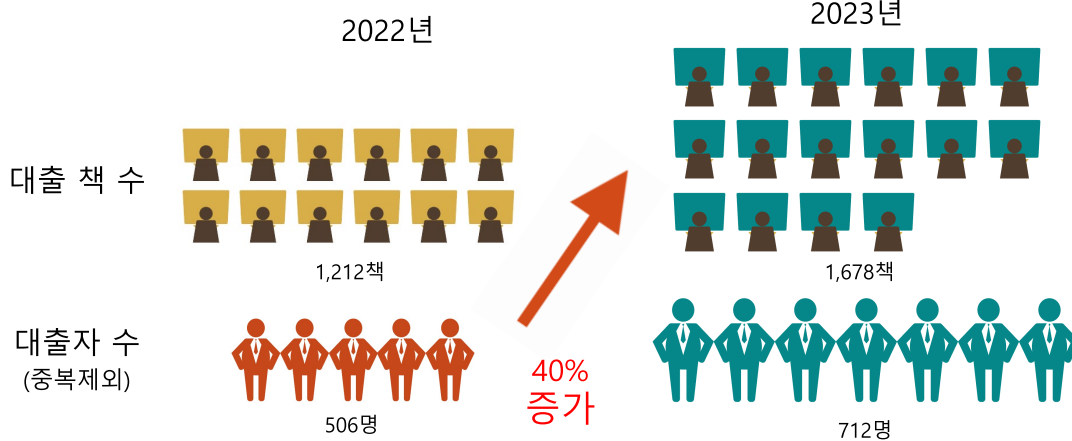
중간성과 보고

LIKE SNU

기준일 2023. 3.(시스템 오픈 후 1개월)

- 총 접속 : 13,033건
- 구성원 접속자 수 : 4,814명 (중복제외)

신입생 도서 대출



가

LIKE SNU 란

- 로그인
- 메인화면
- WHAT IS
- 지식지도
- 지식망원경
- 지식나침반
- 최초 이용 시

서울대학교
SEOUL NATIONAL UNIVERSITY

Why 'LIKE SNU'?

선배 서울대인의 지식 여행 기록은 나의 여행의 네비게이션이 되고
나의 여정은 다시 후배 서울대인의 안내 지도가 됩니다

01 지식 여행을 위한 네비게이터

선배들이 남겨둔 지식 활동의 흔적을 따라가면 어느새 뚜렷한 목표가 생긴 나를 발견할 수 있습니다

02 나의 지식 여정 공유

나의 지식 활동을 다른 사람들에게 공유하여 누군가의 지식 멘토가 될 수 있습니다

03 나만을 위한 맞춤형 지식 콘텐츠

나와 서울대인 전체의 콘텐츠 이용 패턴을 기반으로 나만을 위한 맞춤형 콘텐츠를 제공합니다

7

가

LIKE SNU 란

- 로그인
- 메인화면
- WHAT IS
- 지식지도
- 지식망원경
- 지식나침반
- 최초 이용 시

서울대학교
SEOUL NATIONAL UNIVERSITY

LIKE SNU! 어떻게 시작하나요?

- ▶ 서울대학교 중앙도서관 홈페이지(<https://lib.snu.ac.kr/>)에 접속
- ▶ 로그인 후
- ▶ 마이 라이브러리 > LIKE SNU 를 선택



8

가

LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용 시

9

가

LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용 시

10



LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용 시



● 지식 지도

어떤 책을 보기 전, 후에 읽으면 좋은 책에 대한 독서 가이드를 받을 수 있고, 최근 서울대인의 관심이 집중되는 분야와 연계 분야를 확인할 수 있습니다

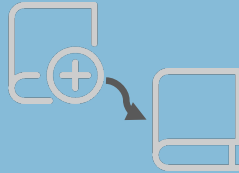
01 분야별 도서 연관 지도

분야별 인기 도서 500권이 만드는 지식 지도를 마치 우주를 여행하는 느낌으로 탐색할 수 있습니다.



02 연관 도서 지도

관심 도서를 읽기 전과 읽은 후에 어떤 책을 읽어야 할지 궁금하다면, 연관 도서 지도에서 확인할 수 있습니다.



03 주목할만한 분야 지도

요즘 서울대인의 관심 분야, 나의 관심 분야의 주제, 나의 관심 분야의 주제, 분야 간의 연계, 융합 현황 등을 이해할 수 있습니다.



LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

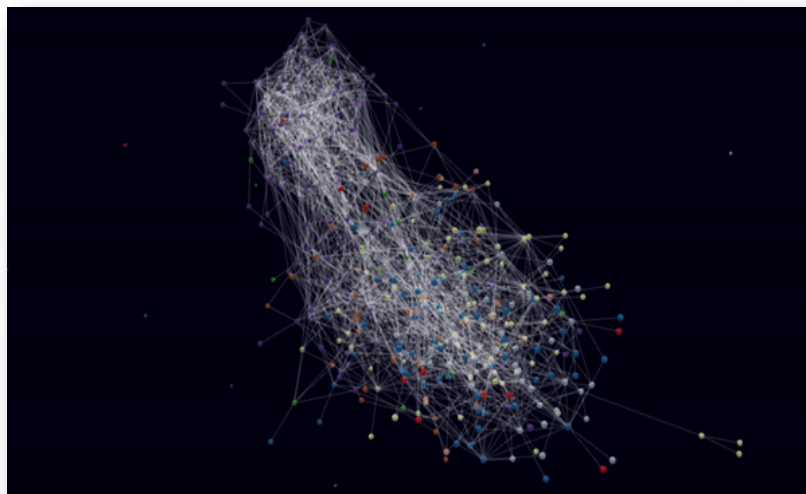
최초 이용 시



지식 지도 내가 관심 있는 분야와 연관된 서울대 구성원의 지식 트렌드를 확인해 보세요

분야별 인기 도서 500선

분야별로 대줄이 많은 500권의 도서 연관 지도를 탐험합니다.



LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용 시

서울대학교
SEOUL NATIONAL UNIVERSITY

지식 지도 — 내가 관심 있는 분야와 연관된 시중에 구성원의 지식 트렌드를 확인에 보세요

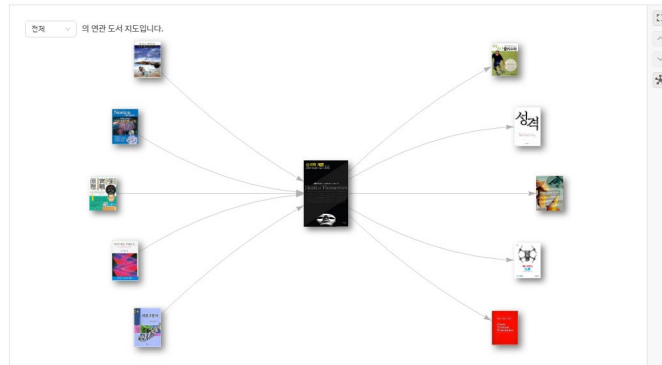
최근 대출 도서

도서관장이 가장 최근 대출하신 도서입니다.



심리학 개론의 연관 도서 지도

이 도서를 읽기 전에 읽은 후에 어떤 도서를 읽는지 확인해보세요.



LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

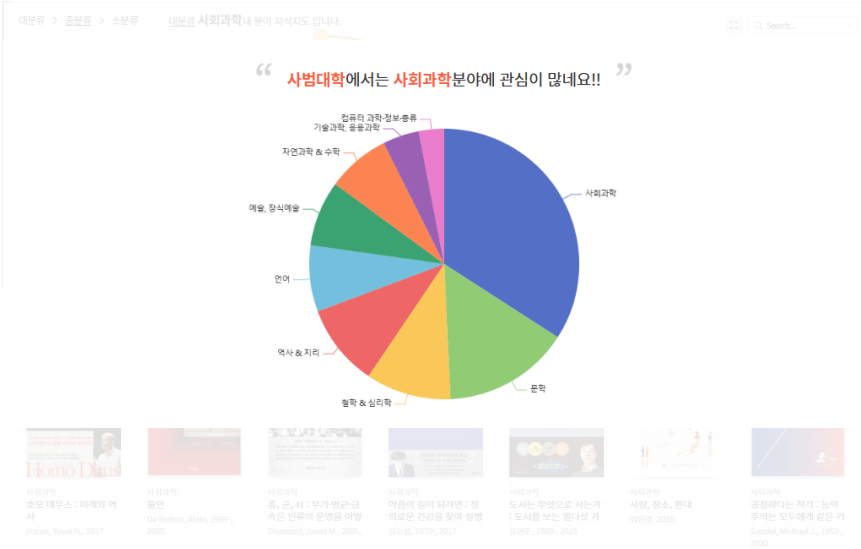
최초 이용 시

서울대학교
SEOUL NATIONAL UNIVERSITY

지식 지도 — 내가 관심 있는 분야와 연관된 시중에 구성원의 지식 트렌드를 확인에 보세요

사범대학에서 관심있는 분야

최근 일주일간 사범대학에서 가장 대출이 많은 분야입니다.



LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용 시



● 지식 망원경

나의 독서 취향에 딱 맞는 도서, 최근 발린 도서와 유사한 도서,
전공이 같은 사람들의 최근 관심 도서 등
다양한 방식으로 추천된 도서를 확인할 수 있습니다


01 맞춤 추천 도서

나의 도서 대출 이력과 서울대인의 대출 이력을 종합적으로 고려하여 나의 관심사에 딱 맞는 도서를 추천합니다.




02 나와 비슷한 사람들 도서

나와 같은 전공인 친구들은 어떤 도서를 많이 읽고 있는지 알 수 있습니다. 또한, 다른 전공, 교수님들의 관심 도서 등도 확인할 수 있습니다.



03 관심 분야의 도서

내가 관심 있는 분야에서는 어떤 도서가 주목 받고 있는지 새로 나온 도서는 무엇인지 알 수 있습니다.



LIKE SNU 란

로그인

메인화면


WHAT IS

지식지도

지식망원경

지식나침반

최초 이용 시










서울대 학생들이 많이 읽은 도서








서울대 대학원생들이 많이 읽은 도서

서울대 교수님들이 많이 읽은 도서








서울대 학생들이 많이 읽고 있는 도서
최근 일주일간 학부생이 가장 많이 대출된 도서입니다.

 서태욱 호모 데우스 : 미래의 역사 Harari, Yuval N., 2017	 서태욱 북 리뷰 2022-2023 이국재, 권위 이국재, 권위, 2021	 서태욱 GPU의 모든 것 장지현 Koselleck, 2009	 서태욱 AI의 모든 것 : 지식, 배드, 영웅, 600-원대 유진광 Person, Karl Gunnar, 1943, 2016	 문석 최후의 강연 : 학생 연극 소설 Person, Karl Gunnar, 1943, 2016	 서태욱 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 서태욱 책 읽는 습관 : 30년간 배운 '서적의 국가' 요한 안드 Ansd, Bohman, 2009
---	--	---	---	---	--	--

서울대 대학원생들이 많이 읽고 있는 도서
최근 일주일간 대학원생이 가장 많이 대출된 도서입니다.

 서태욱 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 서태욱 교육다움 : 이론과 교육의 관계 이국재, 권위 이국재, 권위, 2019	 서태욱 교육다움 : 이론과 교육의 관계 이국재, 권위 이국재, 권위, 2019	 서태욱 이와 사회 : 사회학의 이론과 실천 Harari, Yuval N., 2015	 문석 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 문석 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 서태욱 책 읽는 습관 : 30년간 배운 '서적의 국가' 요한 안드 Ansd, Bohman, 2009
--	---	---	---	---	---	--

서울대 교수님들이 많이 읽고 있는 도서
최근 일주일간 교수님이 가장 많이 대출된 도서입니다.

 문석 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 서태욱 교육다움 : 이론과 교육의 관계 이국재, 권위 이국재, 권위, 2019	 문석 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 서태욱 이와 사회 : 사회학의 이론과 실천 Harari, Yuval N., 2015	 문석 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 문석 최종 결정자의 삶에 얽힌 이야기 Buchholz, Todd C., 2009	 서태욱 책 읽는 습관 : 30년간 배운 '서적의 국가' 요한 안드 Ansd, Bohman, 2009
---	---	---	---	---	---	--

LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용시



● 지식 나침반

현재까지의 나의 지식 좌표를 확인할 수 있습니다

지식 나침반 — 서울대학교에서 나의 지식 좌표를 한 눈에 확인해 보세요.

나의 총 대출권수

29 권

주제별 대출 통계



향후, 서울대인과 나의 독서 지표 비교,
취향 유사 그룹과의 비교 등 기능 확대

17

LIKE SNU 란

로그인

메인화면

WHAT IS

지식지도

지식망원경

지식나침반

최초 이용시




입력한 취향을 활용해 지식 지도에서 관심 분야 표시, 관심 분야 도서 추천

사용자님의 관심사를 분석하여 꼭 맞는 도서를 추천드립니다.

01 · 02 · 03

재미있게 읽은 책 또는 읽고 싶은 책을 5권 이상 선택해주세요.




나의 관심 분야 선택







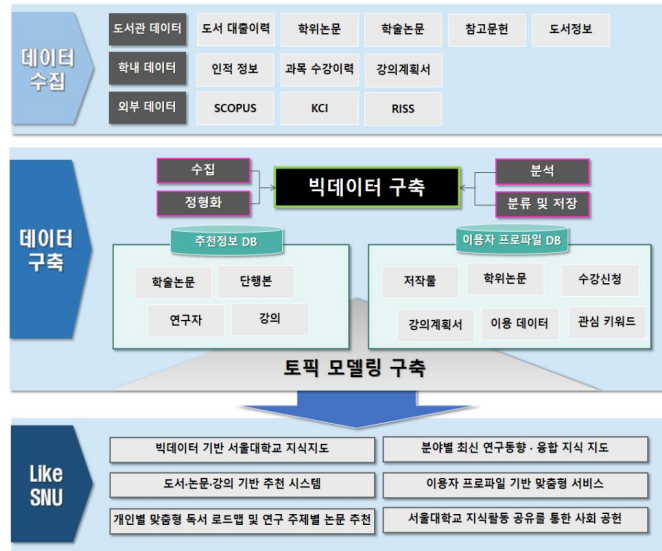
이전단계

추천 결과 보기

18

향후계획(2차년도 사업 추진 중)

데이터범위를 도서에서 학내 생산된 모든 지식데이터로 확장하고
토픽모델과 다양한 부가기능 구축



향후계획(2차년도 사업 추진 중)

○ 학내/외 데이터를 수집·연계하여 빅데이터 플랫폼 서비스 구축

- 도서 대출 이력, 학위논문, 학술논문, 학사데이터 등 데이터 정제를 통한 분석 수행
- 참고문헌, 피인용 데이터를 수집 및 분석하여 인용 선후관계 파악
- 키워드 통제 및 데이터 연동 주기 설정

○ 서울대학교 지식지도를 통해 학내 트렌드 및 관계도 제공

- 독서 주제 및 수강 과목을 활용한 독서 경로 제시
- 학내 인기 연구 토픽 및 토픽별 연구동향 파악
- 인용/피인용 기반 전후관계를 파악함으로써 다인용자료 리스트 제공
- 토픽별 시계열 연구 변화 추이 제공 및 연구트렌드 예측
- 교과목 정보를 활용하여 학제 간 융합 대상 지식 클러스터 도출



향후계획(2차년도 사업 추진 중)

○ 이용자 프로파일 관리를 통한 맞춤형 서비스 제공

- 저작물(단행본, 논문), 수강내역, 강의계획서, 대출 이력, 도서관 이용 로그 등 이용자별 프로파일 생성 및 관리
- 프로파일 기반 유사도가 높은 논문, 단행본, 연구자, 토픽, 교과목에 추천 알고리즘을 적용하고 추천 근거 제공
- 도서 속성 및 과목수강 정보를 기반으로 한 도서 추천 기능 확대
- 주기적으로 추가되는 정보를 반영하여 업데이트 내용 반영

○ 통합 토픽 모델 구축

- 다양한 데이터에 적용할 수 있는 통합 토픽 분류체계 및 네트워크를 구축하여 빅데이터 통합 제공의 기틀 마련
- 향후 데이터 추가·통합 및 서비스 확장의 기준 마련



향후계획(2차년도 사업 추진 중)

○ 최신 시각화 모델 구현 및 플랫폼 확장

- 1차 사업 시 구축된 시각화 모델의 최신화 및 최신 모델 적용
- 향후 분석모델, 분석지표, 분석데이터 확장성을 고려한 시스템 구축
- 개인의 지식활동을 통합한 SNU Collection 생성 및 공유 기능
- 학외 이용자(일반회원)를 위한 별도 프로파일 구축 및 추천 서비스 제공

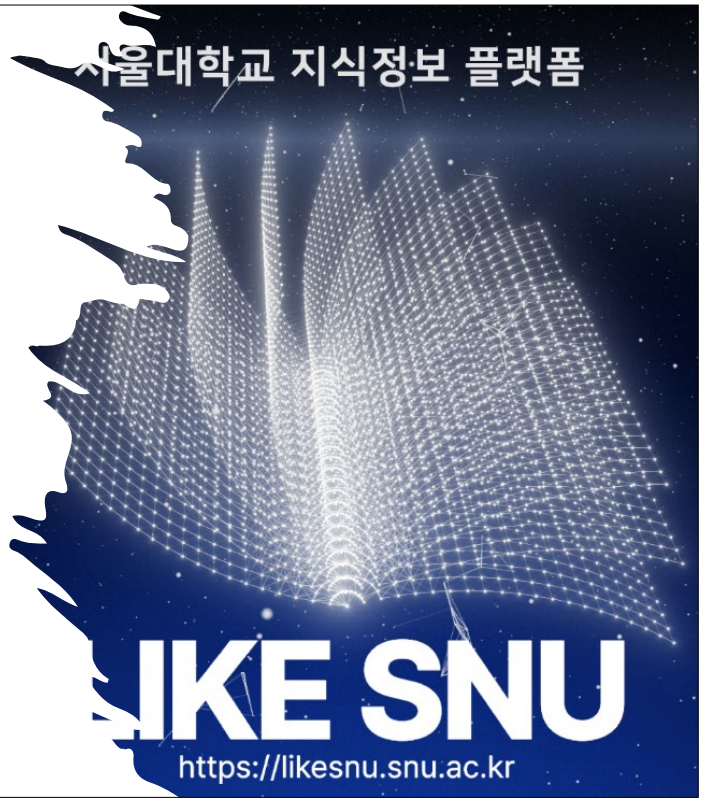
○ 인터페이스 구성

- 기존 빅데이터 플랫폼에 신규 서비스 연계하여 구축
- 마이페이지 구현을 통한 개인화서비스(메일링, 푸시) 제공
- 관리자페이지 구현을 통한 도서/논문 통계 및 이용자피드백 확인

서울대학교 지식정보 플랫폼

대답할 수 있는 질문들

- 무슨 책을/논문을 읽을까?
- 어려운데... 무슨 책을/논문을 미리 읽으면 좋을까?
- 다음엔 무슨 책을/논문을?
- 수업 들으면서 읽으면 더 좋을텐데... 무슨 수업?
- 문과 학생인데 인공지능 분야와 융합하려면... 경로가?
- 나랑 비슷한 관심을 가진 친구들은 어디 있을까? 전공이 다르면 더 좋을텐데
- 융합 과목을 개발해야겠는데... 융합적 지식 클러스터?
- 학부기초대학/첨단융합학부에서는 뭘 가르쳐야 하지?
- 전공설계... 데이터에 기반해서?
- 혁신적 연구의 잭팟은 어디에?
- 교육의 새로운 방향. 기존 학문을 잘게 쪼개서 다시 조립?
- 도서관의 글로벌 경쟁... 어떻게?
- 서울대의 사회공헌... 돈 말고 지식으로. 혼자 말고 다 같이?



서울대답게 서울대처럼

LikeSNU 2023 ALA 국제혁신상 수상 기념

대학도서관 혁신 컨퍼런스

학술대회 발표자료집

02

대학의 지식 빅데이터와 활용 방향

／
(주)사이람
배수진 차장



대학 지식 빅데이터와 활용 방향

- 대학도서관 혁신 컨퍼런스 -

2023.7.6

(주) 사이람 배수진 차장

CONTENTS

1. 대학 도서관의 새로운 역할
2. 지식 빅데이터의 구성과 구조
3. 지식 빅데이터 활용 방향
4. 빅데이터 기반 지식 정보 서비스 구축 사례
5. 기대효과

1 대학 도서관의 새로운 역할

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

능동적 지식 정보 서비스 제공
 (Passive Information Service ▶ Proactive Knowledge Service)
 지식의 생산, 유통, 소비를 지능적으로 지원하는 허브로의 도서관 위상/역할 재정립
 (Improving the Quality of Research Support)

지식 리더십 제시
 대학의 연구 혁신 전략 방향성 도출을 위한
 Best Practice 제시



글로벌 위상 제고
 국내외 지식 생태계(Knowledge Ecosystem)과의
 지식 상호작용 및 협업 주도

2 지식 빅데이터의 구성과 구조

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

- ▶ 대학의 지식 빅데이터는 지식을 생산/소비하는 대학 구성원과 그 대상인 콘텐츠 간의 '행위'로 구성
- ▶ 대학 구성원이 지식 콘텐츠 및 서비스를 생산/이용하는 행위는 '지식 활동'으로 정의할 수 있음

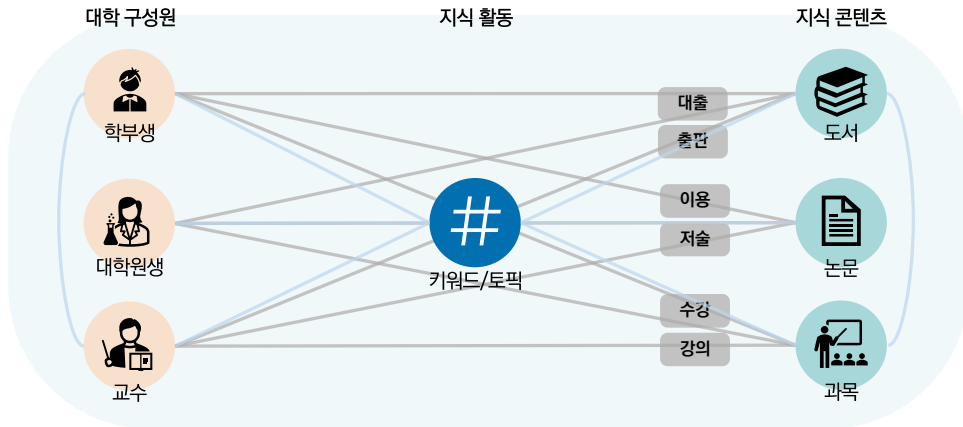


3 지식 빅데이터 활용 방향

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

01 지식 요소 간 연관/연결 구조 구축

- ▶ 대학 구성원과 지식 콘텐츠는 '지식 활동'을 매개로 연결됨
- ▶ 또한 모두 '연관/관심 키워드(토픽)'로 연결할 수 있어, 효과적으로 지식을 탐색할 수 있음



3 지식 빅데이터 활용 방향

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

02 지식 트렌드 분석

- ▶ 지식 콘텐츠의 생산 및 소비 패턴을 통해 학습 및 연구의 트렌드를 이해
- ▶ 특히, 지식 분야 간 융합/연계 현황을 파악함으로써, 향후 학습/연구/강의의 설계 방향을 설정하는데 참조

분야별 관심도 분석

- 도서, 논문, 과목의 생산 및 소비 빈도와 시기를 기반으로 핫토픽을 파악
- 토픽의 시계열적 변화 추이를 토대로 유망 토픽을 예측

국내외의 트렌드 비교

- 글로벌 지식 정보 데이터(예, 논문 서지 정보)를 활용하면, 국내외 토픽 관심도를 표준화하여 비교할 수 있음
- 글로벌 대비 연구 경쟁력을 파악

*출처 https://www.oas.gop/eng/ox/education/annual_report/2019/white_paper_summary_05.html

분야 융합/연계 현황 분석

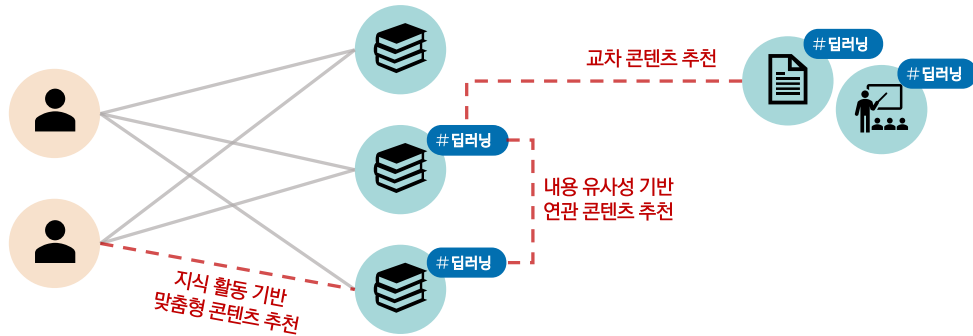
- 이용자 지식 활동을 통해 서로 다른 토픽이 융합되는 현황을 파악
- 연관 토픽을 추천하거나 시각화 함으로써, 융합 연구, 융합 과정 설계 등에 활용

3 지식 빅데이터 활용 방향

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

03 지식 콘텐츠 추천

- ▶ 개인 맞춤형 콘텐츠 추천: 대학 구성원의 '지식 활동'을 기반으로 개인 맞춤형 콘텐츠 추천
- ▶ 연관 콘텐츠 추천: 콘텐츠의 내용 유사성을 기반으로 연관 콘텐츠를 추천
- ▶ 교차 콘텐츠 추천: 도서-논문-과목 간의 연관성을 기반으로 다른 형태의 지식 콘텐츠를 추천



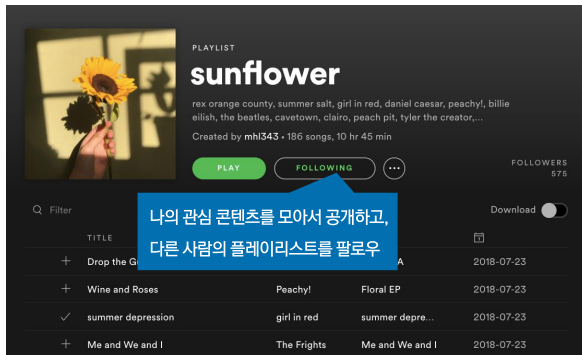
3 지식 빅데이터 활용 방향

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

04 지식 공유 및 지식 매개 상호작용

- ▶ 대학 내에 발생하는 지식 활동을 공유함으로써, 과거-현재, 개인-전체로 지식이 확산될 수 있음
- ▶ 또한, 지식 공유를 통해 구성원 간 상호작용 활성화를 촉진하여, 데이터에 근거한 학내 활동 지원 가능

〈예. Spotify 플레이리스트 기능〉



〈예. 왓챠 나의 취향 분석〉

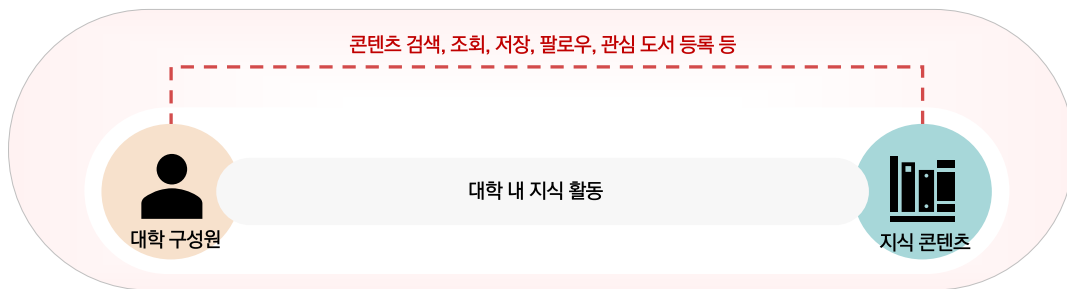


3 지식 빅데이터 활용 방향

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

05 지식 서비스의 로그 데이터를 통한 지식 빅데이터 확장

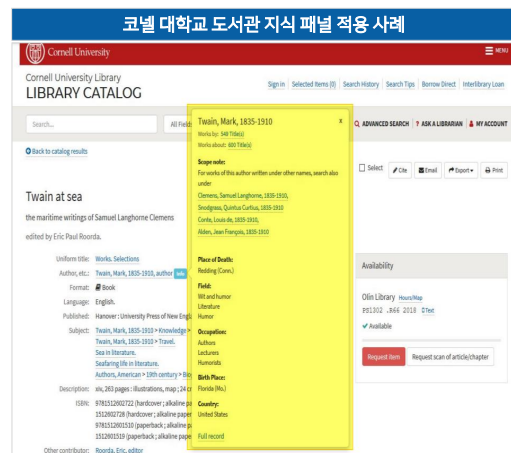
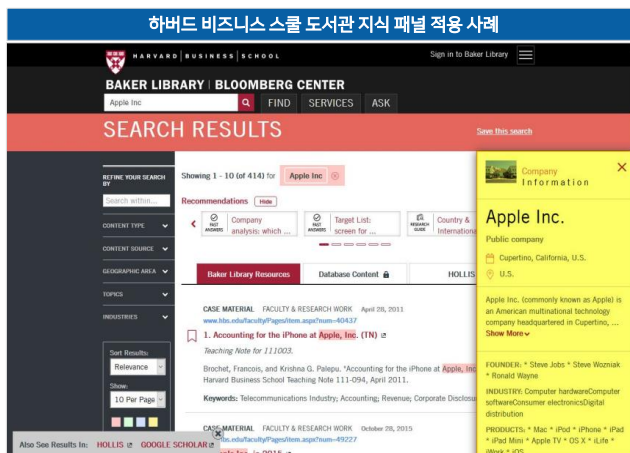
- ▶ 지식 빅데이터를 활용한 지식 정보 서비스를 통해 다양한 서비스 로그를 추가로 수집하여 지식 빅데이터의 범위를 확대
- ▶ 예를 들어, 특정 콘텐츠나 토픽에 대한 구성원의 선호/관심도는 콘텐츠 이용 외에도 검색 기록, 콘텐츠 조회 및 저장 등의 사용자 로그를 통해서도 파악할 수 있음



4 빅데이터 기반 지식 정보 서비스 사례

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

- ▶ (미국) 주요 도서관이 참여하는 LD4P(Linked Data for Production) 프로젝트를 통해 도서관에서의 지식 그래프 활용 방향에 대해 다양한 시도를 적용하고 있음

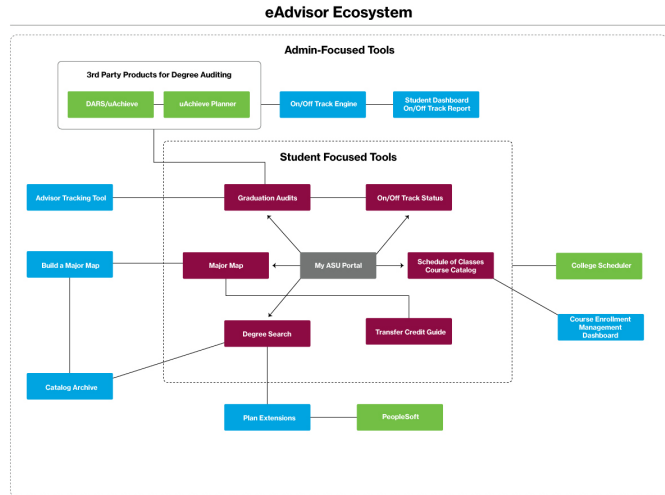
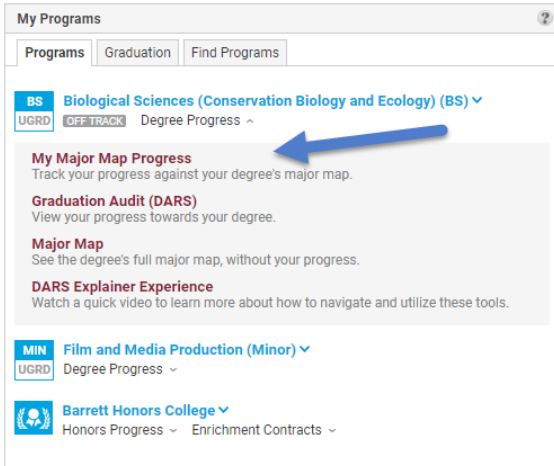


4 빅데이터 기반 지식 정보 서비스 사례

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

▶ Arizona State University의 eAdvisor

- 학생들의 학습 성과를 모니터링 하여, 전공 학습이 미비한 경우 새로운 전공 분야 및 관련 과목 추천

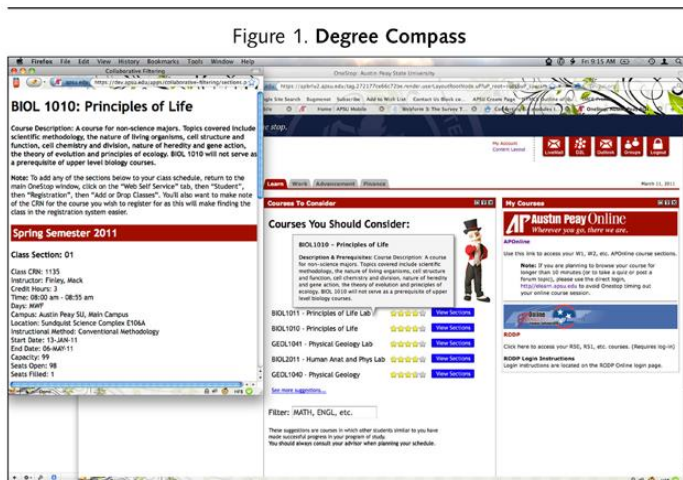


4 빅데이터 기반 지식 정보 서비스 사례

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

▶ Austin Peay State University의 Degree Compass

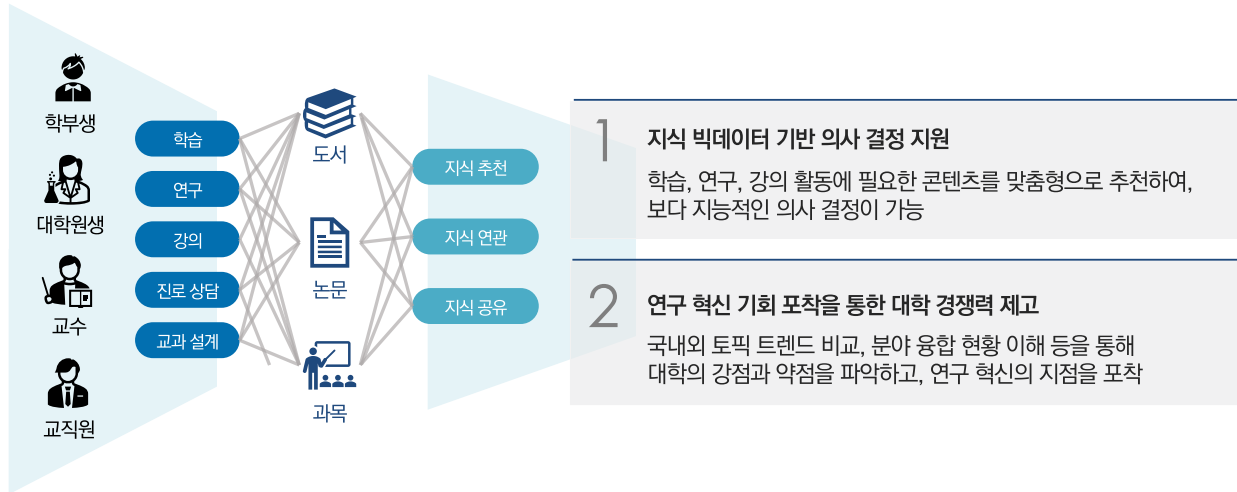
- 전공, 수강 과목 성적 등 학생들의 정보를 활용하여 맞춤형 수강 과목 추천



5 기대 효과

대학 혁신 컨퍼런스: 대학 지식 빅데이터와 활용 방향

지식 주체 지식 수요 지식 콘텐츠 지식 서비스 기대 효과



서울대답게 서울대처럼

LikeSNU 2023 ALA 국제혁신상 수상 기념

대학도서관 혁신 컨퍼런스

학술대회 발표자료집

03

서울대학교 데이터거버넌스 전략과 실행

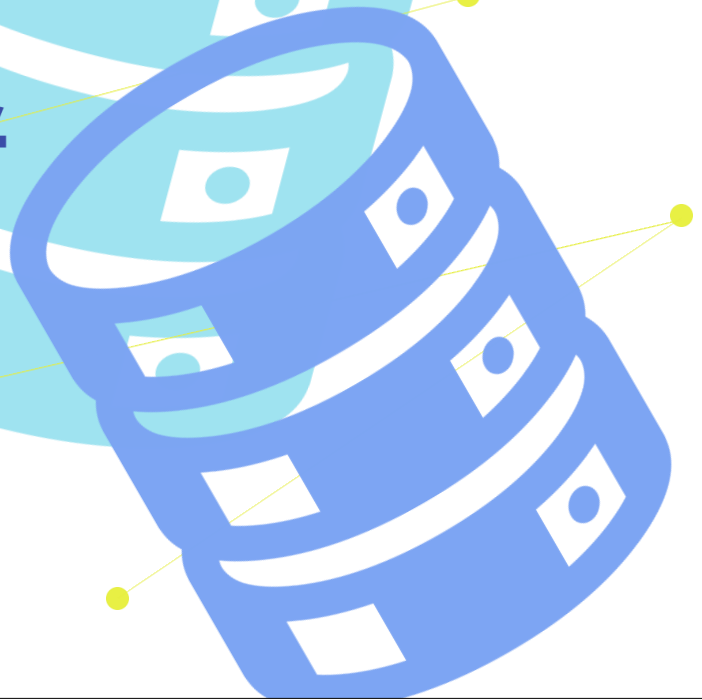
서울대학교 정보화본부
이종표 담당관



서울대학교 데이터거버넌스 전략과 실행

2023. 7.

정보화본부 이종표 담당관



목차

I

서울대학교 데이터거버넌스 전략

1. 그간의 서울대학교 데이터 거버넌스를 위한 노력
2. 기존 노력의 한계
3. 추진 목표 및 과제
4. 데이터기반행정 거버넌스 정립
5. 데이터 통합기반 구축으로 데이터 공동활용 지원
6. 정책활용 및 지능형 서비스 제공을 위한 데이터 분석 활용 지원
7. 데이터 기반활용 역량 강화 및 데이터기반행정 문화 조성

II

서울대학교 데이터거버넌스 실행

1. 추진 체계 구성
2. 서울대학교 데이터 공동이용 활성화 규정 제정
3. 데이터 통합관리 플랫폼 구축 및 활용
4. 공동활용 데이터 수집 및 활용 활성화
5. 데이터를 활용한 정책 및 서비스 창출 강화
6. 서울대학교 데이터 거버넌스 추진 로드맵
7. 서울대학교 데이터 거버넌스 추진 목표

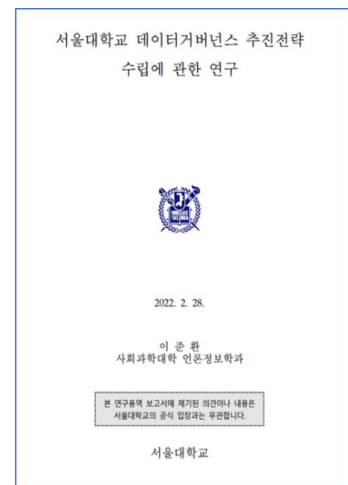
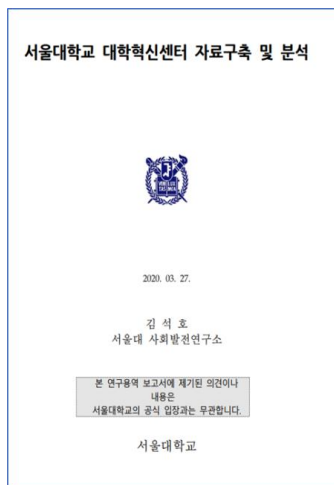
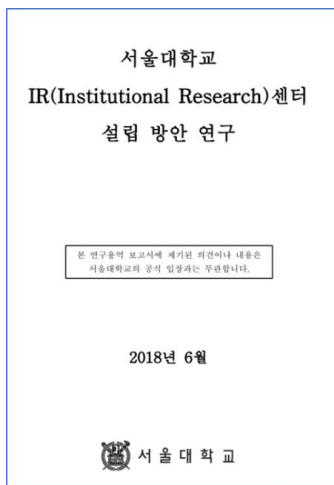
I. 서울대학교 데이터거버넌스 전략

1. 그간의 서울대학교 데이터 거버넌스를 위한 노력
2. 기존 노력의 한계
3. 추진 목표 및 과제
4. 데이터기반행정 거버넌스 정립
5. 데이터 통합기반 구축으로 데이터 공동활용 지원
6. 정책활용 및 지능형 서비스 제공을 위한 데이터 분석 활용 지원
7. 데이터 기반활용 역량 강화 및 데이터기반행정 문화 조성

I. 서울대학교 데이터 거버넌스 전략

1. 그간의 서울대학교 데이터 거버넌스를 위한 노력

정책과제 수행



기타: 빅데이터 관리체계 수립(2022), 온라인 교육 데이터 활용 및 수집방안(2023) 등

I. 서울대학교 데이터 거버넌스 전략

1. 그간의 서울대학교 데이터 거버넌스를 위한 노력

데이터를 활용한 서비스 및 지표 제공

1



서울대학교 통계정보 시스템(2014~)
K-MOOC, SNUon, edX를 통한 지식 콘텐츠

2



스누지니(2021~)
학생중심의 지능형 학사정보 서비스

3



READ LIKE SNU(2022~)
도서관 빅데이터 기반 지식정보 플랫폼

서울대학교 데이터거버넌스 전략과 실행

정보화본부

I. 서울대학교 데이터 거버넌스 전략

1. 그간의 서울대학교 데이터 거버넌스를 위한 노력

데이터 역할 조직 설립 및 운영

정보화본부	대학혁신센터 데이터통합관리부	빅데이터 연구원	모바일 및 빅데이터를 활용한 융합형 보건인재양성 사업단
도시 데이터 사이언스 연구소	데이터사이언스 혁신인재양성 사업단	선도연구진흥센터 데이터지원단	AI 연구원
데이터사이언스 대학원	빅데이터 혁신 공유대학 사업단	빅데이터 혁신 융합대학 사업단	...

서울대학교 데이터거버넌스 전략과 실행

정보화본부

1. 서울대학교 데이터 거버넌스 전략

2. 기존 노력의 한계점

1	빅데이터 전략 수립	데이터를 이해하고, 대학에 적용하려는 인력
2	비표준화 데이터	개별 기관에서 데이터를 관리로 인한 데이터의 표준화 한계로 활용 어려움
3	데이터의 책임주체	데이터에 대한 소유 및 문제 발생시 책임자
4	전담 조직	서울대학교 데이터를 총괄하는 전담 조직 및 인력
5	데이터 공유에 대한 부정적 인식	데이터의 오용 약용 등 우려로 인하여 공유에 부정적
6	기준/규정의 부재	데이터 이용 기준, 활용 등에 대한 규정
7	기관간 데이터 공유 칸막이	기관에서 보유하고 있는 데이터를 타 기관에 미제공
8	활용 가능 데이터 부족	공동 활용을 위한 도출 데이터 부족
9	데이터 리터러시	데이터에 대한 이해 부족

서울대학교 데이터거버넌스 전략과 실행

정보화본부

1. 서울대학교 데이터 거버넌스 전략

3. 추진 목표 및 과제

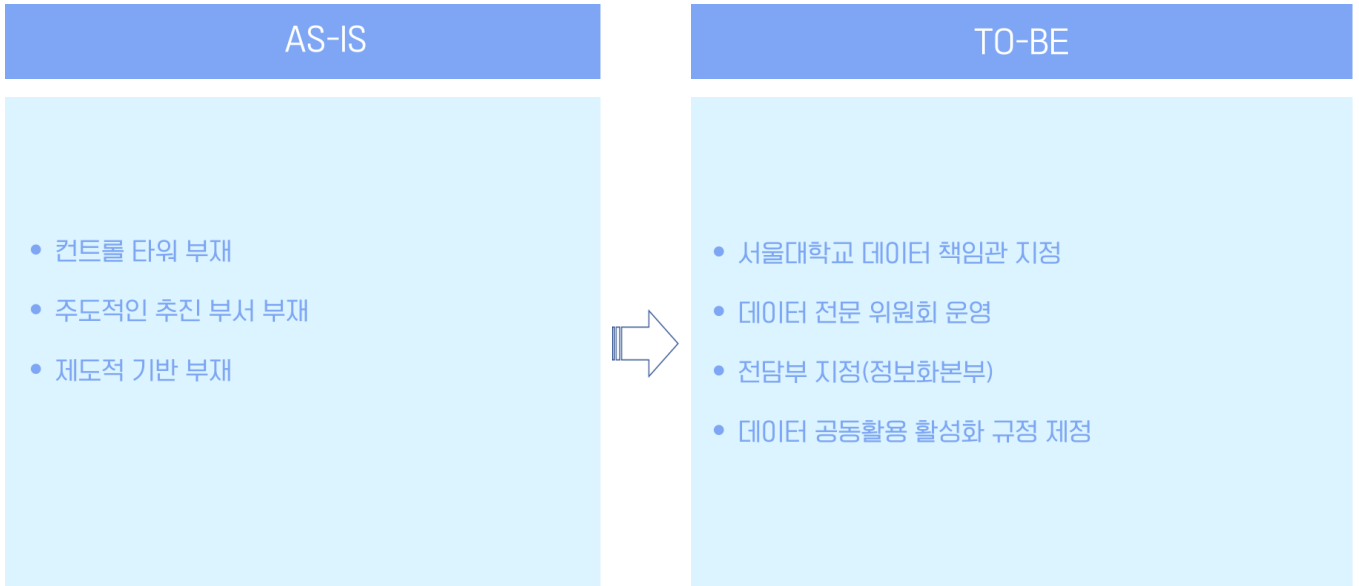
미션	데이터 기반 행정 효율화(Data Centricity) ※ 2023-2025 정보화중합발계획 미션			
추진목표	정책결정에 데이터를 적극 활용하는 과학적 행정 구현		구성원이 신뢰하고 공감하는 지능형 행정서비스 제공	
추진과제	데이터기반행정 거버넌스 정립	데이터 통합 기반 구축으로 데이터 공동활용	정책활용 및 지능 형 서비스 제공을 위한 데이터 분석· 활용 지원	데이터활용 역량 강화 및 데이터기 반행정 문화 조성
세부 추진과제	1. 수행조직및인력구성 2. 제도기반 마련 3. 추진체계 마련 4. 시행계획 수립 5. 실태점검 추진 6. 예산 운용·관리	1. 메타데이터 관리 2. 데이터기본시스템구축 3. 공동활용 데이터 수집·활용 지원 4. 데이터 활용 지원 5. 공동활용 데이터의 안전한 관리	1. 데이터분석과제발굴 2. 데이터분석수행지원 3. 데이터 분석결과 의 정책활용 4. 데이터 분석경험 및 산출물 공유	1. 데이터 역량강화 교육 제공·지원 2. 데이터기반행정 우수사례 발굴·확산 3. 데이터기반행정 활성화 시책 추진

서울대학교 데이터거버넌스 전략과 실행

정보화본부

1. 서울대학교 데이터 거버넌스 전략

4. 데이터기반행정 거버넌스 정립

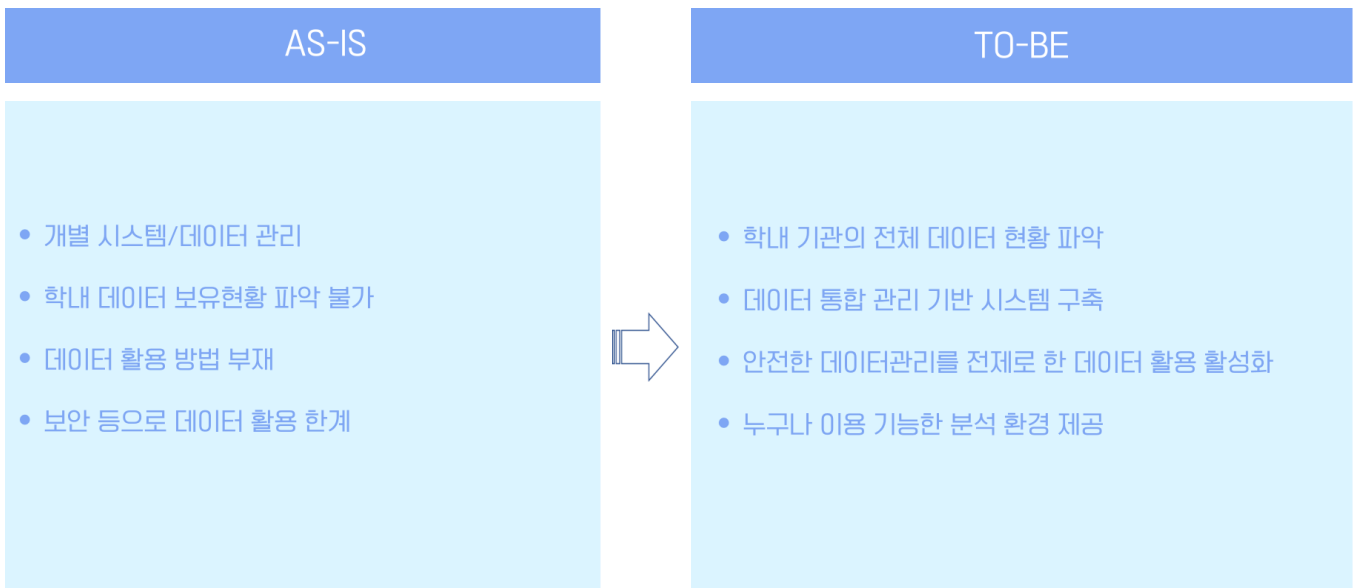


서울대학교 데이터거버넌스 전략과 실행

정보화본부

1. 서울대학교 데이터 거버넌스 전략

5. 데이터 통합기반 구축으로 데이터 공동활용 지원

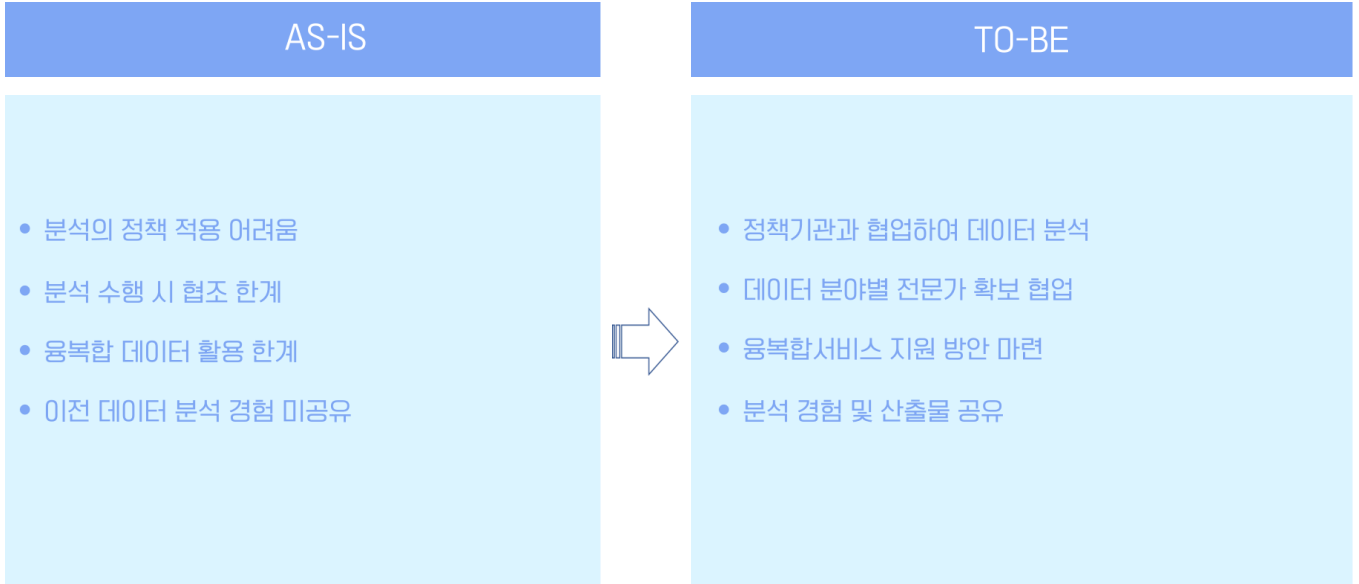


서울대학교 데이터거버넌스 전략과 실행

정보화본부

1. 서울대학교 데이터 거버넌스 전략

6. 정책활용 및 지능형 서비스 제공을 위한 데이터 분석,활용 지원

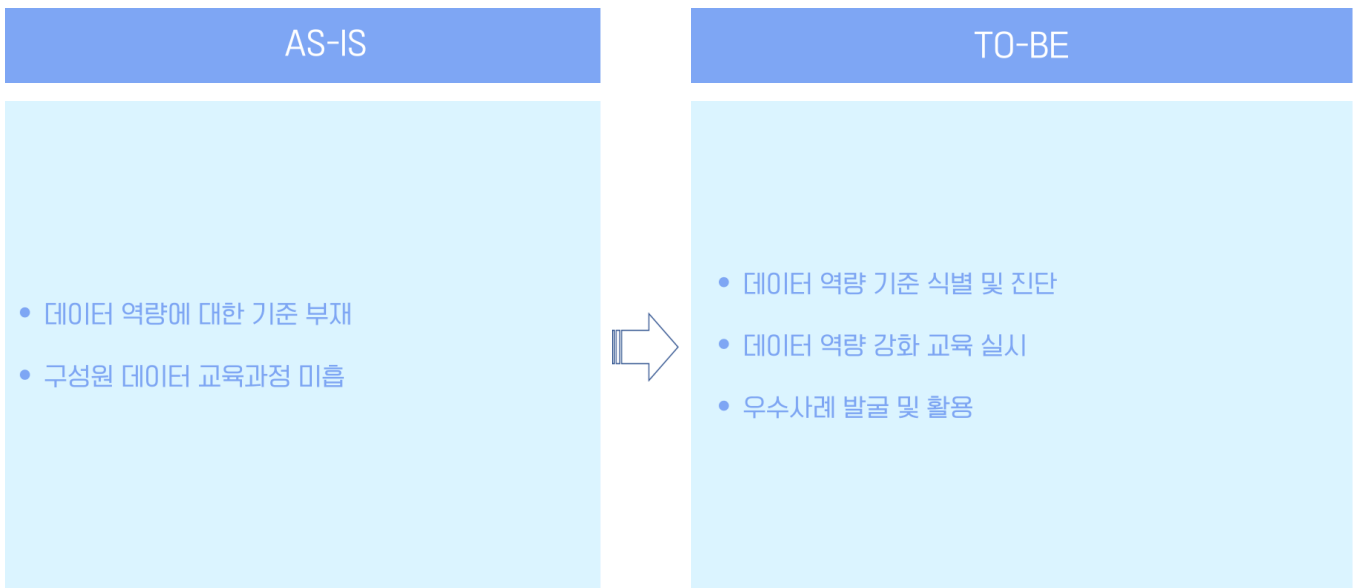


서울대학교 데이터거버넌스 전략과 실행

정보화본부

1. 서울대학교 데이터 거버넌스 전략

7. 데이터 기반 활용 역량 강화 및 데이터기반행정 문화 조성



서울대학교 데이터거버넌스 전략과 실행

정보화본부

II. 서울대학교 데이터거버넌스 실행

1. 추진 체계 구성
2. 서울대학교 데이터 공동이용 활성화 규정 제정
3. 데이터 통합관리 플랫폼 구축 및 활용
4. 공동활용 데이터 수집 및 활용 활성화
5. 데이터를 활용한 정책 및 서비스 창출 강화
6. 서울대학교 데이터 거버넌스 추진 로드맵
7. 서울대학교 데이터 거버넌스 추진 목표

II. 서울대학교 데이터거버넌스 실행

1. 추진 체계 구성

1

전담조직 구성

정보화본부 내 데이터기획 파트 신설 및
전담인력 지정(2023. 7. 1.)

2

데이터 전문위원회 운영 강화

데이터 전문위원회 구성 및 심의 기능 규정화:
'정보화위원회' 규정 개정 추진 중

3

데이터 실무 TF 구성 및 운영

데이터 기획, 분석, 개인정보총괄, 보유부서, 활용 부서
중심으로 정기적으로 의견 교류(2023. 5.~)
※ 정보화본부, 대학혁신센터, 총무과, 학사과, 중앙도서관

II. 서울대학교 데이터거버넌스 실행

2. 서울대학교 데이터 공동이용 활성화 규정 제정(진행중)

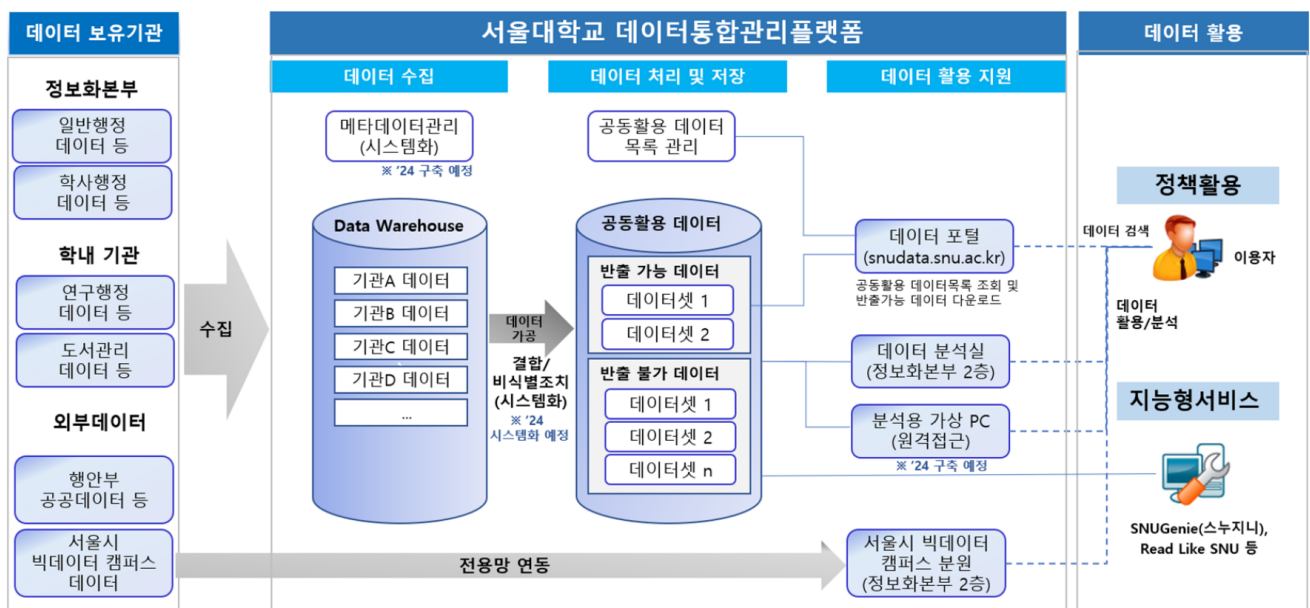
목적 및 기본 원칙	서울대학교 데이터의 공동이용 활성화에 필요한 사항을 정의, 데이터기반의 객관적이고 과학적인 데이터기반행정 지원
데이터 공동이용 활성화 정책 심의·조정	데이터 공동이용 활성화에 관한 사항은 데이터전문위원회에서 심의·조정
공동활용 데이터 발굴 및 수집	(정보화본부) 공동활용 데이터를 발굴하여 데이터 플랫폼에 수집·등록 (데이터보유기관)학내데이터보유기관의공동활용데이터수집협조
이용자의 데이터 이용 절차	데이터 통합관리 플랫폼에 등록된 데이터를 이용하며 등록되지 않은 데이터는 별도 신청
데이터 통합관리 플랫폼 구축 및 운영	데이터의 효율적인 통합관리 및 공동활용을 위하여 데이터통합 관리 플랫폼 구축 및 운영
데이터 책임관 임명	서울대학교 데이터 공동이용 활성화에 관한 업무 총괄을 위하여 정보화본부장을 데이터책임관으로 임명

서울대학교 데이터거버넌스 전략과 실행

정보화본부

II. 서울대학교 데이터거버넌스 실행

3. 데이터 통합관리 플랫폼 구축 및 활용



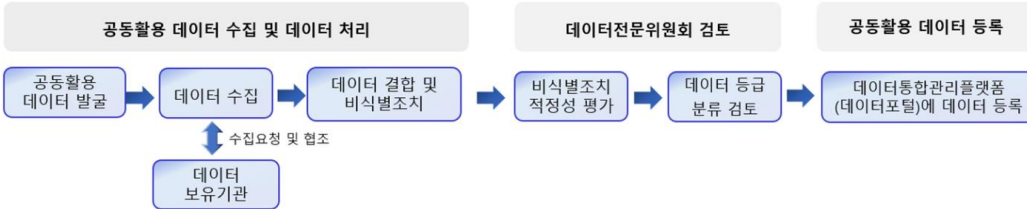
서울대학교 데이터거버넌스 전략과 실행

정보화본부

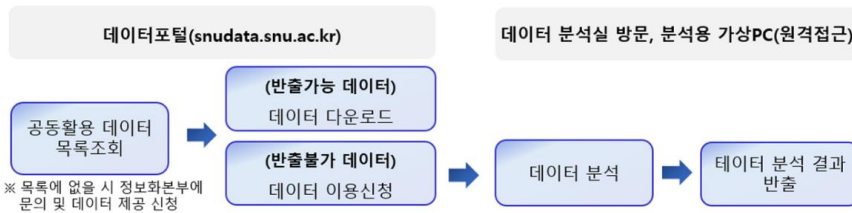
II. 서울대학교 데이터거버넌스 실행

4. 공동활용 데이터 수집 및 활용 활성화

공동활용 가능한 데이터 발굴 수집



이용자의 데이터 활용 절차

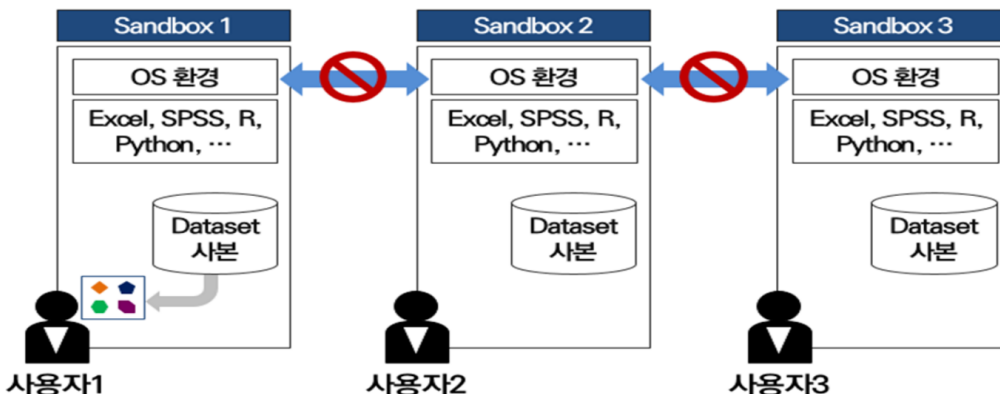


II. 서울대학교 데이터거버넌스 실행

4. 공동활용 데이터 수집 및 활용 활성화

분석지원환경 마련(물리적 공간, 가상pc를 활용)

- 학내에 데이터 활용을 위하여 Sandbox 개념의 정보화본부 내 데이터 분석실 마련



- 서울시 빅데이터 분석 설치(~23. 9.) 및 운영 ※ 서울시 공공데이터를 방문하지 않고 학내에서 사용
- 방문 이용 시 데이터 소개 등 맞춤형 활용 교육 실시

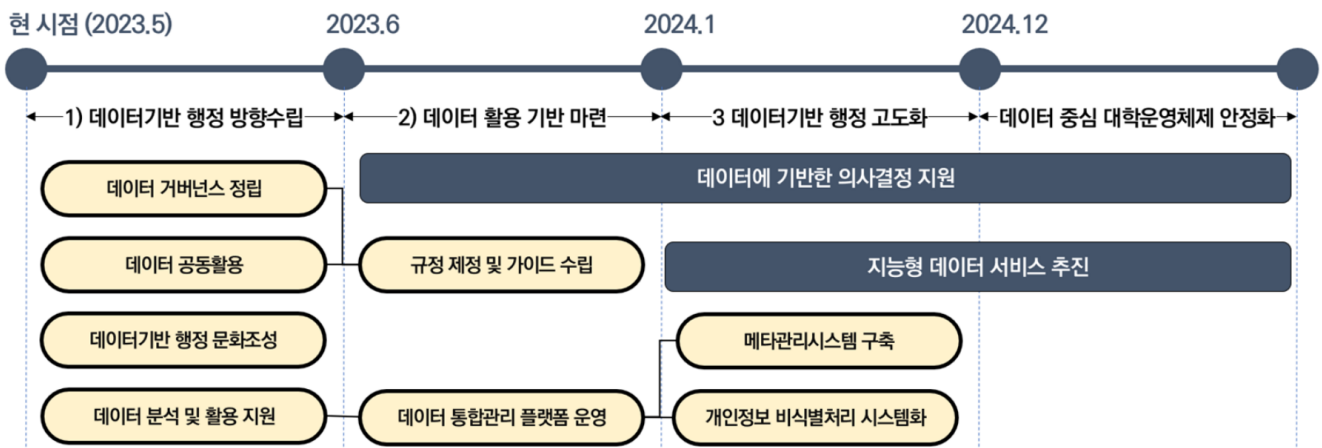
II. 서울대학교 데이터거버넌스 실행

5. 데이터를 활용한 정책 및 서비스 창출 강화



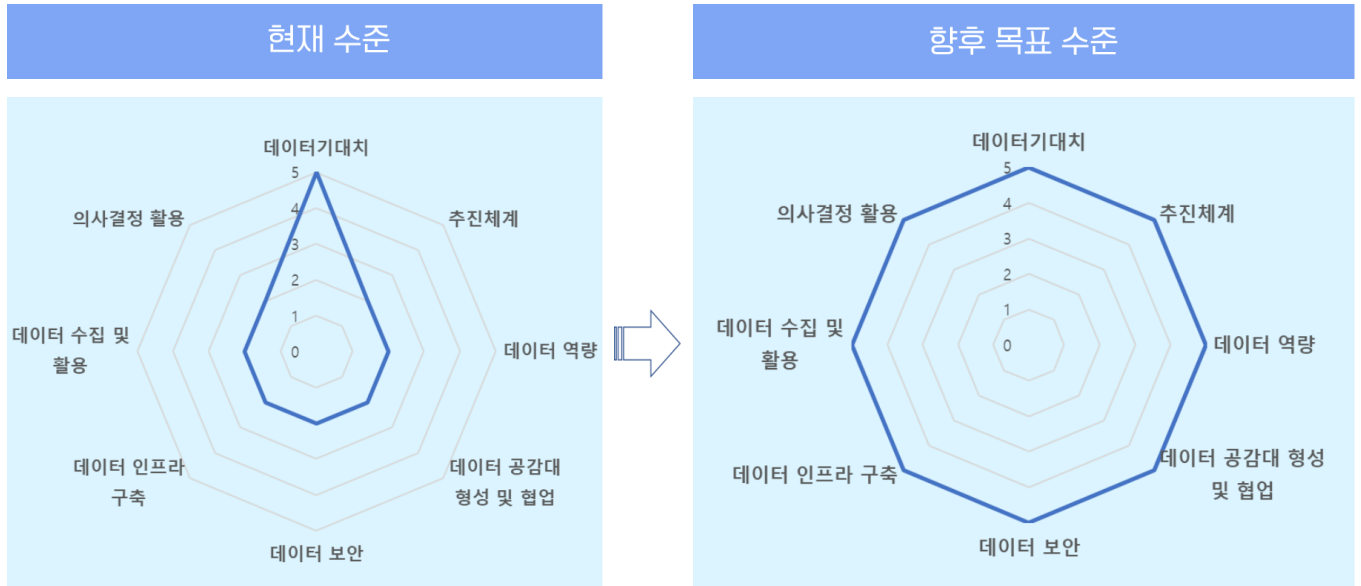
II. 서울대학교 데이터거버넌스 실행

6. 서울대학교 데이터 거버넌스 추진 로드맵



II. 서울대학교 데이터거버넌스 실행

7. 서울대학교 데이터거버넌스 추진 목표



서울대학교 데이터거버넌스 전략과 실행

정보화본부

감사합니다.

서울대학교 데이터거버넌스 전략과 실행

정보화본부

서울대답게 서울대처럼

LikeSNU 2023 ALA 국제혁신상 수상 기념

대학도서관 혁신 컨퍼런스

학술대회 발표자료집

04

서울대학교 중앙도서관 빅데이터 특징과 활용방안

서울대학교 대학혁신센터
서은경 선임연구관



서울대학교 중앙도서관 빅데이터 특징과 활용방안



서울대학교 대학혁신센터

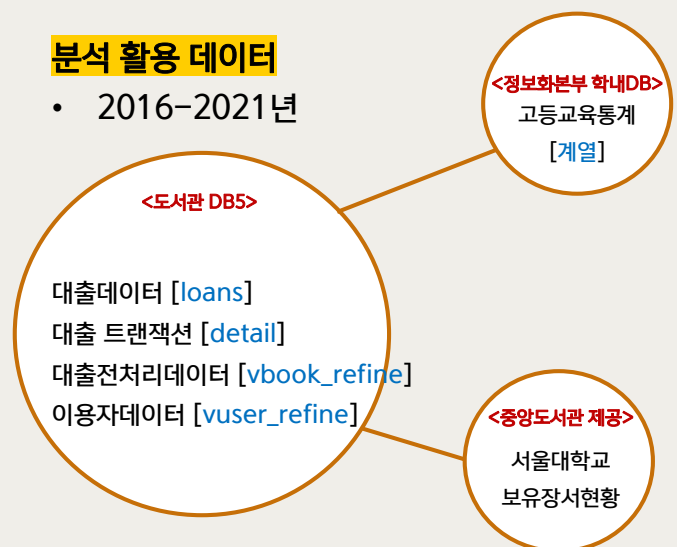
분석 내용

학생들의 단행본 이용행태 분석

- 주제별, 학생 계열별
중앙도서관 소장도서
대출 건수 및 이용정도 분석
- 출판경과연수에 따른 도서 대출
현황 분석

분석 활용 데이터

- 2016-2021년



분석 대상

588,155

2016 - 2021년
 중앙도서관 도서 대출 총 건수

28,645

2016 - 2021년
 학생 이용자 수
 학부생: 15,627명
 대학원생: 13,018명

5

전공계열 수
 인문사회계열: 14,001명
 자연과학계열: 6,050명
 공학계열: 5,847명
 예체능계열: 1,463명
 의학계열: 1,284명



3

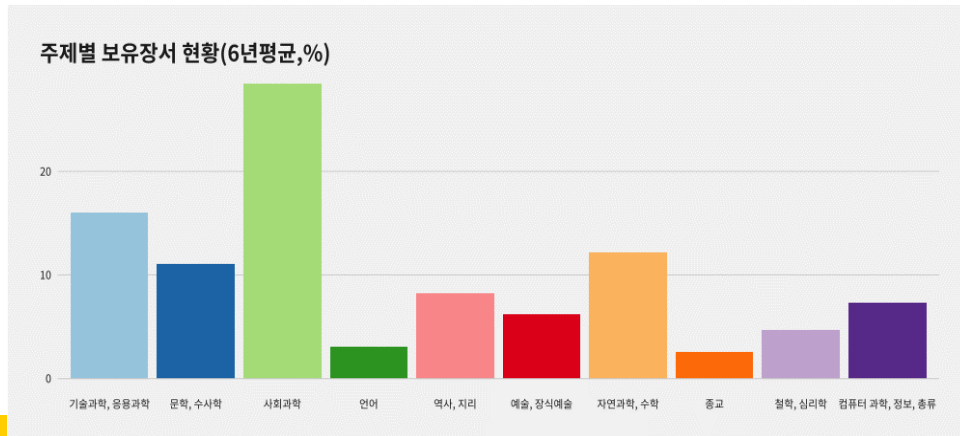
1

주제별 도서대출 현황

서울대학교 도서관 분류체계 대분류 기준 분석

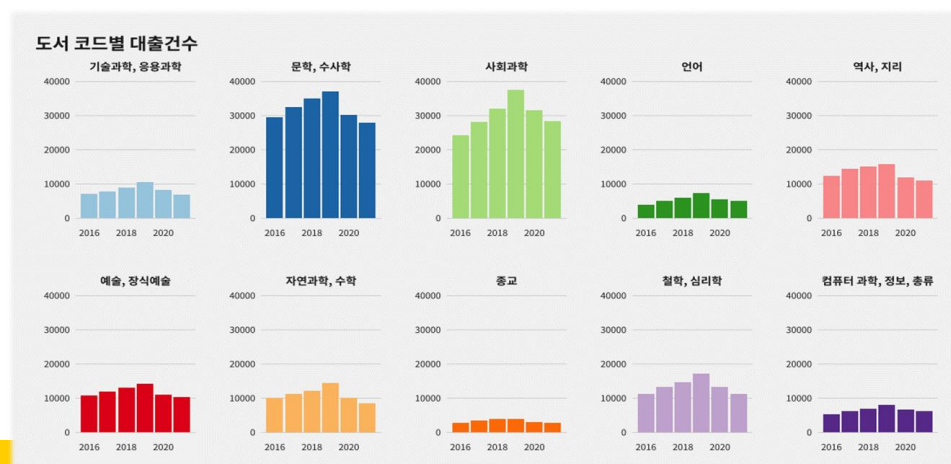
4

주제별 보유도서 현황



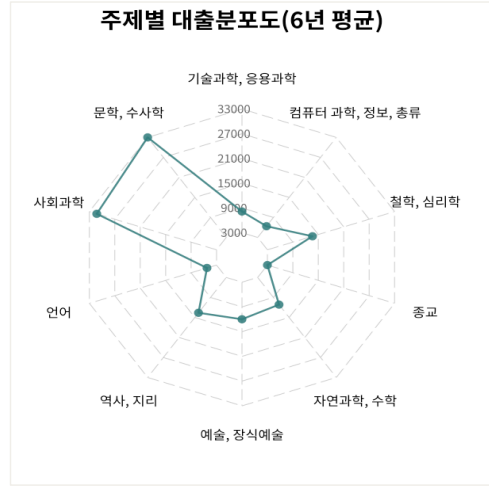
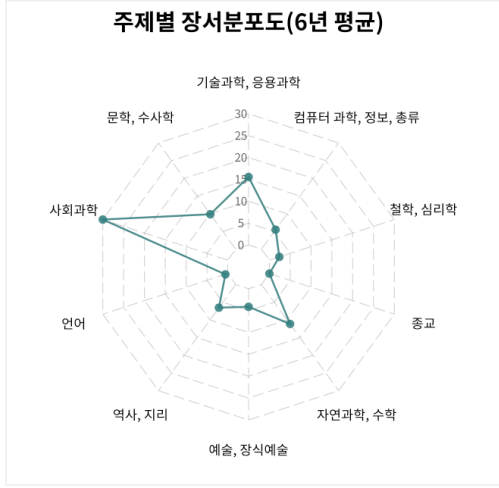
사회과학(28.50%)과 기술과학 및 응용과학(16.03%) 보유 도서가 가장 많으며, 언어(3.10%)와 종교(2.61%)가 가장 적음

주제별 도서대출 현황



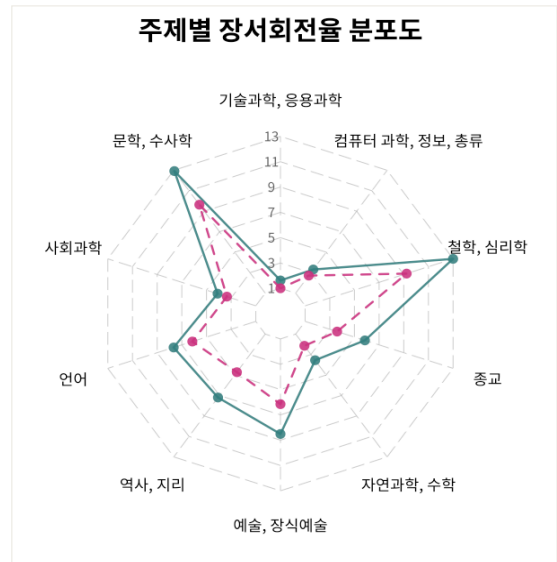
학생 이용자들의 대출건수가 가장 많은 주제는 문학·수사학과 사회과학, 가장 적은 주제는 종교와 언어

주제별 보유도서 vs. 보유도서 대출건수



주제별 소장도서 장서회전율

- 코로나 이전과 이후 시점인 2018년과 2021년을 대상으로 주제별 소장도서 장서회전율 비교:
(자주색: 2018년, 청록색: 2021년)
- 주제별 비율은 비슷하나 코로나19 이후 장서회전율이 전반적으로 증가
- 장서회전율 = 주제별 도서 대출건수 / 주제별 보유도서 수 (값이 클수록 도서 대출수가 많음을 의미)



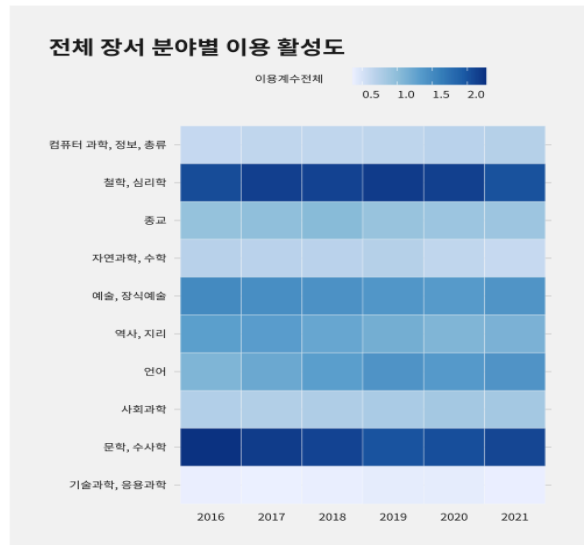
주제별 이용정도 (이용계수 분석)

■ Bonn의 이용계수 (Use Factor)

$$= \frac{\frac{\text{특정 주제 분야 총 대출건수}}{\text{도서관 총 대출건수}} \times 100}{\frac{\text{특정 주제 분야 도서수}}{\text{도서관 총 도서수}} \times 100}$$

이용계수가 1보다 크면 활발히 이용되는 장서

- 가장 활발: 문학·수사학(2.13), 철학·심리학(2.12)
- 가장 저조: 컴퓨터과학·정보·총류(0.66),
기술과학·응용과학(0.38)



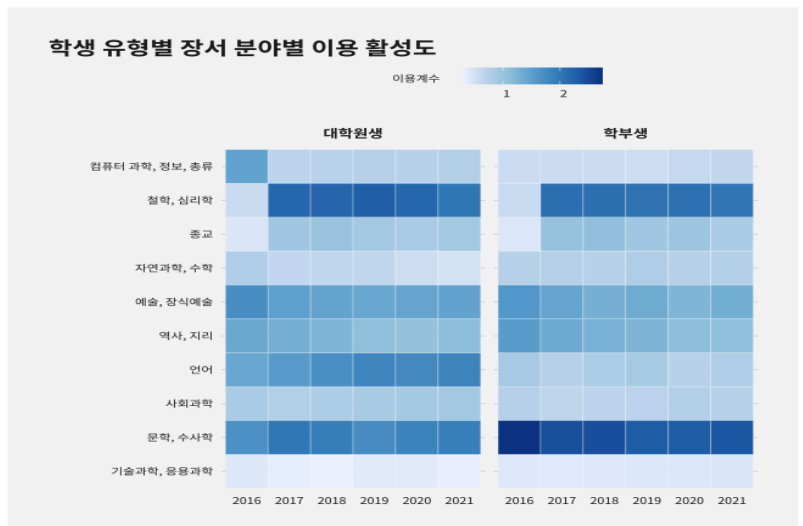
주제별 이용정도 (이용계수 분석)

■ Bonn의 이용계수 (Use Factor)

$$= \frac{\frac{\text{특정 주제 분야 총 대출건수}}{\text{도서관 총 대출건수}} \times 100}{\frac{\text{특정 주제 분야 도서수}}{\text{도서관 총 도서수}} \times 100}$$

이용계수가 1보다 크면 활발히 이용되는 장서

- 가장 활발히 이용 :
(대학원생) 철학·심리학 (6년평균 1.92)
(학부생) 문학·수사학 (6년평균 2.45)



2

전공 계열별 도서대출 현황

5개 전공계열별 도서대출 패턴 분석



계열별 1인당 도서대출 현황

<계열별 1인당 대출건수>

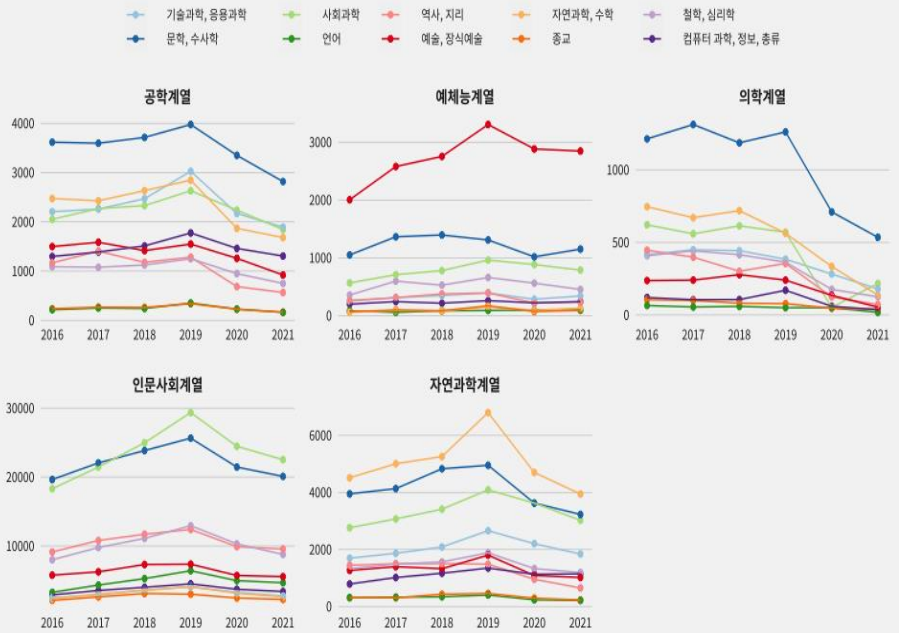
계열별 1인당 대출건수
= $\frac{\text{계열별 총 대출건수}}{\text{계열별 재적학생 수}}$

	인문사회계열	자연과학계열	공학계열	예체능계열	의학계열
2016	5.66	2.23	2.03	2.39	1.74
2017	6.66	2.46	2.12	3.09	1.75
2018	7.56	2.70	2.19	3.19	1.67
2019	8.52	3.18	2.50	3.78	1.63
2020	7.00	2.33	1.87	3.05	0.79
2021	6.44	1.97	1.52	3.04	0.57

주제별 대출건수

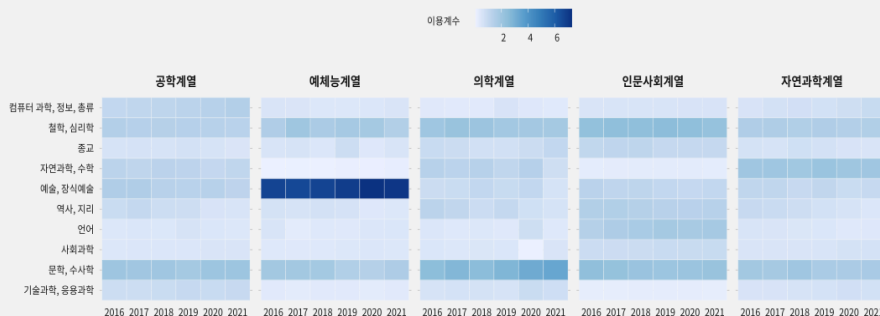
- 예체능계열, 인문사회계열, 자연과학계열은 전공 관련 주제 도서 대출 건수가 높게 나타남
- 공학계열과 의학계열은 문학·수사학 분야 도서를 가장 많이 대출한 것으로 나타남

학생 전공별 도서 코드별 대출건수



주제별 이용정도(이용계수 분석)

학생 전공별 장서 분야별 이용 활성화



- (이용정도가 가장 활발한 주제) 인문사회계열: 철학·심리학(2.40), 자연과학계열: 자연과학·수학(2.02), 공학계열: 문학·수사학(2.01), 예체능계열: 예술·장식예술, 의학계열: 문학·수사학(2.87)
- 예체능계열의 경우 전공 관련 주제인 예술·장식예술 관련 이용도가 매우 활발

계열별 주제편중도 분석 (전체)

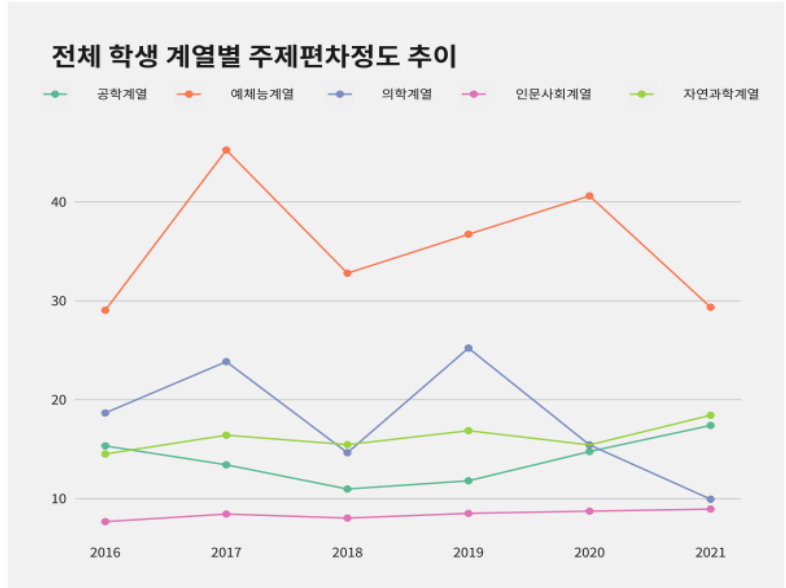
주제 편차정도

$$= \frac{\text{최대 주제편중도}}{\text{최소 주제편중도}}$$

최대(최소)주제편중도

$$= \frac{\text{계열별 최대(최소) 대출 주제 분야 대출건수}}{\text{계열별 총 대출건수}} \times 100$$

주제 편차정도가 작을수록 다양한 주제를
균형적으로 이용함을 의미

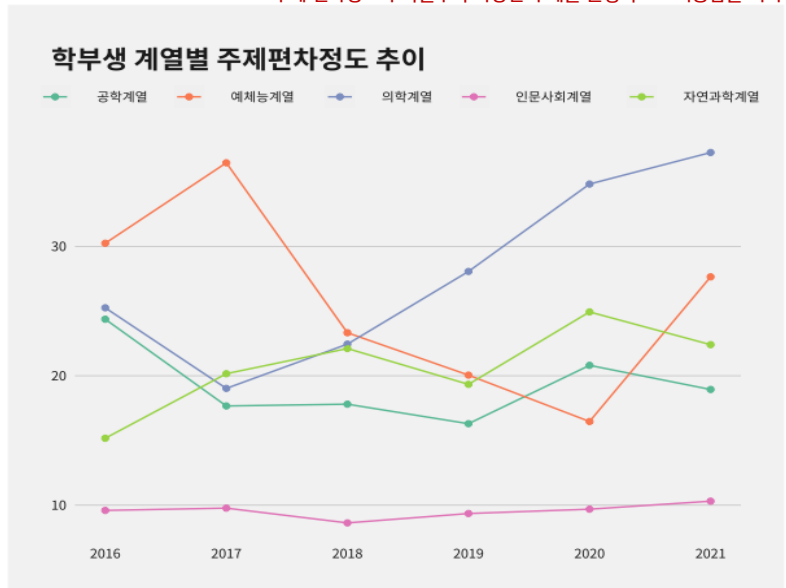


계열별 주제편중도 분석 (학부)

* 주제 편차정도가 작을수록 다양한 주제를 균형적으로 이용함을 의미

<최대-최소 선호 주제 (학부)>

연도	인문사회계열	자연과학계열	공학계열	예체능계열	의학계열	
2016	최대	문학·수사학	문학·수사학	문학·수사학	예술·장식예곡	문학·수사학
	최소	종교	언어	언어	자연과학·수학	종교
2017	최대	문학·수사학	자연과학·수학	문학·수사학	예술·장식예곡	문학·수사학
	최소	언어	언어	언어	언어	언어
2018	최대	문학·수사학	문학·수사학	문학·수사학	예술·장식예곡	문학·수사학
	최소	언어	언어	종교	종교	언어
2019	최대	문학·수사학	자연과학·수학	문학·수사학	예술·장식예곡	문학·수사학
	최소	종교	언어	언어	언어	언어
2020	최대	문학·수사학	자연과학·수학	문학·수사학	예술·장식예곡	문학·수사학
	최소	언어	언어	언어	언어	컴퓨터과학 정보·응용
2021	최대	문학·수사학	자연과학·수학	문학·수사학	예술·장식예곡	문학·수사학
	최소	종교	언어	종교	종교	언어





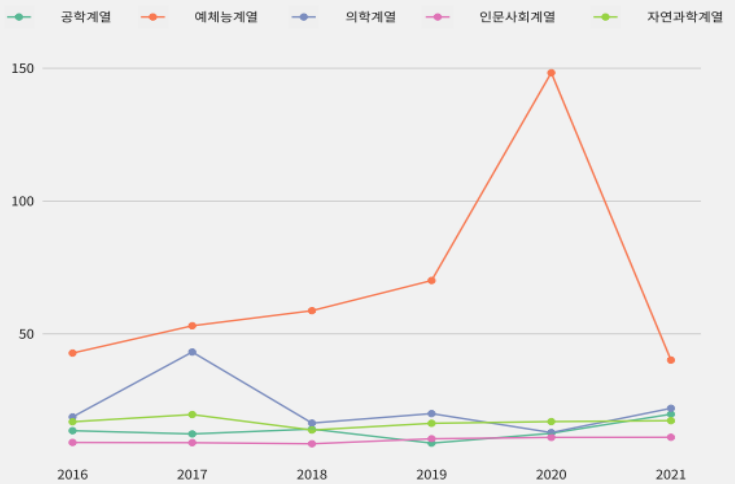
계열별 주제편중도 분석 (대학원)

* 주제 편차정도가 작을수록 다양한 주제를 균형적으로 이용함을 의미

<최대·최소 선호 주제 (대학원)>

		인문사회 계열	자연과학 계열	공학 계열	예술· 체육계열	의학 계열
2016	최대	문학·수사학	자연과학·수학	문학·수사학	예술·장식예술품	문학·수사학
	최소	종교	종교	언어	종교	언어
2017	최대	사회과학	자연과학·수학	문학·수사학	예술·장식예술품	문학·수사학
	최소	기술과학· 응용과학	종교	언어	언어	언어
2018	최대	사회과학	자연과학·수학	문학·수사학	예술·장식예술품	문학·수사학
	최소	종교	종교	언어	언어	언어
2019	최대	사회과학	자연과학·수학	문학·수사학	예술·장식예술품	자연과학·수학
	최소	종교	종교	종교	언어	언어
2020	최대	사회과학	자연과학·수학	문학·수사학	예술·장식예술품	자연과학·수학
	최소	종교	언어	종교	종교	종교
2021	최대	사회과학	사회과학	문학·수사학	예술·장식예술품	문학·수사학
	최소	자연과학· 수학	종교	언어	언어	예술·장식예술품

대학원생 계열별 주제편차정도 추이



3

출판경과년수에 따른 도서 이용 현황 분석

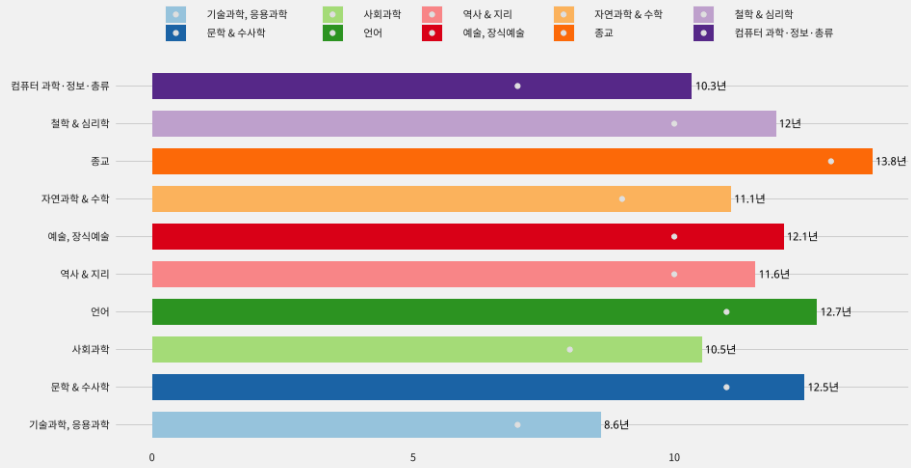
대출 도서 연령 평균

도서 연령 = 대출연도 - 출판연도

- 학생 이용자가 대출한 전체 도서의 출판경과년수 평균: 11.45년
- 주제에 따라 이용되는 자료의 최신성에 차이가 있음

장서 코드별 대출 도서 연령 평균값(전체)

점은 중앙값을 의미한다



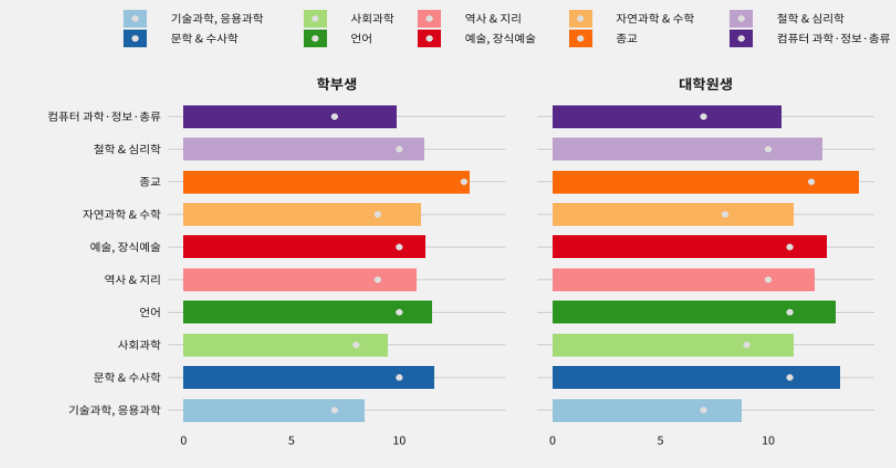
대출 도서 연령 평균

도서 연령 = 대출연도 - 출판연도

- 학부생과 대학원생의 자료 대출 경향 비교
- 대학원생 대비 학부생들이 전반적으로 더 신규 도서를 대출하는 경향을 보임

장서 코드별 대출 도서 연령 평균값

점은 중앙값을 의미한다





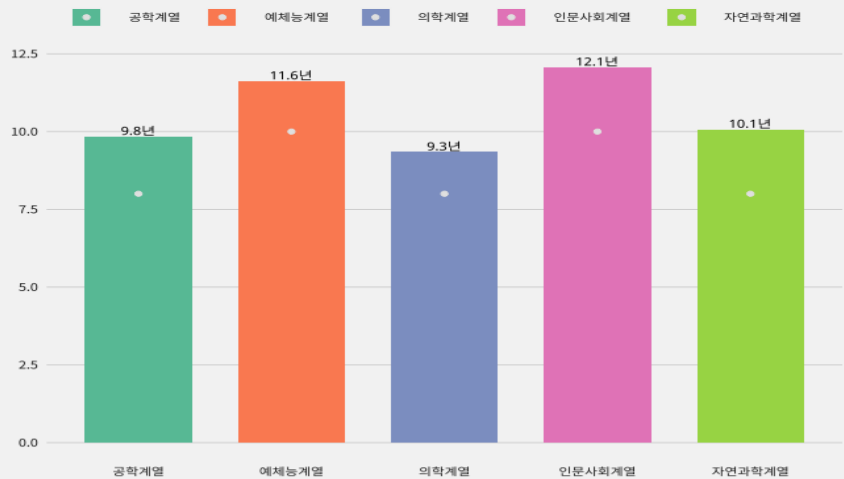
도서 연령에 따른 계열별 자료 대출 현황

도서 연령 = 대출연도 - 출판연도

- 공학계열, 의학계열은 최신 도서 대출 경향이 다른 계열보다 높은 것으로 나타남
- 인문사회계열, 예체능계열은 최신 도서에 대한 선호도가 상대적으로 낮은 것으로 보임

계열별 대출 도서 연령 평균값(전체)

점은 중앙값을 의미한다



도서관 빅데이터 활용방안

1. 대출 데이터 분류체계 개선

- 동일한 제목, 동일한 권수(volume)의 개별 서적들이 동일한 단행본으로 분류될 수 있도록 하는 새로운 분류체계 마련 필요
- (예) 책 볼륨별 별도 고유 번호 부여

	mms_id	barcode	item_loan_id
죄와 벌. 1-2	죄와벌 1 99184732102591	10101506023	10154066750002500
	죄와벌 2 99184732102591	10101448992	
	죄와벌 2 99184732102591	10101553998	10154066750002500
	죄와벌 2 99144636402591	10101075402	

mms_id (= book id) item_id

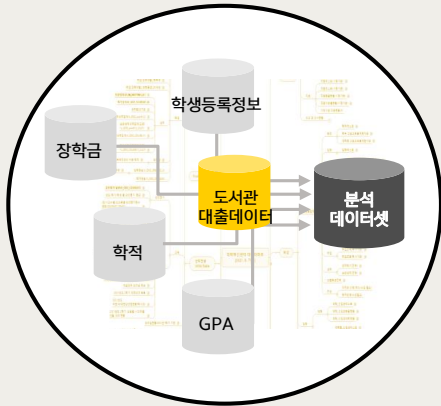
2. 대출 데이터 관리 개선

- 출판연도에 대한 정보 업데이트 필요

item_loan_id	publication_year	publication_date	loan_yr	publicationA	t
10696351150002500	0		1962	2018	1962
10812858530002500	0	25 25 cm.		2018	
11765536700002500	0		1962	2018	1962
12770818590002500	0		1962	2019	1962
13114816320002500	0		1962	2019	1962
13314021770002500	0		1962	2019	1962
16000269950002500	0	民國 72 [1983].		2020	1983
17021999760002500	0	?2003.		2020	2003
18031569630002500	0	明治35[1902]-昭和5[1930].		2020	
18458381400002500	0	New York :		2021	
18939968500002500	0	明治35[1902]-昭和5[1930].		2021	
19208592130002500	0		1984	2021	1984
20501316330002500	0	New York :		2021	

📌 도서관 빅데이터 활용방안

3. 정보화본부 학내 데이터와의 연계
 - (예) 구성원ID 비식별화 방식 통일



- 핵심장서 분석 및 출판경과연수에 따른 대출패턴 분석
 - 대출이 활발한 핵심장서를 선정하고 핵심장서 대출 패턴 분석
 - 이용이 활발한 도서와 이용되지 않는 도서를 중심으로 효율적 도서구매 및 구성 방안 마련
- 장서 수요 대비 공급 현황 분석
 - 강의계획서, 수강생 데이터 등과 연계하여 도서관 장서 구성이 교육과 연구 활동에 도움이 되고 있는지 분석
- 수업 설계 지원을 위한 도서 추천 시스템 구축
 - 교원들이 새로운 교과목 개설 또는 수업 개선에 활용할 수 있는 도서추천 서비스 제공

감사합니다.

서울대답게 서울대처럼

LikeSNU 2023 ALA 국제혁신상 수상 기념

대학도서관 혁신 컨퍼런스

학술대회 발표자료집

05

데이터와 데이터 이용문헌 간 링크의 기대 효과

서울대학교 한국사회과학자료원
구혜란 책임연구원





데이터와 이용문헌 간 링크의 기대효과

2023. 7. 6.
구혜란 (서울대학교 한국사회과학자료원)
대학도서관 혁신 컨퍼런스

목차

- 01 과학에서의 재현성 위기와 오픈데이터
- 02 오픈데이터와 학술정보생태계의 변화
- 03 데이터-문헌 연결의 필요성과 기대효과
- 04 실천 과제



1. 과학에서의 재현성 위기와 데이터

• Diederik Stapel scandal



3 | 1. 과학에서의 재현성 위기와 데이터

Science Integrity Digest

A blog about science integrity, by Elisabeth Bik, for Harbers-Bik LLC. Support my work at [Patreon.com/elisabethbik](https://patreon.com/elisabethbik)

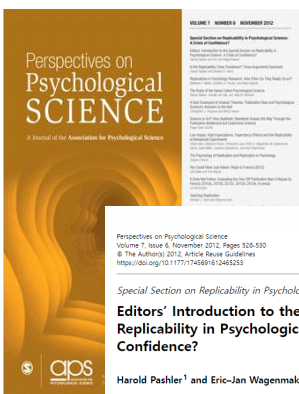
Fabrication: The Diederik Stapel case



One of the most discussed case of fabricated data comes from the Netherlands. [Diederik Stapel](#) was a professor of Social Psychology at Tilburg University and the University of Groningen. Some of his research was featured on news sites, such as his [Science paper](#) that found that people discriminate more in messy environments or another (unpublished) study that [thinking about meat made people less social](#).

Interestingly, Stapel never appeared to share raw data with his students or collaborators. During his tenure, Stapel claimed to be the only person who had access to research participants and facilities, such as schools. So even though he was the professor, he would not let his students collect the data. Instead, he handed them raw data that was collected by himself, and then let his students analyze it. The data often looked close to perfect, and when Stapel was asked about how he collected that data, he would get unusually defensive.

1. Intro: 과학에서의 재현성 위기와 데이터



Perspectives on Psychological Science
 Volume 7, Issue 6, November 2012, Pages 520-530
 © The Author(s) 2012. Article reuse guidelines: <https://doi.org/10.1177/1745691912465231>

Special Section on Replicability in Psychological Science: A Crisis of Confidence?

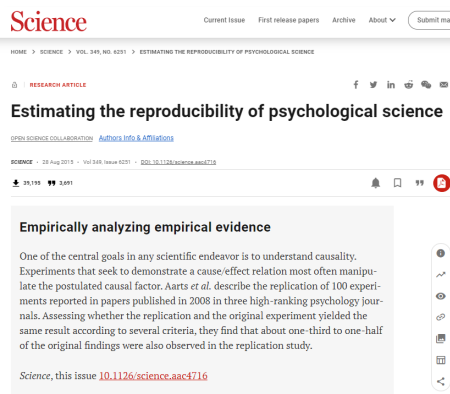
Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?

Harold Pashler¹ and Eric-Jan Wagenmakers²

¹University of California, San Diego
²University of Amsterdam, The Netherlands

Corresponding author(s):
 Hal Pashler, University of California, San Diego, Department of Psychology, 9500 Gilman Drive #0109, La Jolla, CA 92093-0109. E-mail: hpashler@ucsd.edu

Is there currently a crisis of confidence in psychological science reflecting an unprecedented level of doubt among practitioners about the reliability of research findings in the field? It would certainly appear that there is. These doubts emerged and grew as a



[Open Science Collaboration](#), Estimating the reproducibility of psychological science. *Science* 349, aac4716(2015). DOI: [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716)

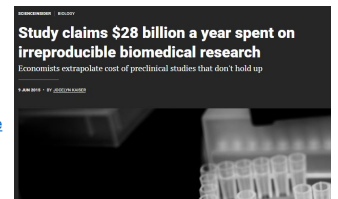
Cancer Research Is Broken

There's a replication crisis in biomedicine—and no one even knows how deep it runs.

BY DANIEL ENGBER APRIL 16, 2016 • 9:21 AM



<https://slate.com/technology/2016/04/biomedicine-facing-a-worse-replication-crisis-than-the-one-plaguing-psychology.html>



Study claims \$28 billion a year spent on irreproducible biomedical research

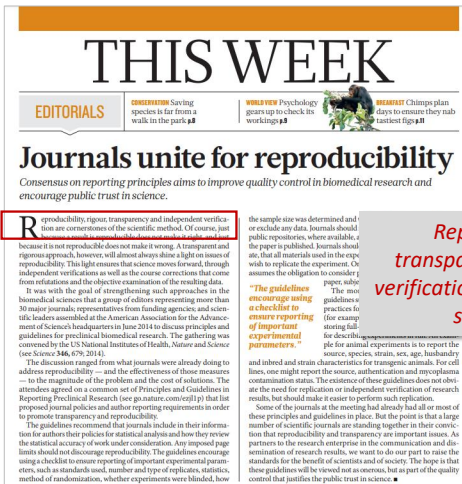
Economists extrapolate cost of preclinical studies that don't hold up

BY ANNE M. BYRNE

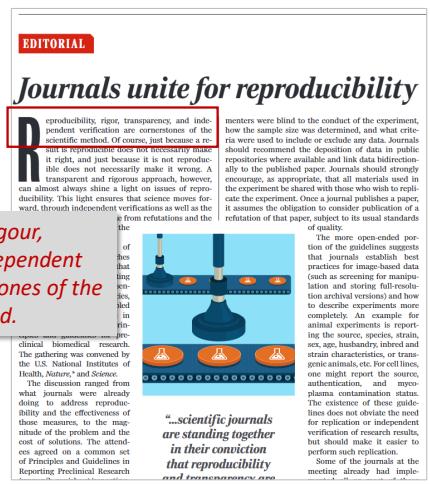
<https://www.science.org/content/article/study-claims-28-billion-year-spent-irreproducible-biomedical-research>

4 | 1. 과학에서의 재현성 위기와 데이터

1. Intro: 과학에서의 재현성 위기와 데이터



Journals unite for reproducibility. *Nature* 515, 7 (2014). <https://doi.org/10.1038/515007a>



Marcia McNutt, Journals unite for reproducibility. *Science* 346, 679-679(2014). DOI: [10.1126/science.aaa1724](https://doi.org/10.1126/science.aaa1724)

1. Intro: 과학에서의 재현성 위기와 데이터

- 데이터 가용성 진술문(data availability statement)
- 2018년 PLOS 발행 논문 21,793건의 약 94%, BMC 발행 논문 31,956건의 약 88%가 데이터 가용 여부에 대한 정보를 제공하고 있으며, PLOS 논문의 21%, BMC 논문의 12%가 데이터가 소장된 리포지토리 정보를 함께 제공하고 있는 것으로 나타남(Colavizza et al., 2020)*.

* Colavizza G, Hrynaszkiwicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. *PLOS ONE* 15(4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>



Data and Code Deposition

As outlined in the **TOP guidelines** above, the *Science* journals generally require all data underlying the results in published papers to be publicly and immediately available. Post-publication embargoes are not permitted, nor are stipulations for readers to contact the authors (rare



"When submitting a manuscript, authors must provide a **Data Availability Statement** ... The Data Availability Statement must list the name of the repository or repositories as well as digital object identifiers (DOIs), accession numbers or codes, or other **persistent identifiers** for all relevant data.

SPRINGER NATURE

Include a **data availability statement** in your research paper.

Data availability statements are important because they support **validation, reuse and citation**



of research data.

<https://www.springernature.com/la/authors/research-data-policy/data-availability-statements/12330880>



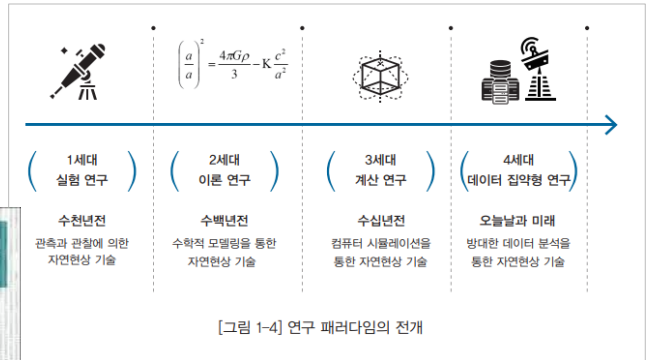
Availability of data and materials

All manuscripts must include an 'Availability of data and materials' statement. Data availability statements should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be

2. 오픈데이터와 학술정보 생태계의 변화

● 학술정보의 핵심 콘텐츠로서의 데이터

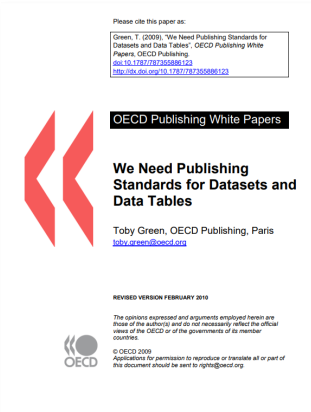
- 정보통신 기술의 발달
- 네트워크 혁명
- 디지털 전환
- 연구패러다임의 변화
- 데이터에 대한 새로운 시각



출처) 최희윤, 서태실. 2020. 오픈사이언스. KISTI. p.17

2. 오픈데이터와 학술정보 생태계의 변화

● 일급 연구 성과물로서의 데이터



“데이터세트는 중요한 학술적 기록의 일부이다. … 데이터는 저자나 독자, 사서들이 저널 논문이나 단행본을 이용하고 관리하는 것처럼 쉽게 이용하고, 찾을 수 있고, 관리할 수 있도록 학술정보시스템에 통합될 필요가 있다.”

Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres

Sarah Callaghan, Steve Donegan & Sam Pepler,
NCAS British Atmospheric Data Centre,
STFC Rutherford Appleton Laboratory

“데이터 인용과 출판은 데이터가 구할 수 있고, 동료간 검토가 가능하며, 인용가능하고, 쉽게 발견하고 재활용할 수 있는 일급 연구산출물로 간주되도록 해줄 것”

Callaghan et al. (2012). Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres. International Journal of Digital Curation. 7. <http://doi.org/10.2218/ijdc.v7i1.218>



(<https://www.cambridge.org/core/services/authors/open-data/data-citation>)

2. 오픈데이터와 학술정보 생태계의 변화

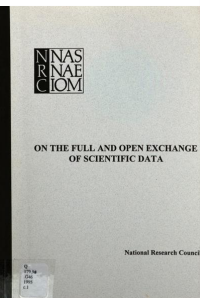
• 공공재로서의 데이터

“연구는 공공재로서 자유롭게 이용되어야 한다”

R. Merton. 1942. "Science and Technology in a Democratic Order." Journal of Legal and Political Sociology 1: 115-26.



National Research Council. 1985. *Sharing Research Data*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/2033>.



National Research Council. 1995. *On the Full and Open Exchange of Scientific Data*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/18769>.



OECD. 2004. Declaration on Access to Research Data from Public Funding



OECD. 2007. Principles and guidelines for access to research data from public funding

G8 open data charter. 2013

The G8 Open Data Charter is a commitment by the governments of the G8 to make their data more open and accessible to the public. It is a landmark document that sets out the principles and standards for open data.

- 1. Timely and Comprehensive**
Open data should be available in a timely and comprehensive manner. Governments should provide data in a timely and comprehensive manner.
- 2. Timely and Comprehensive**
Open data should be available in a timely and comprehensive manner. Governments should provide data in a timely and comprehensive manner.
- 3. Accessible and Usable**
Ensuring that data is machine-readable and easy to find will make data go further. Governments should ensure that data is accessible and usable.
- 4. Comparable and Interoperable**
Data has a multiplier effect. The potential value you can get from data is increased when it is comparable and interoperable.
- 5. For Improved Governance**
Open data helps the world to improve its governance. Governments should use open data to improve their governance.
- 6. For Inclusive Benefits**
Open data should be used to promote inclusive benefits. Governments should use open data to promote inclusive benefits.



The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, Ian Abbott, Greg Behr, Gabrielle Amdur, Mark Anton, Arie Adriaens, Michael Brinkmann, Ian Wilson, Robert Lu, Boris Beitzel, Philip A. Bourne, Abbas Baram, Anthony A. Bradley, Tim Clark, Michael Cooper, Nigel Collier, Oliver Conway, Scott Edwards, Oleg S. Dzuba, Richard Ekin, Alexandra Grotzinger, Bethan Haskell, J.C. Goh, Paul Groth, Carole Goble, Jeffrey S. Gordon, Jean Hendrix, Peter A.C.J. Hoare, Rob Hoober, Stefan Huber, Ines Ischke, Scott J. Lamber, Margaret L. Manley, Albert Mendel, David J. Miller, Deep Shrivastava, Thomas Steinbach, Marco Stoffel, Jean-Luc Stockli, Susanna-Assunta Sonnen, Erik Sjöström, Thore Stierwald, Srdjan Stokich, George Strawn, Monica A. Suter, Mark Thomson, Johan van den Laak, Erik van Mulligen, Jan Velasco, Anika Wärentzen, Peter Wüthrich, Andrew Wilkinson, Su Zhan, & Benedek Zoltai

Scientific Data 3 | Article number 160018 (2016) | <https://doi.org/10.1038/sdata.2016.18>

Wilkinson et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

2. 오픈데이터와 학술정보 생태계의 변화

• 관리의 대상으로서의 데이터

- 영국: RCUK(2013), 공공재원으로 생성된 연구데이터는 공공재이고 공개를 원칙으로 한다
- 캐나다: 공적 기금으로 조성된 연구비의 지원을 받아 생산된 데이터는 오픈 액세스로 공개할 것을 의무화.
- 미국
 - 2022년 백악관 Office of Science and Technology Policy(OSTP) 연구데이터 관리 정책
 - A. 논문과 데이터는 국민에게 개방된 리포지터리에서 공개
 - B. 연천만 달러 이상을 받고 있는 프로젝트는 180일 안에, 연 천만 달러 이하를 받고 있는 프로젝트는 360일 안에 논문과 데이터를 오픈할 수 있는 플랜을 내놓도록 한다.
 - C. 모든 정부 연구 결과물의 경우 반드시 digital persistent identifier (PDI) 와 관련 메타 데이터를 첨부 해야 한다.
 - D. 연구데이터 관리에 관련된 모든 비용을 계산해서 연구계획서에 미리 첨부해야 한다.
 - 2023년 National Institutes of Health(NIH, 미국국립보건원) 새로운 지침 발표
 - a) 연구비 신청시 연구데이터 관리계획 첨부 의무화
 - b) 연구 데이터 관리에 드는 모든 비용을 데이터 플랜에 명기
 - c) 데이터 매니지먼트에 필요한 준비가 잘 되고 있는지를 중간단계에서 NIH에서 모니터링 할 수 있다



Kanza, S., Knight, N.J. Behind every great research project is great data management. *BMC Res Notes* 15, 20 (2022). <https://doi.org/10.1186/s13104-022-05908-5>

2. 오픈데이터와 학술정보 생태계의 변화

- 데이터 관리와 공유를 위한 인프라



Publishing data to accompany your journal articles - Library, Archive & Open Research Services blog (lshtm.ac.uk)

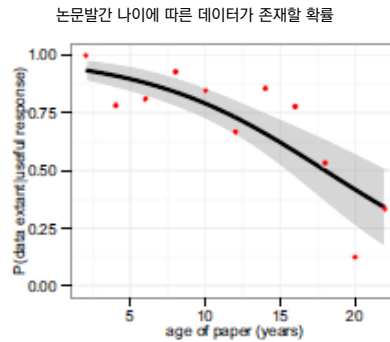
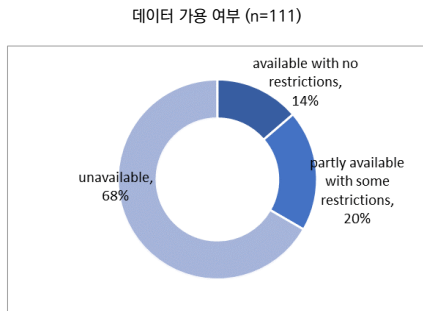
2. 오픈데이터와 학술정보 생태계의 변화

데이터는
일급 연구성과물로 공공재로서
모두에게 공개되고 공유되어
검증과 재현이 가능하도록
관리되어야 한다

- ⇒ 저널, 논문, 보고서, 단행본 등의 전통적인 도서관 서비스 콘텐츠의 확장
- ⇒ 연구데이터 관리와 서비스 방안 마련

3. 데이터-문헌 연결의 필요성과 기대효과

- 그 많은 데이터는 어디에?



Hardwicke와 Ioannidis(2018)*는 심리학과 정신의학 분야에서 2011년에서 2016년 사이에 발간된 논문 중에서 인용이 가장 많이 되는 111건의 논문을 선정하여 데이터 가용성을 확인해 본 결과 68%(76건)의 논문에 활용된 데이터를 구할 수 없었다.

Vines과 그의 동료들(2014)**은 1991년에서 2011년 사이에 발간된 516건의 논문을 분석한 결과, 논문이 발간된 이후 논문에 이용된 데이터가 존재할 확률이 매년 17%씩 줄어든다는 것을 확인

* Hardwicke, T. E., & Ioannidis, J. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLoS one*, 13(8), e0201856. <https://doi.org/10.1371/journal.pone.0201856>

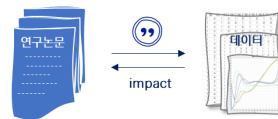
** Vines, T. H., Albert, A., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J. S., Renaut, S., & Rensson, D. J. (2014). The availability of research data declines rapidly with article age. *Current biology : CB*, 24(1), 94-97. <https://doi.org/10.1016/j.cub.2013.11.014>.

3. 데이터-문헌 연결의 필요성과 기대효과

- 데이터-문헌 연결의 기대 효과

1. 연구데이터의 가시성과 발견성 제고
2. 연구데이터의 재활용 활성화
3. 연구의 재현성과 투명성 제고
4. 연구 성과 인정 체계 보완

Colavizza와 그의 동료들(2020)은 논문과 데이터 리포지토리에 있는 연구에 활용한 데이터를 연결해놓았을 때, 평균적으로 25% 이상 논문의 인용 영향력이 높아지는 것을 발견.



Colavizza G, Hrynaskiewicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. *PLOS ONE* 15(4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>



California Digital Library, Crossref, DataCite, DataONE 등 다양한 기관들이 참여하여 일관되고 표준화된 방식으로 연구데이터 이용을 측정하는 데이터 이용 매트릭스(data usage metrics)를 개발하고 데이터 검색과 다운로드, 인용 건수 등을 제공하고 있음. <https://makedatacount.org/>



2012년 톰슨로이터사는 연구데이터 인용 및 색인정보 데이터베이스인 데이터 인용 색인(Data Citation Index, DCI)을 구축하고 데이터 피인용 횟수를 제공하고 있음.

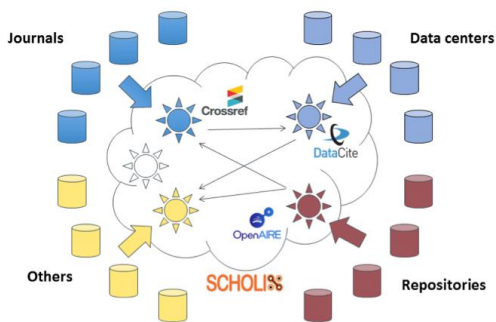
<https://clarivate.com/webofsciencegroup/solutions/webofscience-data-citation-index/>

3. 데이터-문헌 연결의 필요성과 기대효과

- Scholix case

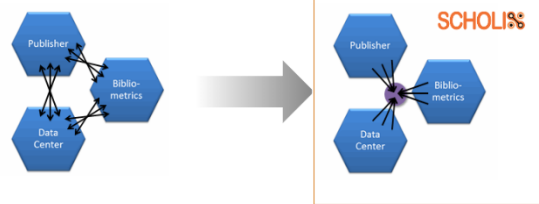
Scholix: a Framework for *Scholarly Link eXchange*

A new framework presenting a vision and guidelines for linking research data and literature using a common, global approach



Objective: move from
 a plethora of (mostly) bilateral arrangements between the different players...

to
 ... a one-for-all cross-referencing framework for articles and data



3. 데이터-문헌 연결의 필요성과 기대효과

3. 데이터-문헌 연결의 필요성과 기대효과

● KOSSDA case

데이터세트: 3,300여개
자료이용문헌: 13,000여건

이용문헌수	데이터수
0	976
1	1,808
2-9	502
10-99	123
100-999	59
1,000+	8

검색결과

자료 검색 Search Tips

불명동

검색

결과수 10 | 정렬 기준 정확도 | 정렬 내림차순

이후

검색결과 75건 중 1-10 (소요 시간: 0.045 초)

양적/질적자료 | 조사자료번호 | **자료이용문헌 (75)**

자료이용형 | 발행연도 | 인용서식

- 2016 안서연. (2016). 한국의 가구 노동소득 불평등 연구: 1999년과 2014년 불평등 요인 분석. 서울대학교.
 - 한국노동개발연구소, 2012: 가구원
 - 한국노동개발연구소, 2011: 가구원
 - 한국노동개발연구소, 2010: 가구원
 - 한국노동개발연구소, 2009: 가구원
 - 한국노동개발연구소, 2008: 가구원
- 2018 김용환. (2018). 소득 불평등 인식과 정부신뢰: 주관적 계층 상승 가능성을 중심으로. 서경대학교 대학원.
 - 한국노동개발연구소, 2012
 - 한국노동개발연구소, 2003-2021 [누적자료]
 - 한국노동개발연구소, 2003-2018 [누적자료]
- 2009 이성병. (2009, December 18). 외환위기 이후 전개된 한국의 부의 불평등: 지역을 중심으로.
 - 한국노동개발연구소, 2009-2018 [누적자료]

자료시리즈

한국노동개발연구소 (KOSSDA)

자료이용서식

김지현. 한국노동개발연구소, 2003-2018 [누적자료] [데이터 세트]. 성균관대학교 사범대학사치센터 [연구수용기관]. 한국사회과학자료원 (KOSSDA) [자료제공기관]. 2019-12-27. <https://doi.org/10.22687/KOSSDA-A1-CUM-0023-V2.0>.

자료이용안내

1) KOSSDA 2003-2018 누적자료에는 기존 공개된 연도별 데이터 중 오류를 수정한 내용을 포함하고 있고, 모든 연도에 적용할 가능한 수가 동일하지 않습니다. 성균관대학교 사범대학사치센터는 누락된 데이터의 적용을 우선하고 있습니다. 2) KOSSDA 2003-2018 누적자료에는 지역정보를 시·도 수준으로 제공합니다. 이는 조사응답자의 개인정보 및 비밀번호를 위한 조치로 보다 세분화된 지역 정보가 필요한 연구자는 SRC(kgsa.src@gmail.com)로 문의하시기 바랍니다.

상세 메타데이터 | 관련 파일 | **자료이용문헌**

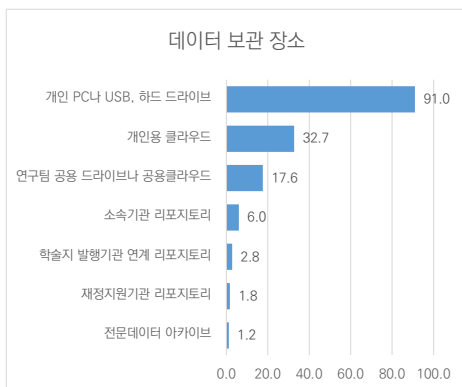
정렬 기준 | 정렬년도 | 정렬 | 내림차순 | 이후 | 다음

자료이용형 | 발행연도 | 인용서식

- 2022 김기룡. (2022). 노동조합과 투로조합: 노동조합 내 투로조합의 사회경제적 격차. 한국노동연구원.
 - 한국노동개발연구소, 2003-2021 [누적자료]
 - 한국노동개발연구소, 2003-2018 [누적자료]
- 2022 이영훈. (2022). 정부신뢰에 대한 무해인식의 영향. 한국정책학회, 27(1), 185-204.
 - 한국노동개발연구소, 2003-2021 [누적자료]
 - 한국노동개발연구소, 2003-2018 [누적자료]
- 2022 Seo G. (2022). Correlates of Contemporary Gender Preference for Children in South Korea. Population and Development Review, 46, 161-188. <https://doi.org/10.1111/jpdr.12408>
 - 한국노동개발연구소, 2003-2018 [누적자료]

4. 실천 과제

● 데이터 공개 및 공유 문화 확산



KOSSDA. 2021년. KOSSDA 연구데이터 관리와 공유에 대한 연구자 인식조사, 미발행자료. (사례수=725건)

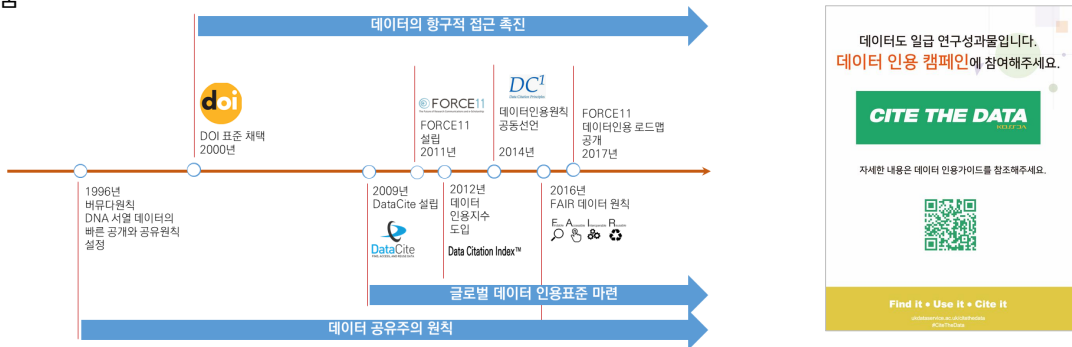
수집, 생산한 연구 데이터를 분실 또는 망실한 경험	%
있다	27.6
없다	72.4

데이터 공유, 제공 경험	%
전혀 공유하지 않음	44.3
지인이나 개인 연구자의 요청으로 직접 제공	74.5
개인 또는 연구팀의 웹사이트	18.1
논문 게재와 함께 학술지 DB를 통해 제공	33.7
소속기관의 리포지토리나 서버에 게시	11.0
전문데이터 기관에 기탁	4.8

4. 실천 과제

• 데이터 인용 규범 체계화

- 적절한 데이터 인용은 연구논문과 데이터를 연결해주고 데이터가 웹 상에서 영구적으로 찾아질 수 있도록 지원함으로써 데이터 발견과 접근성을 높이는 데 기여
- 데이터 인용은 데이터 출처와 데이터 이용 정보를 연결시킴으로써 연구의 투명성을 높여줄 뿐만 아니라 다른 연구자가 연구 결과를 재현하고 검증할 수 있도록 데이터에 접근하게 해줌으로써 연구 재현성 문제 해결에 기여할 수 있음
- 데이터 인용은 데이터 이용과 재이용을 추적할 수 있게 해줌으로써 보다 효율적이고 정확한 데이터 영향력 측정을 가능하게 해줌



4. 실천 과제

• 연구데이터 관리 체계 마련

연구데이터 관리 가이드라인 (Research Data Management Guidelines)

등록일 : 2022-06-16 조회수 : 2935

연구데이터 관리 가이드라인 (Research Data Management Guidelines)

- 연구데이터 관리 가이드라인 (Research Data Management Guidelines)
- 연구데이터 보존 가이드라인 (Research Data Preservation Guidelines)
- 연구데이터 관리 계획(DMP) 가이드라인 (DMP(Data Management Plan) Guidelines)
- 연구데이터 라이선스 가이드라인 (Research Data License Guidelines)
- 연구데이터 윤리 가이드라인 (Research Data Ethical Guidelines)
- 연구데이터 인용과 활용지침 가이드라인 (Guidelines for Indicators of Citation and Utilization of Research Data)
- 연구데이터 관리 규정(안) (Research Data Management Regulations(draft))
- 연구데이터 리포지터리 활용 지침서 (Research Data Repository Utilization Guidelines)
- 메타데이터 설계지침서 (Metadata Design Guidelines)

사회과학 연구데이터에 가치를 더하다

KOSSDA 연구데이터 관리 이해하기

자료기탁 가이드

데이터 인용 가이드북

2020년 8월 Ver. 1.0

2022년 9월 Ver. 1.0

4. 실천 과제

● 학술정보 기관 간의 협력

Scholix is supported by the following organizations:



감사합니다.



- 📍 0826 서울시 관악구 관악로 1 서울대학교 아시아연구소 101동 250호
- 🌐 <https://kossda.snu.ac.kr>
- ✉ kossda@snu.ac.kr
- ☎ 02.880.2111
- 👍 <https://www.facebook.com/kossda.snuac>
- 📷 https://www.instagram.com/kossda_snu
- 📺 <https://www.youtube.com/c/KOSSDA한국사회과학자료원>

