

## 결과보고서 요약

사업기관	인문학연구원		
사업명	북한 인문학 시맨틱 데이터 아카이브 구축		
사업책임자	성명	소속	직위
	김종일	인문학연구원/ 고고미술사학과	부원장/ 교수
사업기간	2021.3.1.~2022.1.31	사업비	

### 1. 사업 목적

- 북한 인문학을 대상으로 한 최초의 학제적 연구 수행.
- 디지털 인문학 방법론을 활용한 북한 인문학 아카이브 편찬.

### 2. 사업 내용

- 한국 문학 및 해외 문학 분야 기초 데이터 구축  
북한에서 발표된 인문학 분야의 학술논저(논문과 단행본) 약 15,000여 건의 자료를 3년간 수집·정리·분석·입력하여 디지털 데이터베이스 구축. 제2차 연도(2021)에는 ‘한국문학’과 ‘외국문학’ 영역의 지식 정보를 데이터로 편찬.
- 인물, 문헌, 기관 등을 중심으로 북한 인문학 위키 페이지 작성  
북한 인문학을 이끈 주요 인물, 핵심 문헌, 관계 기관 등의 위키 페이지를 작성. 타임테이블, 전자지도, 네트워크 그래프 등의 다양한 확장 기능을 활용하여 아날로그 환경의 선형적 읽기를 뛰어넘는 정보의 시각화를 구현.

### 3. 사업 성과

- 시맨틱 데이터 베이스 구축  
기 구축된 ‘인문학 시맨틱 데이터 아카이브’의 데이터 6,300여 건(Node: 2,900여 건, Edge: 3,400여 건)을 XML 형식으로 고도화.
- 위키 페이지 작성  
인물 7명, 학술지 10건, 기관 1건, 문헌 10건 등.
- 학술회의 조직 및 발표  
학술대회 세션 조직 및 발표 1건, 학술 강연회 조직 및 발표 2건, 발표 1건.
- 논문 집필  
KCI 등재지 북한 인문학 관련 논문 1편 게재 (통일기반구축사업 사사표기)

## 인문학연구원

### 1. 사업 배경

#### □ 개요

- 사업명: 북한 인문학 시맨틱 데이터 아카이브 구축
- 기관명: 인문학연구원

#### □ 추진 배경 및 목적

- 추진배경
  - 북한 인문학에 대한 기초 연구를 디지털 환경에서 체계적으로 진행.
  - 북한 인문학에 대한 관심을 환기하고 북한 연구에 대한 호응을 유도.
  - COVID-19 사태로 인한 비대면 연구 및 교육 환경에 적극 대응.
  - 디지털 인문학 교육 플랫폼으로서 북한에 대한 대중적 이해 확산.
- 목적
  - 북한 인문학을 대상으로 한 최초의 학제적 연구 수행.
  - 디지털 인문학 방법론을 활용한 북한 인문학 아카이브 편찬.
  - 1차 연도 사업성과(2020)를 토대로 데이터를 확충 및 재편.

### 2. 사업 내용

#### □ 사업 추진 내용

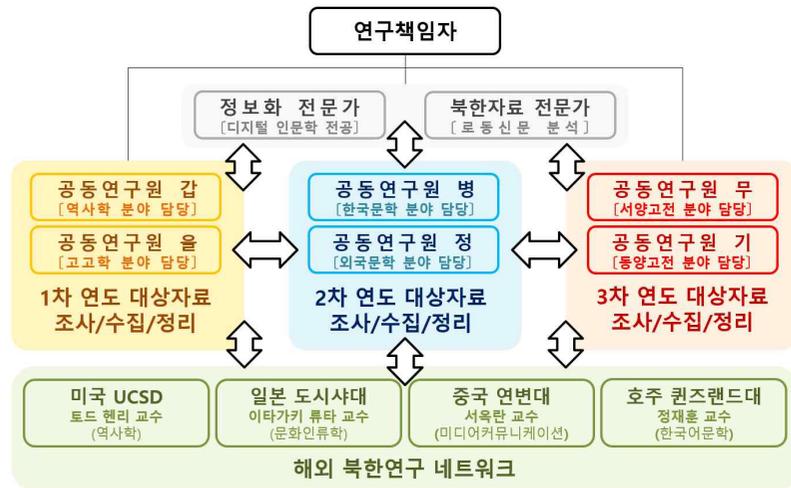
- 연구 개요
  - 연구 과정  
본 사업은 북한에서 발표된 인문학 분야의 학술논저(논문과 단행본) 약 15,000여 건의 자료를 3년간 수집·정리·분석·입력하여 목록화하고, 연구 과정에서 도출된 모든 지식 정보를 디지털 데이터베이스로 구축하는 연구·편찬 사업임. 통일을 대비하여 체계적 역량을 증진하기 위한 기초적 성격의 연구로서, 학술적 차원

에서 남북 교류 협력을 유도할 수 있는 선구적인 작업임. 연구과정은 ① 자료 현황 파악 및 수집 ② 수집된 자료의 목록화 ③ 자료 분석을 통한 온톨로지 설계 ④ 데이터베이스 구축 ⑤ 연구 결과물로서의 웹서비스 구현으로 구성됨.

- 연차별 계획

제1차 연도(2020)에는 북한 인문학 중 ‘역사학’ 과 ‘고고학’ 영역의 지식 정보를 정리하고 데이터로 편찬. 그 과정에서 얻은 노하우를 적극 활용하여 제2차 연도(2021)에는 ‘한국문학’ 과 ‘외국문학’, 제3차 연도에는 ‘동양고전’ 과 ‘서양고전’ 영역의 지식 정보를 데이터로 편찬.

○ 연구조직



○ 연구방법과 활용기술

- 연구방법

- 연구 전 과정에 디지털 인문학적 방법 적용 : 연구의 기획 단계에서부터 디지털 인문학의 방법을 도입해 연구의 효율성을 제고하고 성과의 사회 환원을 촉진하고자 함.
- 인문학자와 디지털 인문학자의 협업을 통한 데이터 구축 : 인문학자와 디지털 인문학자의 협업을 통해 데이터 연구와 데이터베이스 편찬 업무의 분리·이원화를 경계하며, 텍스트 및 콘텍스트 분석의 성과를 반영한 시맨틱 데이터 아카이브를 구축하여 ‘인문학의 디지털적 전환’ (digital turn in humanities)을 도모함.
- 온라인 협업 및 공유 공간 운용 : 자료 수집과 정리 및 목록화 그리고 데이터 구축 등의 업무를 수행함으로써 결과가 자동으로 축적되는 데이터베이스 서버를

온라인상에서 호스팅하여 운영함.

- 웹 서비스로서의 연구 결과물 구현 : 형식적 차원의 엑셀 파일 목록을 만드는 등 수동적인 데이터베이스 구축에 머물지 않고, 연구 결과물이 웹에서 바로 공개 서비스되도록 함.

- 활용기술

- Wiki 소프트웨어 활용 : 디지털 협업 환경 구축을 위해 Wiki 소프트웨어를 활용해 데이터를 축적하고 커뮤니케이션을 진행함.
- RDB(SQL)와 Python 및 Vis.js 활용 : 데이터베이스 편찬 및 데이터 시각화를 위한 ODBC 구현 및 데이터 시각화 결과물 구현에 활용함.



□ 일정별 추진 사항

일 정	추진 내용 (데이터)	추진 내용 (위키)
2021.03~2021.04	자료 수집	원문 및 서지 정보 수집
2021.05.~2021.06.	자료 목록화	학술사적 맥락 검토
2021.07.~2021.09.	데이터 모델 조정	위키 페이지 편찬
2021.10.~2021.11.	데이터 구축 및 DB 편찬	위키 페이지 데이터 검수
2021.12.~2022.01.	데이터베이스 업데이트 및 위키 페이지 확충	

□ 참여 인력

○ 책임자

성 명	대 학	학 과	직 급
김종일	인문대학	고고미술사학과	교수
연구소(원)	구 내 전 화	휴대전화	이메일
인문학연구원	4083		

○ 주요 참여 인력

성명	소속	직급	
최윤영	인문대 독어독문학과	교수	
류준필	인문대 중어중문학과	부교수	
김현	인문학연구원	부교수	
홍종욱	인문학연구원	부교수	
김도민	인문학연구원	선임연구원	
고일홍	아시아연구소	HK연구교수	
장문석	경희대 국어국문학과	조교수	
류인태	한중연 디지털인문학연구소	연구원	

```

58 </기사>↓
59 ↓
60 </북한인문학정보>↓
61 <북한인문학정보>↓
62 ↓
63 <기사><!--Class:Article-->↓
64 <기본정보>↓
65 <대표명>조선 문학과와 한문 문학</대표명><!--한국어면 현대 한국어 맞춤법에 따라 한글로 표기. 두음법칙
66 도 적용. 다른 언어면 상당해서 결정-->↓
67 <식별자>A20002</식별자><!--A20001 부터 번호 부여-->↓
68 <원제목>조선 문학과와 한문 문학</원제목><!--언어, 맞춤법, 한자 노출 여부, 띄어쓰기 등등 완전히 원문
69 그대로-->↓
70 <범주>논문</범주><!--예시) 소설, 비평, 논문 등-->↓
71 <번역원전></번역원전><!--원가 있으면 상당해서 결정-->↓
72 <특집사항></특집사항><!--원가 있으면 상당해서 결정-->↓
73 </기본정보>↓
74 ↓
75 <시간관계>↓
76 <집필연월일><!--관계어-->↓
77 <대표명>1955년</대표명><!--예시) 연도:1929년, 연월:1929년03월, 일자:1929년03월17일-->↓
78 <식별자>D1955</식별자><!--예시) 연도:D1929, 연월:D192903, 일자:D19290317-->↓

```

그림x. 신규 XML 데이터 구축 예시(『역사과학』)

- 수록 논문을 기준으로 집필연월일, 집필자, 번역자, 집필기관, 번역기관 등 유관 관계 정보를 모두 XML Schema 설계에 반영해서 정리함.
- 구축한 XML 데이터를 RDB에 적재해서 SQL 기술을 이용해 Data Table을 구성해서 EXCEL 포맷의 데이터로 추출함.
- XML과 EXCEL 포맷의 Dataset을 모두 북한인문학 아카이브에 업로드함으로써 데이터를 공유함.(<http://dh.aks.ac.kr/~nkx/wiki/index.php/Dataset>)

3. 사업 성과

□ 사업 성과 (구체적 내용 기술)

○ 북한 인문학 시맨틱 데이터베이스 신규 구축

- 『역사과학』, 『역사제문제』, 『조선문학』 대상 데이터의 XML화 진행.
- 『역사과학』 수록 논문 1113건, 『역사제문제』 수록 논문 111건 업로드 완료. 『조선문학』 입력 및 검수 작업 중.

```

1 <?xml version="1.0" encoding="UTF-8"?>↓
2 <?xml-stylesheet type="text/xsl" href=""?>↓
3 ↓
4 ↓
5 <북한인문학정보>↓
6 ↓
7 <기사><!--Class:Article-->↓
8 <기본정보>↓
9 <대표명>강화조약의 역사적 교훈과 당시의 국내외 정세</대표명><!--한국어면 현대 한국어 맞춤법에 따라 한글
10 로 표기. 두음법칙도 적용. 다른 언어면 상당해서 결정-->↓
11 <식별자>A20001</식별자><!--A20001 부터 번호 부여-->↓
12 <원제목>江華條約의 歷史的敎訓과 當時의 國內外情勢</원제목><!--언어, 맞춤법, 한자 노출 여부, 띄어쓰기
13 등등 완전히 원문 그대로-->↓
14 <범주></범주><!--예시) 소설, 비평, 논문 등-->↓
15 <번역원전></번역원전><!--원가 있으면 상당해서 결정-->↓
16 <특집사항></특집사항><!--원가 있으면 상당해서 결정-->↓
17 </기본정보>↓
18 ↓
19 <시간관계>↓
20 <집필연월일><!--관계어-->↓
21 <대표명>1948년 07월 30일</대표명><!--예시) 연도:1929년, 연월:1929년03월, 일자:1929년03월17
22 일-->↓
23 <식별자>D19480730</식별자><!--예시) 연도:D1929, 연월:D192903, 일자:D19290317-->↓

```

그림x. 신규 XML 데이터 구축 예시(『역사제문제』)

○ 북한 인문학 위키 데이터 신규 구축

- 북한 인문학과 관련된 주요 인물, 핵심 문헌, 관계 기관 등을 대상으로 위키 페이지를 작성.
- 인물 7건(김광진, 이청원, 도유호, 김석형, 한설야, 이극로, 신구현), 학술지 10건(역사제문제, 역사과학, 문화유물, 문화유산, 고고민속, 조선고고연구, 민족문화유산, 문화전선, 조선문학, 문학예술), 기관 1건(과학원), 문헌 10건(역사과학, 조선문학, 고고민속 게재 주요 문헌).
- 타임라인, 전자지도, 네트워크 그래프 등 디지털 환경에서의 다양한 확장 도구를 활용해 비선형적 형식으로 학술 정보를 들여다볼 수 있는 다양한 시각화 결과물 구현 작업을 시도함.
- 하이퍼텍스트 기반의 학술 위키 페이지 구축은 인문학적 전문성을 기반으로 한 북한 관련 학술 정보를 웹상에서 적극적으로 개방하고 공유하기 위한 실험적 시도로서의 의미를 지님.



그림x. 신규 위키 데이터 구축 예시(이극로)  
 (URL: <http://dh.aks.ac.kr/~nkh/wiki/index.php/이극로>)

### ○ 학술회의 조직 및 발표

- 2021.3.17. University of Queensland 한국학연구소와 함께 기획한 International Joint Research Seminar Series on “North Korean Studies in the Asia-Pacific Region” 의 일환으로 “Digital Humanities and North Korean Studies: Construction and Use of the “North Korean Humanities Digital Archive” 개최.
- 2021.5.26. 상허학회 주최 집담회에서 “북으로 간 강좌파: 이청원의 생애와 사상” 으로 발표.
- 2021.6.9. University of Queensland 한국학연구소와 함께 “북한 인문학 데이터 아카이브의 인물 위키페이지 구축” 이라는 학술 강연회를 개최.
- 2021.11.30. 통일기반구축사업 연합학술대회에서 “북한 인문학 시맨틱 데이터 아카이브 구축” 이라는 세션을 조직.

### ○ 북한 인문학 관련 논문 KCI 등재지 게재

- 박형진, 2021, 「아시아라는 흔적: 이청원의 조선연구와 김남천」, 『상허학보』 63, 179-201쪽. (통일기반구축사업 사사표기)

### □ 관련분야 기여도

- 북한 지역 연구의 새로운 지평을 제시
  - 관련 학술회의를 통해 역사학, 문학, 고고학을 넘나드는 북한 인문학에 대한 학제적, 포괄적 이해를 제시.
  - 문헌, 연구자, 기관 등의 연관을 보임으로써 연구사 정리를 넘어 인문학 인프라 자체를 연구의 대상으로 재정립.
- 디지털 인문학 연구 및 교육에 새로운 방식을 제안
  - 코로나 시대 비대면 연구 및 교육의 토대가 될 수 있는 디지털 인문학 플랫폼으로 기능.
  - 서울대학교가 추진 중인 혁신공유대학의 ‘역사정보 마이크로 디그리 과정’ 설계의 토대를 제공.

## 5. 자체평가 및 건의사항

### □ 사업평가

- 사업의 적절성 및 효율성
  - 북한 인문학 자원을 디지털 콘텐츠화 한다는 계획에 맞춰 공동연구원, 연구보조원이 역할 분담하여 시맨틱 데이터 아카이브를 구축.
- 사업의 영향력
  - 본 연구로 축적된 지식과 정보를 ‘북한 인문학 시맨틱 데이터 아카이브’ 웹사이트 통해 사회에 환원. 예컨대 ‘이청원 북한’으로 구글링하면 동 사이트가 두 번째로 표시됨.
- 사업의 발전가능성
  - 제3차 연도까지 연구를 수행하여 1단계 사업을 마치면 북한 인문학 연구 및 교육을 위한 플랫폼 구현 가능. 이후 혁신공유대학의 교육 플랫폼으로 활용하며 유지, 확충할 계획.

### □ 미흡한 점(한계) 및 개선(보완)할 점

- 시맨틱 데이터의 XML 형식으로서의 가공은 노동집약적 연구활동으로 공동연구원 혼자만의 노력으로는 한계가 있음
- 고도화된 데이터 시각화 및 분석(네트워크 분석 등)을 위해서는 시맨틱

데이터의 추가적인 처리가 요구됨

□ **향후 계획**

- 북한 디지털 인문학 학문 후속 세대 양성을 통한 시맨틱 데이터 생산의 지속성 확보
- 북한 디지털 인문학 전문가 양성 워크숍 정기 개최 (2022-23년 통일기반 구축사업 예산 및 서울대 동아연구소 매칭 펀드 활용)
- 워크숍 참여 인력 활용을 통한 아카이브 확충
- 북한 인문학 시맨틱 데이터의 새로운 시각화 및 분석을 위한 데이터 프로세싱 시스템 개발
- ‘PDF → Plain Text → XML → JSON’ 과정의 자동화 추구

□ **건의사항**

- 없음

**6. 성과 관련 지표**

□ **대표 사업 실적**

- ‘북한 인문학 시맨틱 데이터 아카이브’ 구축
- 연구활동으로 수집·구축된 모든 데이터를 공유를 목적으로 설계되었으며, DB의 역할과 사업결과 확산 플랫폼의 역할을 수행하고 있음.

□ **성과 자율 지표**

- 기구축된 ‘인문학 시맨틱 데이터 아카이브’의 데이터 6,300여건(Node: 2,900여건, Edge: 3,400여건)을 새로운 형태로 가공함
- XML 형식으로 시맨틱 데이터 가공
- ‘인문학 시맨틱 데이터 아카이브’ 데이터의 공유를 위해 생성된 ‘위키 페이지’ 확충 (2020-21년 대비)
- 위키페이지 문서 주제의 확충 (인물, 기관)
- 위키페이지 콘텐츠의 확충 (1개 문서, 2274 단어 → 7개 문서, 24,303 단어)