도서관 빅데이터 분석활용 체계 구축 사업 참여 안내

2018. 3.

국립중앙도서관

KiSTi www.kisti.re.kr 한국과학기술정보연구원 Korea Institute of Science and Technology Information

Ⅰ. 사업 소개

1. 사업 개요

○ 사업명 : 도서관 빅데이터 분석활용 체계 구축

○ 사업기간

(1차년도) 2014. 2. 1. ~ 2014. 12. 31.

(2차년도) 2015. 2. 1. ~ 2015. 12. 31.

(3차년도) 2016. 1. 1. ~ 2016. 12. 31.

(4차년도) 2017. 1. 1. ~ 2017. 12. 31.

(5차년도) 2018. 1. 1. ~ 2018. 12. 31.

○ 주최기관 : 국립중앙도서관

○ 수행기관 : 한국과학기술정보연구원

2. 세부 추진내용

○ **도서관 빅데이터 공유/활용 플랫폼 구축 :** 도서관 데이터, 공공데이터, 소셜 데이터 등의 다양한 데이터를 수집·분석할 수 있는 플랫폼을 구축하고, 도서관 분야 데이터의 공유·확산을 위한 체계 구축

○ **도서관 빅데이터 서비스 안정화 :** 빅데이터 분석 기술을 개발하여 이를 적용한 분석 플랫폼 구축과 도서관 빅데이터 활용 서비스 안정화 추진

○ **도서관 빅데이터 연구환경 조성** : 빅데이터 수집 및 연구를 위한 국내 협력 네트워크를 구성하고, 빅데이터 플랫폼 활용 세미나, 빅데이터 전문가 양성을 위한 교육, 도서관 빅데이터 활용 사례 공모 등을 개최함으로써 도서관 빅데이터의 활용도 제고

3. 도서관 빅데이터 서비스 소개

○ 도서관 정보나루

□ 도서관 빅데이터 사업을 통해 수집된 데이터를 개방 공유함으로써, 도서관 정책가, 사서, 연구자, 민간사업자 등이 이를 활용하여 새로운 콘텐츠 생산, 서비스 개발 및 연구를 수행하도록 지원하는 서비스

□ 주요 기능

- 스마트 대시보드
  - 참여도서관 수집 데이터 통계, 지역·연령별 인기 대출 장서 현황, 데이터 활용 현황 통계 제공
- 도서관 관련 공개 데이터 제공
  - 데이터 공개에 동의한 참여기관의 장서 및 대출빈도 데이터 제공
  - 텍스트 형식(CSV파일) 및 오픈 API 방식 지원

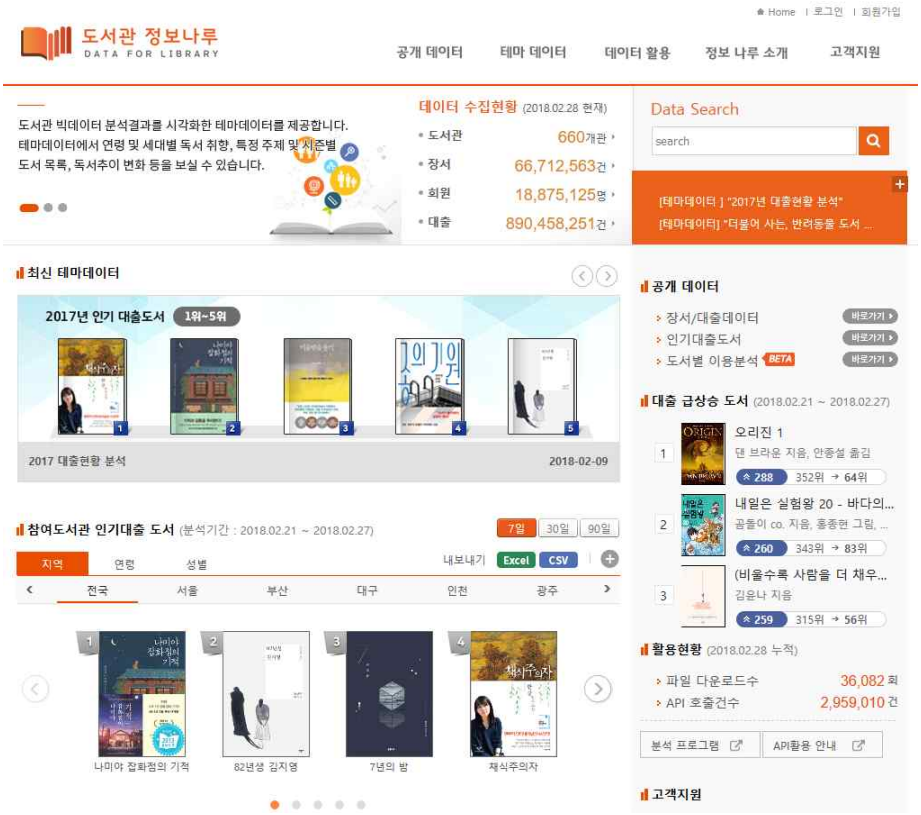
  ※ 개인정보를 비롯한 민감정보는 제외
- 테마 데이터 제공
  - 수집된 빅데이터를 활용하여 시사점이 있거나, 흥미를 유발하는 다양한 분석 테마를 선정하고 분석결과 제공
  - 분석결과 시각화 이미지, 분석에 활용한 소스데이터, 시각화 이미지 파일을 동시 제공

○ 테마 데이터

<데이터로 보는 4차산업혁명>

<2017 대출현황 분석>

□ 서비스 인터페이스



□ 주요 메뉴 구성

메뉴	서비스 내용
스마트 대시보드	• 수집데이터 현황판(참여도서관, 장서, 회원, 대출/일간 & 누적) • 지역별, 연령별, 성별 인기 대출 도서 분석(최근 90일 기준) • 데이터 활용 현황 : 파일 다운로드 수, API 호출건수
공개 데이터	• 데이터 공개 참여도서관 현황(도서관 디렉토리) • 참여도서관의 장서목록 및 대출 빈도 공개 • 참여도서관의 다대출도서공개(기간, 성별, 연령, 주제 등 다양한 검색가능) • 도서별 이용 분석(베타서비스) • 텍스트 형식(CSV파일) 및 오픈API 방식, CC BY라이선스
테마 데이터	• 수집데이터 기반 데이터 분석, 시각화 이미지 및 소스 데이터(엑셀)제공
데이터 활용	• 오픈API 활용방법(회원 가입 후 인증 키 발급) • 텍스트(CSV파일) 활용방법 • 사용자 매뉴얼 및 오픈API 매뉴얼 제공
정보 나루 소개	• 도서관 정보 나루 소개 • 도서관 빅데이터 사업 소개: R 분석 툴 Radar 제공 • 빅데이터 활용사례 소개
고객 지원	• 공지사항, 질의응답

○ 사서의사결정지원서비스(서비스명 : Solomon)

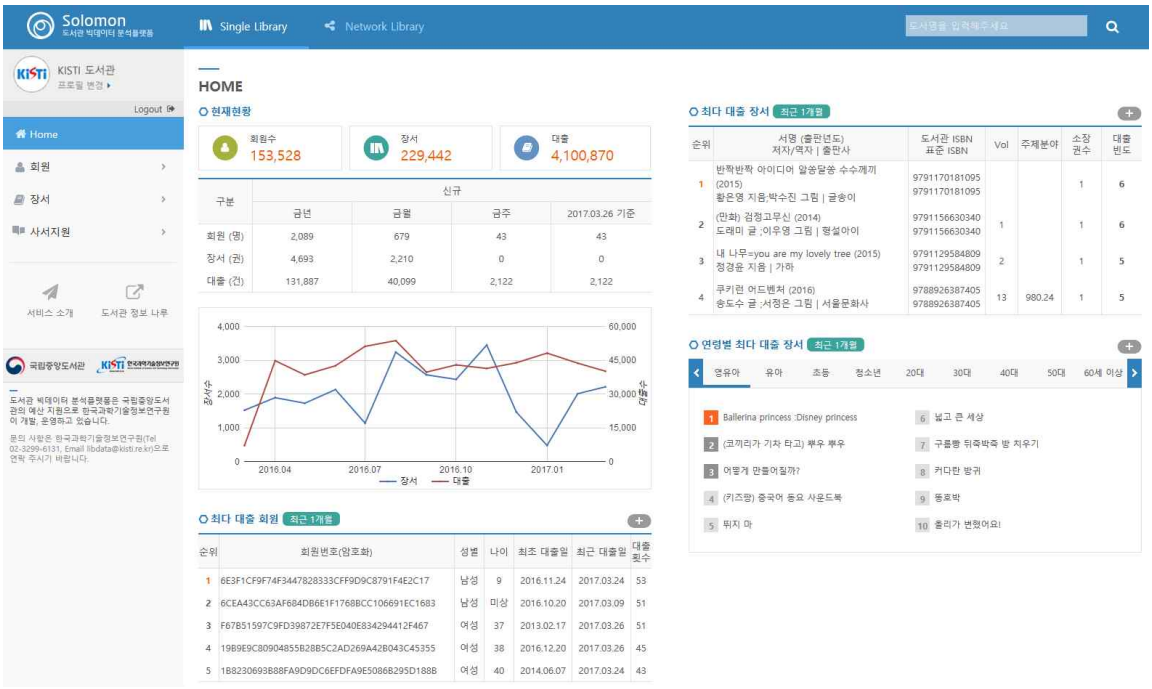
□ 도서관 내부에서 수집 가능한 장서 데이터, 이용자 데이터, 대출 데이터와 온라인 서점 데이터, 공공 데이터 등 도서관 외부 데이터를 수집 및 분석하여 도서관 운영을 위한 사서 업무를 지원하는 웹기반의 의사결정지원 서비스

□ 자관 특성에 적합한 수서 및 업무 지원, 타관과의 연계 분석 지원 가능

□ 주요 기능

- 도서관별 회원분석, 장서분석 결과 제공
- 기간별, 분류별 선택을 통한 데이터 조회
- 분석 결과에 대한 시각화 기능 제공
- 외부 데이터(타관, 서점) 연계 분석에 기반한 장서 추천
- 수집 도서관 전체 데이터 분석결과 제공(네트워크 라이브러리)
- 현재 330여 개 도서관에 계정 발급

□ 서비스 인터페이스



□ 주요 메뉴 구성

메뉴	세부메뉴		서비스 내용
Home			• 현재 누적 및 신규 데이터 현황 • 최근 1개월 최다 대출 장서(통합, 연령대별) 및 회원 현황
Single Library	회원	회원통계	• 성별, 연령, 지역 구분으로 연도별 신규/누적 회원 수 제공 • 성별&연령, 성별&지역, 연령&지역에 대한 교차분석
		대출회원 분석	• 성별·연령별 대출 분포, 자관 평균 대출건수를 전국 평균과 비교 • 지역별 대출회원수, 대출건수, 회원1명당 평균 대출건수 • 성별, 연령, 주제, 대상 독자별 회원 대출 순위 확인 • 성별, 연령, 대출 기간별 연체 회원 및 연체 빈도 확인
		자관/타관회원 분석	• 자관 회원과 타관 회원의 연도별 대출 현황 및 대출 건수 조회 • 자관 회원이 타관에서 대출한 건수 조회
		회원검색	• 대출 정보, 최근 1년간 대출 분석, 대출 성향, 회원별 추천도서 목록 등 제공
	장서	장서통계	• 연도별 신규/누적 장서의 주제별 분포 • 금년, 금월, 금주, 설정 기간에 따른 장서별 주제 분포 • KDC 분류에 따른 장서별 분포의 계층별 확인
		대출장서 분석	• 장서별, 주제별 대출빈도 순위 및 기간별 대출주제 분포 • 월별, 요일별, 시간대별 대출/반납 현황 분포
		상호대차	• 관내 상호대차 현황 및 대출대비 비중 분석, 이용 현황
		미대출장서 분석	• 주제분야별 미대출 장서 현황 확인 및 목록 제공
		장서평가	• 장서별, 주제분야별 장서회전율
		장서검색	• 대출 추이, 주요 키워드, 함께 대출된 도서 시각화 • 등록기간, 대출기간, 반납여부, 연체여부, 주제, 도서명, ISBN
Network Library	도서 추천		• 대출장서순위 목록 : 사업 참여도서관 전체의 대출 건수 확인 • 베스트셀러, 신간도서 조회(매일 갱신 - 알라딘 제공)
	외부데이터 연계		• 2014년 이후 기상데이터를 자관 대출데이터와 연계 분석 • 서점조합연합회 소속 지역서점 판매량 연계 분석(예정)
	관내 도서관 간 비교 분석		• 도서관 간 회원 비교 분석(자관 회원은 최초 가입한 도서관 기준) • 대출회원 및 대출장서 비교 분석 • 연도별/장서별 장서 현황 비교 분석
	관내 도서관 통합분석		• 관내 도서관 전체의 회원, 장서, 대출 현황 파악 및 통합회원의 대출분석
	타 도서관과 자관 비교 분석		• 자관의 장서 수와 비슷한 도서관들과의 대출 건수, 주제별 장서 수, 장서회전율 비교 분석
	전국 단위 대출 분석		• 광역지자체 단위 대출 분포
	광역지자체별 도서관 분석		• 시군구 단위의 회원 현황, 대출 분석, 대출 장서 분석

○ Open API

□ 추천도서 목록을 비롯하여 도서관 정보 나루에서 제공하는 데이터를 API를 통해 공개하고 민간에서 누구나 해당 기능을 활용하여 서비스 구현 및 활용할 수 있도록 개방

□ 주요 기능

- 연구자, 개인, 기관들이 활용할 수 있는 빅데이터 분석 내용 지원
- 도서관, 장서/대출 데이터, 인기대출도서 조회 기능 제공
- 동시대출도서 및 추천도서 목록 조회 기능 제공
- 도서 상세 조회 및 키워드 목록 조회 기능 제공

□ Open API 활용방법

<오픈 API 이용 절차>

1) 회원 가입 : 도서관 정보나루(https://data4library.kr/)에 접속 후 회원 가입 신청
2) 인증키 신청 : 가입한 회원계정으로 로그인 한 후 인증키 발급 신청
3) 인증키 발급 : 인증키 발급신청 후 인증키 부여 받기
4) Open API 활용 : Open API 관련 매뉴얼을 참조하여 데이터 활용

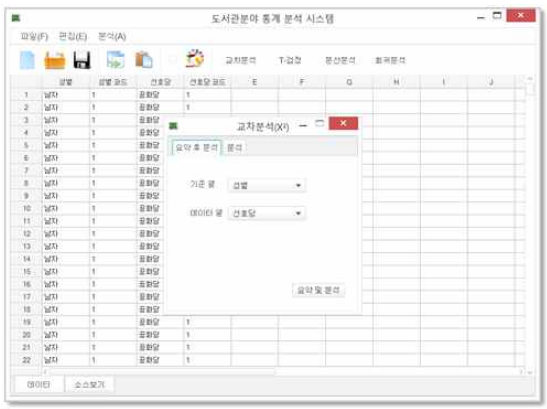
○ **R 기반 통계 분석 도구(서비스명 : Radar)**

□ 이용자가 한 번에 PC에 다운받아 활용할 수 있는 R 기반 통계 분석 도구로써, 비전문가도 손쉽게 활용할 수 있는 통계분석 지원 프로그램

□ 주요 기능

- 도서관 운영 및 관련 연구 활동을 지원하는 통계 서비스 제공
  - 기술통계, 회귀분석, 교차분석, ANOVA, t-test 등 핵심적인 통계기능 제공
  - 위계적 회귀분석, 네트워크 분석 등 고급 통계기능 제공
  - 손쉽게 사용이 가능한 GUI 기반 데이터 편집 및 분석결과 제공

□ 서비스 인터페이스



4. 기대 효과

○ 도서관 운영, 장서개발, 서비스 개선에 필요한 정책 결정 기반 데이터 확보

○ 빅데이터 분석을 통해 실효성 있고 선제적인 서비스 운영 및 개발, 주요 의사 결정 지원

○ 차세대 정보서비스 기술 구현으로 국가 지식 경쟁력 강화 도모

Ⅱ. 사업 참여 방법

1. 개요

○ 빅데이터 분석 플랫폼을 활용하고자 하는 도서관은 자관이 소장하고 있는 **서지, 이용자, 대출 데이터를 국립중앙도서관과 KISTI에 제공**하여야 함.

○ 제공해야 할 서지데이터, 대출데이터, 이용자 데이터 필드는 <참고> 참조

○ 2018년 참여 가능 공공도서관 수 : **100개 도서관**

<참고> 2017년 현재 분석시스템 활용 대상 기관

광역시도	도서관명	도서관 수
서울	강남구립도서관, 강서구립도서관, 관악구립도서관, 광진구립도서관, 도봉문화정보도서관, 동대문구립도서관, 서초구립도서관, 성동구립도서관, 성북구립도서관, 서울특별시교육청 도서관, 서울도서관, 송파구립도서관, 중구구립도서관, 중랑구립도서관	122
부산	금정도서관, 기장도서관, 남구도서관, 부산광역시립시민도서관, 부산광역시립구포도서관, 화명도서관, 사상도서관, 다대도서관, 영도도서관, 부산광역시립중앙도서관, 부산영어도서관, 부산광역시립해운대도서관, 재송어린이도서관	14
대구	달성군립도서관, 동구구립도서관, 대구광역시립북부도서관, 대구광역시립서부도서관, 수성구립도서관, 대구광역시립중앙도서관, 대구광역시립대봉도서관, 대구광역시립남부도서관, 대구광역시립두류도서관, 달서구립도서관, 대구광역시립달성도서관, 대구광역시립동부도서관, 대구광역시립수성도서관, 2.28도서관	46
인천	부평구립도서관, 인천광역시부평도서관, 인천광역시북구도서관, 서구구립도서관, 영종도서관	14
광주	광주광역시립도서관, 남구구립도서관, 서구구립도서관, 북구구립도서관	10
대전	한밭도서관 외 대전광역시 공공도서관, 관평도서관	25
울산	울산남부도서관, 북구구립도서관	21
세종	국립세종도서관	1
경기	경기평생교육학습관, 고양시도서관센터, 광명시립도서관, 군포시중앙도서관, 남양주시립도서관, 부천시립도서관, 수원시립도서관, 안산시립도서관, 안성시립도서관, 안양시립도서관, 양주시립도서관, 오산시립도서관, 느티나무도서관, 의왕시립도서관, 이천시립도서관, 파주시립도서관, 포천시립중앙도서관	173
강원	원주시립도서관, 강원도교육청 도서관	23
충남	금산군립도서관, 서산시립도서관, 예산도서관, 천안시립도서관, 충청남도남부학습관, 충청남도서부학습관, 충청남도학생교육문화원	21
충북	충청북도교육청 도서관, 담양군립도서관, 제천시립도서관, 청주시립도서관	32
경남	거창군립한마음도서관, 고성도서관, 김해통합도서관, 밀양시립도서관, 양산시립도서관, 창원시 문화도서관사업소, 통영시립도서관, 함양도서관	59
경북	경북도립상주도서관, 경북도립안동도서관, 안동시립도서관, 영천시립도서관, 포항시립도서관	47
전남	전남도립도서관, 순천시립도서관, 여수시립도서관	15
전북	순창군립도서관, 완주군립도서관, 익산시립도서관, 전주시립도서관, 전라북도청도서관	22
제주	제주특별자치도 공공도서관	15
계		**660**

2. 참여방법

○ <붙임 2>의 “공공도서관 빅데이터 제공 및 참여 신청서” 작성 후 이메일 혹은 공문 제출 (팩스 접수 불가)

※ 도서관 정보 나루(https://www.data4library.kr)에서도 다운로드 가능

○ 신청서 접수마감 : **2018년 4월 27일(금)**까지 도착분에 한함.

3. 도서관 선정 기준

○ 국립중앙도서관과 KISTI는 본 사업에 참여를 신청한 도서관을 대상으로 아래와 같은 기준을 가지고 **총 100개 기관** 선정 작업 진행

○ 도서관 선정 기준

구분	기준	배점	만점
과거 데이터 제공 기간	개관 시부터 현재까지(전체)	30	30
	최근 5년 ~ 현재	20	
	최근 3년 ~ 현재	10	
	최근 1년 ~ 현재	5	
데이터 제공 주기	실시간	30	30
	일 1회	20	
	주 1회	10	
	월 1회	5	
유사시 DB 접근 권한	원격 접속	20	20
	방문 접속	10	
데이터 공개 여부	도서관 정보나루 내 데이터 공개	10	10
	도서관 정보나루 내 데이터 비공개	0	
과거데이터 제공	과거데이터 벌크 제공(외장하드 등)	10	10
	과거데이터 벌크 제공 불허	0	
합계			100

※ 과거 데이터를 제외한 모든 데이터는 원격 접속을 통해 제공해야 하며, 자세한 수집 데이터 내역은 “<참고> 수집 데이터 예시” 참조.

4. 선정절차

○ 선정 관련 일정

4.11~ 4.12	**도서관 빅데이터 분석활용 교육 및 사업설명회**	- 1차(서울) : 2018년 4월 11일(수), 10:00~17:00, KISTI 서울분원 3층 대회의실 - 2차(광주) : 2018년 4월 12일(목), 13:30~16:00, 아시아문화전당 문화정보원 지하4층 강의실1
4.16~ 4.27	**참여신청서 접수**	2018년 4월 27일(금), 18:00 까지
5.4	**선정결과 통보**	2018년 5월 4일(금) 공문 발송 예정 ※ 선정 통보 일정은 상황에 따라 다소 늦춰질 수 있음.
5.14~ 6.29	**선정도서관 방문 및 데이터 수집**	2018년 5월 14일(월)~ 6월 29일(금)(예정) - 선정 기관 데이터 담당자와의 협의를 통해 진행 예정 - 선정 도서관 방문횟수 약 2~3회 ※ 도서관의 상황에 따라 변동 가능

○ 플랫폼 활용 시기

- 2018년 데이터 수집 완료 즉시 사서의사결정지원서비스 이용 가능
- 플랫폼 사용 방법에 대한 교육 지원

<참고> 수집 데이터 예시

1. 서지데이터

가. 단행본 종 정보

컬럼 명	자료유형	길이	설명
레코드 KEY	NUMBER		[PK][Unique]레코드KEY
KDC 분류기호	VARCHAR2	44	
KDC 버전	VARCHAR2	30	
입력일자	DATE		
수정일자	DATE		
MARC	LONG		

나. 단행본 책 정보

컬 럼 명	자료유형	길이	설명
레코드 KEY	NUMBER		[PK][Unique]레코드KEY
종정보 KEY	NUMBER		
대표책여부	CHAR	1	
등록번호	CHAR	12	
수입구분	CHAR	1	
관리구분	CHAR	2	
매체구분	CHAR	2	
이용제한구분	CHAR	2	
이용대상구분	CHAR	2	
저작자	VARCHAR2	4000	저자정보(AUTHOR)
편/권차	VARCHAR2	150	권 정보(VOL)
발행년	VARCHAR2	100	PUBLISH_YEAR
딸림자료	VARCHAR2	100	
낱권ISBN	VARCHAR2	50	EA_ISBN
낱권ISBN부가기호	VARCHAR2	60	
분류기호구분	CHAR	1	
분류기호	VARCHAR2	22	
도서기호	VARCHAR2	22	별치기호 없을시 추가수집
권책기호	VARCHAR2	100	
복본기호	VARCHAR2	22	
본표제	VARCHAR2	4000	제목(TITLE)
등록일자	DATE		
수정일자	DATE		

2. 회원데이터

컬럼 명	자료유형	길이	설명
사용자 Key (자동생성 번호)	VARCHAR2	44	암호화 대상 (PK, Rec_Key)
가족 아이디	VARCHAR2	44	암호화 대상
생년	DATE		
우편번호(집)	VARCHAR2	24	
우편번호(근무지)	VARCHAR2	24	
성별	CHAR	1	
회원 가입일	DATE		
정보 수정일	DATE		
가입 도서관	VARCHAR2	30	
회원 등급	VARCHAR2	30	

3. 대출데이터

컬럼 명	자료유형	길이	설명
레코드 KEY	NUMBER		[PK][Unique]레코드KEY
대출책의 KEY	NUMBER		
대출자의 KEY	NUMBER		암호화 대상, 사용자 아이디
대출형태	CHAR	1	
반납형태	CHAR	1	
대출일	DATE		
반납예정일	DATE		
예약일	DATE		
예약만료일	DATE		
반납일	DATE		
대출상태	CHAR	1	
관리구분	CHAR	2	도서관 코드
대출장소 (관리구분)	CHAR	2	도서관 코드
반납장소 (관리구분)	CHAR	2	도서관 코드
대출한곳	VARCHAR2	50	
반납한곳	VARCHAR2	50	

4. 기타 주제분류, 별치코드 등 시스템 코드정보

※ 위 데이터는 예시이며, 참여 도서관의 자료관리시스템 및 그 버전에 따라 다소 달라질 수 있음.

5. 데이터 암호화 및 개인정보 처리 방안

○ 개인정보보호법 및 정보통신망법 준수

- 공공도서관에서 수집하는 이용자 데이터는 개인정보보호법에서 정한 바대로 이용자로부터 “제3자에 대한 정보제공 동의”를 받지 않은 이상 KISTI는 개인정보를 수집할 수 없음.
- 따라서 공공도서관으로부터 수집하는 이용자 데이터는 개인정보보호법에서 정한 바에 따라 특정인을 식별할 수 없도록 조치하고 있음. (법률 자문, 행정안전부 확인 완료)
  - 예) 사용자 key를 복호화 할 수 없도록 암호화 하고, 이외의 정보는 생년, 성별, 거주지(직장) 우편번호, 대출기록만을 수집하고 있음.

○ 개인정보 보호 방안

- 외부 보안솔루션 모듈 : 암호화 프로그램 제공
- KISTI : 암호화 키 생성
- 개발업체 : 암호화 코드를 소스에 반영

○ 데이터 암호화 및 보안 대책

- Oracle, MySQL, MS-SQL의 경우 DBMS에서 암호화 방법 제공
- 기타 DBMS의 경우에는 수집 프로그램에서 암호화 수행
- 사용자 번호(대출증 바코드), 이름, 휴대전화 번호 등의 개인정보는 수집 대상이 아님