

구술사연구 제15권 1호

ISSN : 2233-8349(Print) 2713-9328(Online)

구술아카이브의 데이터 검색과 공유의 한계점과 개선 방안에 관한 연구

이정연

To cite this article : 이정연 (2024) 구술아카이브의 데이터 검색과 공유의 한계점과 개선 방안에 관한 연구, 구술사연구, 15:1, 47-72

① earticle에서 제공하는 모든 저작물의 저작권은 원저작자에게 있으며, 학술교육원은 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

구술아카이브의 데이터 검색과 공유의 한계점과 개선 방안에 관한 연구*

이정연**

〈국문초록〉

지난 30년 동안 구술채록의 양적 증가와 구술사 연구의 확대에도 불구하고 국내 구술자료의 검색과 활용이 용이하지 않은 원인을 살펴보고자 하였다. 국내 구술아카이브의 경우 구술채록의 결과물인 구술원자료의 데이터베이스 구축과 검색이 용이하지 않으며, 개별 구술레코드 중심의 선형적 기술과 기관의 자체 개발 보존 중심의 시스템으로 운영되고 있는 상황이다. 반면 해외의 경우에는 구술 원자료와 데이터 구조화 작업과 검색시스템이 데이터베이스화 되어 있으며, 검색 도구와 통합검색이 도서관과의 연계로 서비스가 제공되고 있었다.

앞으로 국내 구술자료들이 데이터로 검색되고 활용되기 위한 개선 방안으로 다음과 같이 제언하였다. 구술자료 라이프사이클 단계별로 생산되는 구술자료가 표준화된 데이터 구조화 기반의 구술아카이브 모델링 설계와 데이터베이스 구축이 이루어져야 한다. 내용의 데이터화를 하는 전거파일, 색인 등의 전문화된 작업과, 개별 메타데이터 요소별 통합 색인 데이터베이스 기반의 검색시스템 개발이 필요하다. 우리나라 구술자료의 소재를 파악하기 위한 종합목록이 필요한데, 기존의 기초조사 목록을 기반으로 각 기관에서 목록을 업데이트하여 상호 공개하고 접근할 수 있도록 한다. 온라인 서비스의 이용자들에게 교육 및 연구 등에 대한 제한 범위 등을 고려한 <온라인 이용 동의서>와 구술자료의 학술적인 인용 방법론을 제공하는 것이 필요하다. 구술자료는 데이터화 하여 디지털보존과 활용을 고려하여, 데이터 수집, 구축, 분석과 해석을 포괄한 학제적 연구가 정책적, 실무적 협력으로 논의되어야 함을 제언하였다.

주제어: 구술사, 구술 아카이브, 디지털 아카이브 플랫폼, 데이터 구조, 검색 시스템, 색인, 디지털인문학, 융합 연구

* 이 논문은 2023년 한국구술사학회 추계학술대회에서 발표한 것을 수정한 것이다.
이 논문은 2022년 대한민국 교육부와 한국연구재단의 일반공동연구지원사업의 지원을 받아 수행된 연구이다(NRF-2022S1A5A2A03050638).

** 한양대학교 글로벌다문화연구원 연구위원. jyonlee@gmail.com

1. 들어가며

클로드 란치만(Claude Lanzmann) 감독의 Shoah(쇼와, 1985)는 홀로코스트 생존자들과 관리자들의 기억을 인터뷰 내용을 영화로 제작하여, 대중들에게 홀로코스트를 알리기 시작하였다. 구술은 구술자와 면담자와의 상호작용으로 산출되는 작업물이며 대화라는 작업 속에서 기억의 사건과 장소, 인물, 관계 등을 불러오게 된다. 구술자료는 그 목적에 따라 1차 자료를 구술채록을 통해 생산하여 연구를 수행하기도 하고, 다큐멘터리, 영화, 멀티미디어 역사 자료, 박물관 큐레이션 등 2차 저작물과 전시 등 콘텐츠의 생성으로 이어지기도 한다. 구술 연구는 대부분 텍스트로 전환된 것을 기반으로 인용되고 해석되는데, 실제로 구술이라는 특성은 텍스트성보다는 몸짓, 표정 등 감성 이입이 표현되어 유용하게 해석될 수 있다. 구술자료는 1차적으로 구술자와 면담자의 대화를 영상(음성)으로 저장하고 보존하면서 자료로 생성되며, 1차 자료는 2차 자료의 새로운 창작물로 활용되기 위하여 검색되고 이용되어야 그 가치가 발현된다. 그러나 현재 우리나라 구술 채록은 1차 자료 생산에만 집중되어 있으며, 과거 역사의 진상 규명을 위한 목적으로 시작되었기 때문에 2차 자료인 단행본 출간에 그치고 있어 1차 원자료가 소실되고 있는 상황이다(배은희 2018; 선영란 2019; 윤택림 2023). 또한 구술프로젝트를 주관하는 기관이 학술적 연구나 혹은 콘텐츠로 활용을 위한 원자료 검색하기 위한 과정과 보존 및 서비스 시스템으로 재구성되어 있지 않은 경우가 대부분이다. 구술자료만이 가질 수 있는 구술성, 맥락성, 시대의 사회적 배경을 이해할 수 있는 기초자원으로 인식되지 못하여 문화 역사적 자원을 놓치고 있다.

본 연구는 지난 30년 동안 한국에서 구술채록 프로젝트의 양적 증가와 구술사 연구가 확대되고 있음에도 불구하고 구술 원자료를 이용하기 위한 접근과 활용이 용이하지 않은 이유를 살펴보고자 하였다. 본 연구를 수행

하기 위한 연구방법론은 문헌 연구와 국내외 구술아카이브 플랫폼 웹 서비스 사례 연구를 중심으로 비교 조사 분석하였다. 이를 통해 국내 구술아카이브의 특성과 데이터 검색의 한계점을 밝혀내고 그 해결 방안을 제언해 보고자 하였다.

2. 구술자료 라이프사이클에 따른 생산 자료 및 자료 유형

구술채록은 면담자와 구술자의 대화 형식으로 기록되는데, 구술프로젝트를 통해 구술채록이 수행되면서 구술자료가 생산된다. 구술자료는 어떤 질문을 하고 대답하는지에 관한 구술 내용 뿐 아니라 언제 누가 어떤 질문을 왜 하게 되는지, 구술프로젝트의 기획 주제와 목적은 무엇인지, 참여하는 기관과 참여자는 누구인지에 대한 맥락 정보는 구술채록 당시의 상황을 해석할 수 있는 유용한 데이터가 된다. 일반적으로 면담 일지나 면담 후기에 채록 일시, 장소 및 구술자와 면담자의 상세정보가 기술되어 있으나 이러한 정보들이 표준화된 데이터 구조화를 통한 설계와 내용의 데이터 구축을 통해 검색이 될 수 있도록 하는 것이 필요하다.

구술자료 라이프사이클은 <표 1>과 같이 구술프로젝트에 대한 기획 및 관리 영역, 구술채록 사전 조사, 구술채록 진행, 구술채록 후 데이터화, 구술자료 활용을 위한 서비스 등 전 과정을 말한다. 구술자료 라이프사이클 전 영역에 걸쳐 생산되는 자료 매체는 텍스트, 음성, 영상, 이미지 등 다양한 멀티미디어 유형이다.

〈표 1〉 구술자료 라이프사이클에 따른 생산 자료 및 자료 유형

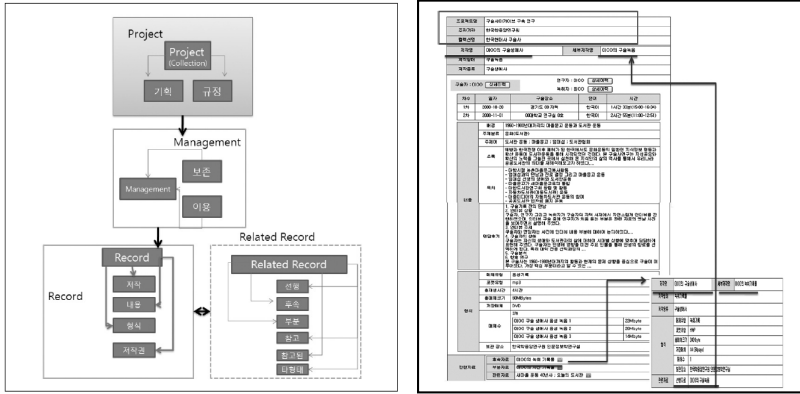
구술자료 라이프사이클	생산 자료	자료 유형
구술프로젝트 기획 및 관리	구술프로젝트 관리 메타데이터 (프로젝트명, 컬렉션명, 저작권임자, 프로젝트 기획 및 결과물, 내용, 기간)	텍스트
	구술컬렉션 관리 메타데이터 (컬렉션명, 프로젝트명, 저작권임자, 컬렉션 기획 및 결과물, 내용, 기간)	텍스트
구술채록 사전조사	사전 인터뷰 질문지	텍스트
	구술자 선정 (구술자 및 면담자 데이터)	텍스트
구술채록 진행 중	구술 음성	음성
	구술 영상	영상
	관련 사진 및 문서	이미지, 텍스트
	구술동의서	텍스트
구술채록 후 데이터화	구술 녹취문 (상세 목차)	텍스트
	면담일지, 감독확인서, 비공개 내역서, 구술서비스동의서 등	텍스트
	자료 유형별 메타데이터, 색인	텍스트
	편집 구술 영상 (콘텐츠)	영상
	구술 2차 저작물 (단행본 출간, 디지털 콘텐츠, 영화 등)	텍스트, 영상
	구술 동영상(음성) 및 녹취록 매칭 메타데이터	텍스트, 음성, 영상
구술자료 활용서비스	프로젝트 구술 목록 및 상세정보 검색 인터페이스	텍스트, 영상, 이미지, 음성 등
	필드별 메타데이터 색인 데이터베이스	
	주제별 색인 및 통합 검색 데이터베이스	
	공개 및 비공개 내역 검색 (온라인 이용동의서)	

구술자료 라이프사이클 각 단계별 생산되는 자료와 자료 유형을 살펴보면 다음과 같다.

첫째, 구술프로젝트는 구술프로젝트를 주관하는 기관의 목적과 기간,

저작책임자, 수행기관, 기획서 및 결과보고서 등이 문서가 생산되며, 구술 프로젝트는 다양한 컬렉션을 기획하면서 구술자료들을 구축하게 된다. 구술프로젝트의 컬렉션별 특성을 기록하는 것은 구술자료의 생산 맥락을 알 수 있는 근거가 된다. 따라서 구술프로젝트 관리 영역 단계에 대한 자료 생산이 필요하다. 프로젝트를 수행하는 주체와 목적, 주제, 시기, 개괄적인 내용 등에 관한 데이터를 축적해야 한다. 둘째, 구술채록 사전 조사 단계에서 사전인터뷰 질문지가 생산되며, 구술자 선정을 통해 구술자 및 면담자 정보가 생산된다. 셋째, 구술채록 진행 중에는 구술자와 면담자의 구술 음성, 구술 영상 및 관련 사진 및 관련 문서와 구술동의서 등이 생산된다. 넷째, 구술 채록 후에는 통상적으로 구술 녹취문과 녹취문 검독확인서가 생산되며 면담일지에 구술 채록의 맥락 정보 및 저작권과 관련된 문서가 도출된다. 생산되는 자료 유형별로 메타데이터를 작성해야 하며, 구술 자료 내용의 주제어, 인명, 사건, 시대 등의 색인 작업이 이루어진다. 또한 1차 구술 원저작물을 편집한 구술 영상이나 단행본 출간, 디지털콘텐츠와 영화 등 2차 저작물을 생산한다. 구술자를 중심으로 한 구술채록의 다양한 형태의 자료들의 연관 관계를 매핑 설정해야 하는데, 구술자의 메타데이터와 산출되는 자료유형들(영상 및 음성 원본과 녹취록, 관련 사진, 텍스트 등)의 연관 관계를 매핑 설정한다. 이를 위해서는 구술저작물의 구술자 단위레코드 메타데이터와 구술레코드 저작 간의 관계 설계에 기반을 두어야 한다. 다섯째, 구술자료 활용서비스 단계에서는 구술자료 검색 및 보존 인터페이스 개발을 통해 구술프로젝트별로 구술목록, 상세정보, 주제 검색, 공개된 원본 내용 검색 등이 이루어져야 한다. 이를 위해서 <그림 1>과 같이 프로젝트, 관리, 각 구술자 레코드 간의 관계를 통해 데이터와 원본 자료 형태와의 연계가 데이터베이스로 구축되어야 한다. 주제 색인이나 분류가 부여되어야 정확한 검색 결과를 도출할 수 있다. 이를 토대로 구술자료의 목록이 생성되고, 항목별 검색과 키워드 검색이 가능하다.

표준화된 메타데이터 스키마에 맞게 구술자료의 데이터가 기술되고, 데이터 요소별로 자동 색인 데이터베이스가 생성되어야 한다. 이러한 표준화된 구술자료 메타데이터 요소 항목¹⁾에 대한 각 기관별 구술프로젝트 관리 및 구술레코드 단위의 데이터가 구축된다면, 타 기관의 목록과도 연계하여 데이터를 통합할 수 있다.



〈그림 1〉 출처 이정연(2009). 구술프로젝트 중심의 구술사기록물 메타데이터 구현 p.174, 181

3. 해외 구술아카이브 시스템 사례 및 이용 연구

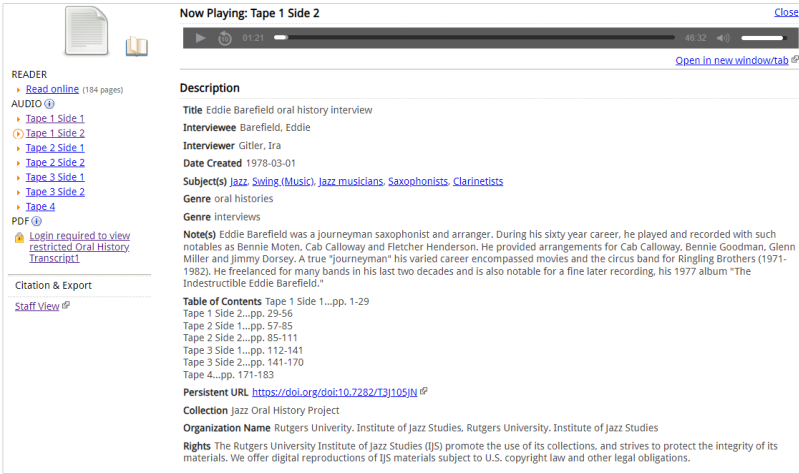
1) 럿거스대학교 구술아카이브

럼저스대학교(Rutgers, New Jersey)는 1994년부터 Rutgers Oral History Archives(ROHA)를 통해 뉴저지의 사회와 문화적 역사에 관한 사람들의 스토리에 대하여 아카이빙 하고 있다. White, K 와 C. Radiack(2017)는 국립예

1) 구술자료 메타데이터 작성 지침(국립중앙도서관 2023)을 참조

술기부금 펀드로 뉴욕시의 비영리기관인 Jazz Interaction에서 관리하였던 JOHP(재즈구술역사프로젝트)가 럿거스대학교 ROHA로 이관된 사례를 자세히 설명하고 있다. NEH(National Endowment for the Humanities)의 지원을 받아 1972년부터 1983년 사이에 재즈뮤지션들의 구술자료들의 릴테잎, 오디오 카세트테이프(122명 아티스트, 1,450시간)을 디지털화 하였는데 이 과정에서 디지털 아카이빙의 필요 가치, 권리 문제, 자원 평가에 관한 분석과 평가가 있었다. 끈끈하게 변해버린 오디오 녹음을 디지털로 재생 복원하면서 디지털 오디오 파일을 생성하였고, 이를 녹취록의 페이지 번호와 연계하였다. 녹취록은 문자인식 소프트웨어를 통해 실행되어 웹에서 녹취록을 검색할 수 있게 하였다. 아날로그 오디오는 표준화된 매체변환 가이드라인에 의하여 매체 변환을 하였다(오디오: 96kHz, 44.1kHz, 16비트, 보존용 마스터오디오: wave)(Rutgers University 2010).

Rutgers University Community Repository에 JOHP 컬렉션을 추가한 것은 럿거스대학이 주립대학도서관의 역할을 통해 지역 사료들 중 하나로 구술 자료를 채택하였기 때문이다. 럿거스대학은 인터뷰 진행자와 구술자 간에 협상된 모든 권리를 상속하였고 온라인 액세스와 관련한 권리계약을 통하여 접근 방법을 해결하였다. <그림 2>에서 나타나듯이 표준화된 메타데이터 스키마에 따라 서명, 저작자, 날짜, 주제, 장르, 주기, 목차, 컬렉션명, 권리조항 등의 데이터 구조화 및 데이터 작업을 수행하였으며, 인터뷰 저작에 대한 고유 번호인 DOI(Digital Object Identifier)를 부여하고 있다. 구술 자료의 검색은 레코드 필드별로 키워드 검색이 가능하다. 주제, 카테고리, 인명, 날짜, 제목, 컬렉션명 검색이 가능하며, 구술자, 면담자, 악기별로도 브라우징 검색도 가능하다. 주제명에 관한 통일성은 LCSH(Library of Congress Subject Headings)를 부여하고 있다.



〈그림 2〉 JOHP 구술프로젝트 아날로그 테이프의 디지털오디오 변환 구분 및
녹취록 시퀀스별 메타데이터 사례

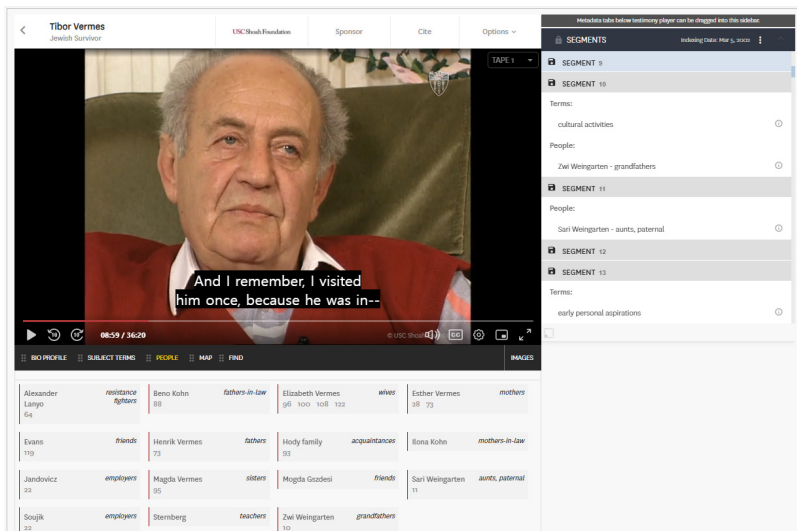
2) USC SHOHA Foundation

스티븐스필버그 감독은 ‘쉰들러 리스트’ 영화 제작 후에 유대인 생존자와 증인이 이야기를 통해 증언한 홀로코스트 구술채록을 USC 대학도서관과 협력하여 SHOAH 재단을 설립하였다.

인류 대학살의 역사적 사건을 증언을 후손들에게 알리기 위하여 아카이빙 하여 학술 및 교육적으로 활용하기 위하여 출발하였다. 1994년부터 홀로코스트 컬렉션을 시작으로 캄보디아, 아르메니안, 르완다, 난징, 로잉하, 북시리아, 과테말라 학살 등 인류의 대량 학살 생존자들의 인터뷰를 온라인 아카이빙 하여 검색 및 이용 서비스를 VHA(Visual History Archive)로 제공하고 있다.

VHA의 특징은 증언 구술영상을 디지털화하고 기본적인 구술자료의 맥락정보를 보존용으로 기술할 뿐 아니라 데이터로서 검색되고 활용할 수

있도록 작업을 수행하였다. 구술인터뷰 내용을 세그먼트로 구분하고 상세하게 검색될 수 있도록 색인 작업을 통한 주제 및 인물, 지역, 시대, 키워드를 부여하였다. 〈그림 3〉은 홀로코스트 생존자 Tibor Vermes의 10번째 내용 구분(세그먼트)의 인터뷰로 해당 내용의 주제 용어와 언급되는 인물, 구술자와 인물과의 관계를 속성으로 기술하여 상세하게 이해할 수 있도록 제공하고 있다.

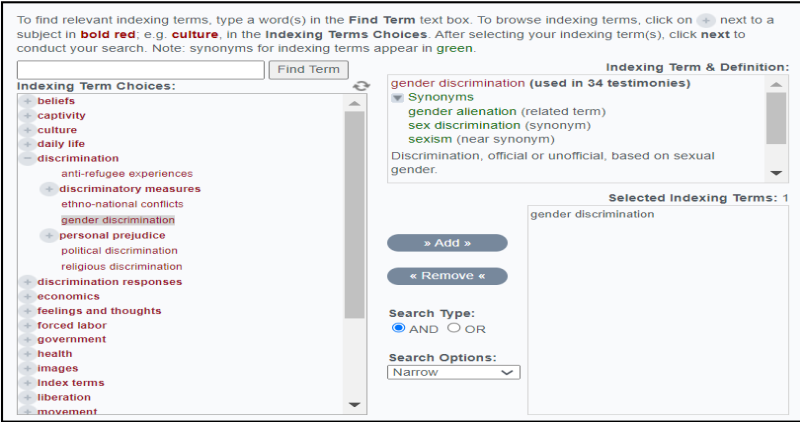


〈그림 3〉 유대인생존자 Tibor Vermes 구술인터뷰 세그먼트별 색인어, 인물 등 정보 접근점 메타정보(VHA)

색인 작업은 단행본의 각주에 해당 내용의 상세한 설명을 용어로 표현하는 것으로 구술자료를 시스템화 하기 위하여 해당 구술컬렉션에 공통적으로 필요한 용어를 추출하고 용어 간의 관계를 도출하여 작업을 수행한 결과라고 볼 수 있다(그림 4 참조)²⁾. 개별 구술레코드의 세그먼트 구분 및 주제, 인물,

²⁾ VHA의 색인은 USC SHOAH의 구술프로젝트에 적합한 색인 용어를 추출하여 자체

시대, 장소별, 색인화 작업 뿐 아니라 레코드들을 통합하여 DB 검색의 접근점을 구축, 서비스하고 있다. 색인어의 집합은 이용자에게는 검색의 접근점으로 제공하기 때문에 적합한 구술자료를 신속하게 찾을 수 있게 된다.



〈그림 4〉 USC SHOAH Visual History Archive 색인어 검색 화면

3) 구술아카이브³⁾ 이용 및 접근에 관한 연구 사례 (OHA: Oral History Association)

구술사협회(OHA)에서는 구술컬렉션의 내용과 맥락들을 대중들이 검색하고 접근할 수 있는지를 파악하기 위하여 2016~2019년에 총 2차례에 걸쳐

개발하였음

- 3) 구술채록사업을 하는 기관이나 단체에서 구술자료를 생산, 수집 및 기증을 통해 구축한 아카이브를 ‘구술아카이브’라고 한다. 과학기술 및 정보통신기기의 발달과 더불어 구술아카이브가 발전하게 되었다(녹음 기술, 텍스트 전환, 음성 및 영상 구술채록을 디지털 데이터로 저장 및 보존). 기록학 용어사전에 의하면, 아카이브는 ① 개인이나 조직이 사적으로 또는 공적으로 생산하거나 접수한 기록 중에서 역사적으로 보존할 가치가 있거나 증거로서 보존할 필요가 있다고 평가, 선별된 영구보존기록물들이다. ②영구 보존기록물을 전문적으로 보존하는 조직 또는 이를 위한 시설 및 장소를 말한다.

연구를 수행하였다(Oral History Association 2020). 북미 및 유럽 국가의 사서, 구술역사가, 아키비스트, 지역사회 프로젝트 관리자, 언론인, 영상기록자, 미디어 전문가 등 실무자 58명 대상으로 설문조사 및 인터뷰를 하였다. 연구 결과, 구술자료를 디지털 저장소에 보관하고 아카이브를 운영하는 기관 실무자의 수가 증가하고 있으며, 콘텐츠와 맥락에 대한 접근을 위하여 상세한 접근 방식이 필요하다고 하였다. 주로 사용하는 표준과 도구는 Dublin Core, 자체 개발 시스템, DACS, LCSH, MARC, EAD 등의 순이었다. 아카이브 시스템 및 메타데이터가 일반적으로 도서관 및 아카이브에서 사용하는 범용적인 시스템 보다는 구술자료의 특별한 성격과 지역적 임무, 목표, 자원의 균형을 맞추는 시스템 구축과 접근 방식이 필요하다고 하였다.

인터뷰 맥락에 대한 중요한 정보를 기록하기 위해서는 인터뷰 이전부터 종료되고 이관되는 전 과정에 모두 일어나야 한다고 하였다. 메타데이터를 조기에 적용하지 않으면 중요한 상황별 정보가 손실될 수 있다고 지적하였다. 다시 말하면 구술 자료에 대한 설명(메타데이터)의 기록은 주로 인터뷰 이후에 수행되지만, <표 1>과 같이 구술자료 라이프사이클 전 단계에 걸쳐서 메타데이터 요소를 수집해야 정밀한 검색의 접근점을 제공하여 활용할 수 있다고 하였다.

위에서 살펴본 바와 같이 해외 구술아카이브 서비스의 경우는 다음과 같이 요약해 볼 수 있다. 첫째, 구술채록 프로젝트와 연계하여 도서관에서 전문화된 검색 및 보존과 이용서비스 시스템을 운영하고 있으며 아날로그 구술채록에 대한 디지털 복원화와 디지털아카이브 시스템에 편입하는 작업을 병행하고 있다. 둘째, 구술자료 이용에 대한 법적인 저작권을 양도 처리하고, 학술 연구지원을 위한 제반 서비스를 제공하고 있다. 셋째, 구술자료 검색을 위하여 컬렉션 중심으로 개별 구술인터뷰가 집중될 수 있도록 목록화하고 연관 레코드를 연계하고 있다. 넷째, 표준화된 인코딩 방식⁴⁾과 메타데이터 작성을 통해 인명, 주제, 유형, 프로젝트, 컬렉션별로

정보 접근점을 제공하고 있다. 다섯째, 구술자료 라이프사이클 단계별로 메타데이터 요소를 부여하여 맥락 정보를 제공하고 있다.

4. 국내 구술자료 관리와 온라인 구술아카이브 서비스 현황

1) 국내 구술자료 관리 현황

국내 구술프로젝트는 1980년대 말 사회단체 중심으로 과거 역사의 진상 규명을 위한 구술 증언을 수집하면서 시작되었다. 1990년대부터는 구술 채록이 국가기관으로 확대되었고, 일제강점기 강제동원, 제주 4.3사건 등 문헌 기록에 남겨지지 못한 근현대 주요 사건에 의한 구술 자료들이 생산되었으며, 2000년대 후반에는 지역의 문화와 정체성을 복원하기 위한 지역과 마을 관련 구술 사업이 증가하기 시작했다.

그러나 배은희(2018), 선영란(2019), 윤택림(2023)의 연구에 의하면, 구술 채록 원본이 축적되지 못하고 대부분 단행본으로 출간하고 있으며, 구술자료를 외부에 온라인 및 열람서비스를 제공하는 기관은 소수에 불과하다고 지적하고 있다. 또한 자체 기관의 기본 목록은 구비되고 있으나 통합하여 구술 목록을 검색할 수 있는 플랫폼은 미비하다고 연구한 바가 있다 정유진(2022)은 연구를 수행하면서 신문기사, 구술채록 원본 및 단행본, 지방자치 역사기록 단행본, 통계자료 및 등으로 지역사에 있어 유의미한 자료들이 여기저기 흩어져 제대로 활용되지 못하는 상황이 답답하다고 하였다.

이와 같은 배경에는 구술프로젝트의 목적이 시대적 사건의 진상 규명과

4) 예를 들어 음악의 경우에는 음악 장르(예: 재즈 컬렉션인 경우 스윙, 비밥, 모던, 쿨, 핫, 리얼, 보사노바, 랙타임 등)를 미리 구분해서 표준화된 메타 정보를 데이터화 하면 추후 음악 장르별로 검색이 가능함

지역 역사 서술에 머물러 있기 때문에 대부분 단행본으로 간행되고 있으며, 연구 혹은 콘텐츠 활용으로서의 원본 구술채록을 관리하고 외부에 공개하여 정보서비스로 제공하는 경우는 매우 희박하다고 볼 수 있다. 윤택림(2023)은 구술자료 공개는 저작권법과 개인정보보호법과 관련이 있다고 설명하면서, 법적 자문과 더불어 구술자료 서비스를 위한 <구술자료 외부 이용동의서>가 필요하다고 하였다.

구술 채록기관 뿐 아니라 학계 및 연구자들은 이와 같은 문제점을 인식하고 정책적인 방향을 연구한 바가 있다. 국사편찬위원회에서는 2018년과 2019년에 걸쳐 각각 전국 각 기관 및 공공단체, 학회 및 연구소, 민간기관, 출판사, 개인 연구자의 구술자료 수집 및 출간 현황을 조사하였다(국사편찬위원회 2018, 2019). 중앙 및 지방기관, 공공단체들이 구술채록 사업을 하는 기관은 78개관이며, 민간기관의 경우 22개 기관, 대학 및 연구소는 29개 기관으로 나타났다. 이 조사 결과는 구술채록 생산물에 대한 기관별, 연도별, 주제별 공동목록집⁵⁾으로 활용될 수 있으나 일회성으로 출간과 데이터 처리가 불가능한 문서 형태로 공개되고 있어 온라인 검색 접근점으로 활용되지 못하는 한계가 있다.

반면 구술아카이브를 종합적으로 관리할 수 있는 국가 차원의 센터를 설립하고 운영하고자 하는 연구를 수행하였는데, 국가기록원(2007)은 『구술기록물종합관리체계방안 및 구술방법론』에서 국가기록원 산하 ‘구술사아카이브’설립을 제안하였으며, 국립중앙도서관(2019)은 『구술자료수집 타당성 및 제도개선 방안 연구』를 통해 국립중앙도서관 산하 ‘한국구술자료 통합정보센터’ 설립 제안을 한 바가 있다.

5) 기관별로 주제, 구술자, 면담자, 생산연도, 구술시간, 공개여부, 생산기관, 지원기관에 대한 요소 항목을 조사하였다.

2) 국내 온라인 구술아카이브 시스템 서비스 현황 분석

국내 기관에서 구술 원자료를 온라인으로 서비스하는 대표적인 기관인 국회기록보존소, 국사편찬위원회, 한국학중앙연구원, 아르코예술기록원, 경기도사이버도서관 등 5개 기관을 선정하여, 이 기관들의 온라인 구술자료 서비스를 분석하였다. 구술 주제, 온라인 목록 형태, 구술 컬렉션 저작별 연계, 검색 서비스와 자료 이용서비스로 구분하여 살펴보았다. 특히 검색 서비스는 기본 목록, 내용 구분 검색, 색인 검색, 주제 및 키워드 검색, 인명 검색, 구술인터뷰 맥락(구술 장소, 면담 후기 등)에 관한 내용을 검색할 수 있는지를 조사하였다. 국내 온라인 구술아카이브시스템 정보서비스 현황분석은 <표 2>와 같다.

첫째, 구술자료 온라인서비스를 제공하는 5개의 조사 대상 기관의 구술자료는 구술자 인터뷰 단위의 기본 목록과 브라우징 검색이 가능하다. 그러나 기관의 구술자료는 구술자별로 채록된 과정과 내용을 검색할 수는 있어도 주제, 시대, 지역 등의 특정 컬렉션별로 개별 저작들 간의 연계되어 있는 기관은 없었다. 이 현상은 개별 구술자료별로 메타데이터와 원자료는 매핑 기술되어 있으나, 구술프로젝트가 카테고리별로 저작들이 집합될 수 있도록 설계 및 데이터화가 되어 구축되어 있지 않았음을 의미한다. 표준메타데이터 요소를 기반으로 개별 구술자료들의 데이터가 통합될 수 있도록 시스템 모델링 설계가 이루어지고 데이터가 구축이 되어야 주제별, 시대별, 인물별 전체 조망과 연관 구술자료까지 확장해서 검색할 수 있다.

둘째, 구술자료 내용에 대한 검색 기능을 살펴보면, 개별 구술채록 레코드별로 텍스트 검색이 가능하였던 기관은 한국학중앙연구원과 국회기록보존소였다. 한국학중앙연구원은 개별 구술채록의 녹취록에서 자연어로 검색이 가능하였으나, 구술자료들의 통합 검색은 제공되지 않았다. 국회기록보존소는 내용을 세그먼트로 구분하여 사건, 지명, 인명 등이 색인되어 있었으나

〈표 2〉 국내 온라인 구술아카이브 시스템 정보서비스 현황 분석

기관	온라인 서비스명	구술 주제	온라인 목록 (브라우징)/ 컬렉션 및 프로젝트 정보 제공	컬렉션별 저작 연계	온라인 검색 서비스							구술자료 온라인 서비스			기관 방문 이용
					구술 단위 기본 목록 검색	내용 색인 검색	주제, 키워드 검색	인명 검색	인터뷰 관련 매체 목록	구술 원자료 온라인 이용	온라인 이용 동의서	인용방법 제공			
국회기록 보존소	구술기록	국회의장단 등 국회 관련 주요 인사	○	×	○	×	○	○	○	○	○	×	×	×	
국사편찬 위원회	전사자료관 (구술자료)	한국근대사의 인물, 정책, 사건 관련 인물	○	×	○	×	×	×	×	×	×	×	×	○	
한국학중앙 연구원	현대한국 구술자료관	현대 한국의 민주화·산업 화 인물	○	×	○	×	△	○	○	○	○	×	×	×	
아르코예술 기록원	한국예술디지털 타이카이브: 구술채록	한국근현대사 원로예술 인물	○	×	○	×	×	×	○	○	○	△	○	○	
경기도사이 버도서관	경기도메모리	경기도민이야기 '도서관과 나'	△	×	○	×	×	×	×	×	×	×	×	×	

내용 구분 세그먼트를 ‘영상별 검색’으로 명명하여 그 의미를 이해하기 어려웠으며, 색인어로 검색하여 접근할 수 있는 방법이 없어 활용이 용이하지 않았다. 반면, 아르코예술기록원, 국사편찬위원회와 경기도사이버도서관은 내용 검색이 제공되지 않았다.

셋째, 조사 대상 기관이 기술 수준은 상이하더라도 구술자별로 상세목록, 인물, 사건, 해제, 주제어 등의 메타데이터를 구축해 놓았으나 검색 효용성이 있는 메타데이터를 통합하여 색인화시키고 검색이 될 수 있게 서비스로 개발하고 있지 않았다. 각 기관들의 메타 공통요소들을 추출하여 통일된 표목으로 설정하고 데이터를 통합하여 검색할 수 있는 시스템을 만든다면 이용자 관점에서 효율적으로 구술자료를 검색하여 활용할 수 있을 것이다. 즉 개별 구술 레코드의 색인어를 통합하여 통합 색인 데이터베이스를 생성하고 색인어 집합을 카테고리화 하여 접근할 수 있게 인터페이스를 구축한다면, 주제 키워드로 접근하여 관련 구술레코드를 접근할 수 있는 효율적인 검색시스템이 될 것이다.

넷째, 구술자료의 온라인 서비스를 살펴보면, 원자료를 획득하는 방법으로 온라인에서 직접 이용할 수 있는 기관은 3개 기관이었으며, 2개 기관은 구술자료의 목록 검색 후 직접 기관 방문 이용서비스를 제공하였다. 온라인으로 이용하는 경우 이용자들에게 온라인 이용동의서와 구술자료의 인용 방법을 제공하는 기관은 1개 기관이었다. 구술자료의 내용적인 특성상 저작권과 개인정보보호에 대한 법률적인 문제를 고려해야 한다. 이를 위해서는 온라인 서비스의 이용자들에게 교육 및 연구 등에 대한 제한 범위 등을 고려한 〈온라인 이용동의서〉와 구술자료의 학술적인 인용 방법론을 제공하는 것이 필요하다.

3) 국내 도서관의 구술자료 수집 및 검색 서비스

국립중앙도서관(2019)과 정연경(2022)에 따르면 구술자료수집의 경험이 있는 국내 공공도서관은 파주공공도서관⁶⁾, 노원휴먼라이브러리, 영종도서관, 부전도서관, 용학도서관, 평택시립도서관, 하남시립도서관이다. 지역 역사기록물 수집의 일환으로 지역도서관과 시민채록단을 구성하고 교육하여 구술 인터뷰 음성 및 동영상 자료 및 사진 자료와 녹취록을 생산하고 있다. 그러나 구술 원자료의 목록화 등의 정리 기술(description)은 거의 전무하여 구술 목록과 검색 및 원자료 서비스 제공이 불가능하며 대부분 구술채록 기관과 유사하게 2차 저작물인 단행본을 생산하여 도서관 목록 시스템을 통해 단행본으로 제공할 뿐이다(국립중앙도서관 2019).

국가도서관인 국립중앙도서관의 경우 최근 구술자료에 대한 수집과 생산이 시작되었다. 국립중앙도서관에서 제작한 ‘2021~2022년 동송동 구술채록’과 ‘2022년 이어령컬렉션’ 있으며 ‘신시대 21’과 ‘문화체육관광부’에서 수행한 구술채록사업 결과물인 구술자료를 수집하고 있다. 2023년에는 『구술자료 메타데이터 작성지침』을 개발하여 구술프로젝트, 구술컬렉션, 구술자, 면담자, 구술자료, 구술자료 매체 표준화 형식의 메타데이터 표준 형식을 예시와 함께 제시하고 있으며, 구술자료 관련 서식을 부록으로 수록하고 있다(국립중앙도서관 2023).

이와 같이 국립중앙도서관에서 부분적으로 구술 원자료를 수집하여 정리하고 있는 상황이며, 국내 도서관들은 공공도서관에서 프로젝트 형태의 구술채록사업을 통해 단행본 출간에 그치고 있으며, 도서관 검색시스템에 구술원자료를 축적하여 서비스하고 있지 못하고 있다. 또한 대학도서관에

6) 경기도 파주중앙도서관은 『민간기록물수집 및 관리 활용』조례에 따라 시민채록단 운영을 통해 채록한 구술기록물을 동영상 뿐 아니라 사진, 물증자료를 기증받아 디지털 기록정보서비스를 제공하기 위한 아카이브 시스템을 준비하고 있다.

서 체계적으로 구술자료를 생산하거나 수집하여 서비스를 제공하는 경우는 거의 없는 실정이다.

4) 데이터 검색과 공유를 위한 구술아카이브 시스템의 한계점과 개선 방안

위에서 살펴본 바와 같이 우리나라는 일부 구술채록기관에서 개별 구술자 중심의 자료 생산물에 대한 메타데이터를 구축하고 있으나, 기관별로 보존의 관점에서 구술 레코드 단위 기록물로 구술아카이브가 구축되어 있었다. 색인 데이터베이스에 근거한 통합검색 접근점은 마련되지 못하고 있으며 전국적으로 산재되어 있는 구술자료들에 대한 종합목록 구축과 검색시스템이 없는 실정이다.

반면, 미국의 경우 구술자료가 역사기록물 장서 구축과 연구자료 제공을 목적으로 컬럼비아, 코넬, 버클리, UCLA 등과 같이 연구중심 대학을 중심으로 도서관과 협력하여 시작되었기 때문에 서비스 중심의 목록과 검색 도구들을 통하여 접근점을 만들고 구술사를 편지, 일기, 메모 등과 같이 특별 유형의 장서로 구분하고 있다(Richie 2014). 구술프로젝트 초기부터 데이터 전문가와 협력한 시스템 설계와 메타데이터 및 색인 검색 등 이용자 중심의 검색서비스가 이루어져 왔다. 더불어 타 기관에서 생산한 구술자료를 도서관으로 이관하여 자료의 수집, 검색을 위한 정리, 보존, 이용에 대한 전문성을 활용하여 제반 부가가치적인 작업을 수행하여 다각적인 검색의 접근점을 제공하고 있음과 대비되는 상황이다.

이와 같이 국내와 해외의 구술자료의 서비스 현황을 비교 분석하여 종합해보면 국내 구술자료를 서비스하는 구술아카이브의 한계점은 다음과 같다.

첫째, 국내 구술채록 및 서비스는 학술적인 목적과 역사기록물로 구축하기 위한 구술채록과 자료 이용 및 보존이 그 목적이 아니었기 때문에 구

술성이 나타나는 구술기록물 원자료를 데이터화 하는 과정이 상당히 생략되어 있다. 둘째, 구술채록기관 중심의 구술아카이브 서비스는 주제, 시대, 유형 등 검색 중심 보다는 기술적인 구술자 레코드 단위의 메타데이터 중심의 시스템이며 아카이브의 대주제를 구술기록물에 부여하는 형태로 웹을 구축하고 서비스하고 있다. 이러한 시스템은 브라우징은 가능하나 주제 키워드와 상세 검색이 통합적으로 이루어지기 어려운 상황이다. 셋째, 구술아카이브 서비스 플랫폼이 표준화된 메타데이터 스킴과 설계로 이루어진 시스템을 운영하는 것이 아니기 때문에 기관 간의 기본 목록 데이터 교환과 공유가 어려워 통합목록과 검색이 불가능한 상황이다. 넷째, 해외의 경우, 도서관과 구술기관과의 협력으로 구술자료의 자료 수집, 생산, 정리, 보존 그리고 이용에 이르는 사이클에 데이터 가공과 정리가 지속적으로 업데이트되어 서비스를 제공하고 있다. 반면 국내는 구술채록기관이 서비스를 직접 제공하면서 생산과 수집에 중점을 두고 있었다. 구술자료 데이터 검색을 위한 구술자료의 복합적인 특성을 데이터로 만들어 낼 수 있도록 전문 인력 및 인프라의 지원과 지속성이 필요하며 도서관과의 협력이 거의 이루어지고 있지 않은 상황이다. 다섯째, 구술자료 서비스를 위한 법적, 윤리적 기반 마련이 부족한 상황이다.

이를 위한 방안으로 본 연구에서는 다음과 같은 실천 방안을 제시하고자 한다.

첫째, 단행본 중심의 구술자료 편찬을 벗어나 구술자료만이 가질 수 있는 원자료의 구술성, 맥락성, 사회적 배경이 반영될 수 있는 시스템을 구축해야 한다. 이를 위해서는 구술프로젝트의 복잡한 워크플로우가 반영되는 구술자료 라이프사이클 단계별로 구술프로젝트의 기획 및 관리, 구술자료의 사전조사, 구술채록 및 정리 영역, 저작 영역 등의 서술적 기술 뿐 아니라 구술자료 활용을 위한 서비스 영역을 포함하여 구술자료의 생산 맥락을 알 수 있도록 한다.

둘째, 구술 자료를 데이터로 검색하기 위해서는 구술컬렉션으로 군집화하고 각 저작물 간의 관계가 시스템적으로 설계되어야 한다. 표준화된 데이터 구조화 기반의 구술아카이브 모델링 설계와 데이터베이스를 구축하고 데이터 관리와 서비스가 이루어질 수 있도록 한다.

셋째, 구술 원자료를 검색을 위한 내용의 데이터화를 하는 전문화된 작업이 필요하다. 표준화된 국내외 메타데이터 스킴과 인코딩방식을 활용하고 개별 구술자료의 메타데이터 구축을 데이터 요소별로 통합하고, 통합 색인 데이터베이스와 검색시스템을 개발하는 것이 필요하다. 주제, 사건, 인물 등에 관한 구체적인 요소들의 색인 작업이 필요한데, 기존에 개발된 통제어휘 전거데이터⁷⁾를 사용하거나 특정 사건과 관련된 추가적인 용어 작업을 수행하여 지금까지 수행해 왔던 구술채록사업에 대한 범위와 내용을 포괄할 수 있는 한국 근현대사에 적합한 주제 키워드(용어)와 인물, 사건들을 포함할 수 있는 주제어 어휘를 개발하는 것도 필요하다. 기본적으로 국립중앙도서관의 주제명 전거파일을 기반으로 하는 것을 추천한다. 구술자 레코드별 색인 및 레코드 내의 세그먼트별 주제(subject heading) 통합 색인이 이루어진다면, 구술아카이브의 전체적인 주제 조망과 더불어 주제, 인명, 시대, 유형 등 데이터 중심의 검색이 가능하게 될 것이다.

넷째, 우리나라 구술자료의 소재 여부를 파악하기 위한 기초 종합목록이 필요하다. 국사편찬위원회(2018)에서 수행한 기초 조사를 중심으로 구술채록기관이 소장하고 있는 목록을 데이터로 전환하여 종합목록시스템을 구축하는 것이 시급하다. 나아가 기존의 방식인 인쇄형이 아닌 기본 표준목록 스킴에 기반하여 기관에서 업데이트하고 상호 공개하여 접근할 수 있도록 마련하는 것이 요구된다. 각 기관에서 구술채록 원자료 공개를 위한 준비 단계이더라도 기관에서 생산 및 수집된 구술자료에 대한 목록을

7) 전거데이터는 도서관 목록에서 사용되는 저자명, 서명, 총서명, 주제명을 일관성 있게 관리하고 사용하기 위해 작성된 데이터를 말한다.

공유하여 종합목록을 지속적으로 만들어갈 수 있다. 우리나라 구술자료의 주제와 소장기관을 전체적으로 조망하여 학술적, 대중적 자료 이용 접근을 도모하여 풍부한 연구와 2차 저작물로 활용될 수 있는 통로를 마련할 수 있을 뿐 아니라 중복 구술 채록을 지양하는 결과를 가져올 수 있다.

다섯째, 구술자료의 저작권과 관련하여 이용과 보존 영역, 비공개 부분 등의 법적, 기술적인 부분이 고려되어야 한다. 서비스 기관은 〈온라인 이용동의서〉를 통해 제시하는 이용 범위(예: 연구목적으로 범위 한정)를 동의 후 원자료를 활용할 수 있는 시스템과 서비스를 개발해야 한다.

여섯째, 구술자료의 통합관리와 서비스를 위한 정책적, 실무적인 협력체가 요구되는 상황이다. 앞서 조사한 바와 같이 구술프로젝트는 구술채록으로만 완성되는 것이 아니라 구술자료를 데이터화 하여 디지털 보존과 활용을 고려해야 하므로 데이터 구축과 서비스에 대한 예산과 인력이 배정될 수 있도록 정책적인 방안을 제시해야 한다. 나아가 디지털인문학⁸⁾의 큰 범주 내에 연구가 다학제적으로 이루어져야 하며, 구술채록기관과의 협력적 연구와 교육프로그램⁹⁾ 개발이 필요하다. 이를 위해서 구술사학회를 중심으로 구술채록기관의 협력체인 구술사네트워크, 국립중앙도서관 등이 협력하여 관련 연구와 논의가 지속적으로 이루어져야함을 제언한다.

8) 디지털인문학은 인문정보의 디지털데이터 수집, 데이터베이스 구축, 데이터 분석과 해석을 모두 포괄하는 것이다. 디지털 객체의 내용과 형식을 기술하고 표준화된 디지털 포맷, 데이터 교환과 보존될 수 있는 시스템으로 구축하여 디지털 매체로 표현하여 웹으로 표현하고 생산하여 문화유산으로 남기는 것을 목적으로 한다. 디지털인문학과 구술사와의 연계는 Douglas A. B., M. Larson(2014)에서도 언급하였듯이 녹취록 기반의 연구를 벗어나 구술 매체의 디지털화 데이터베이스 구축과 통합화를 통한 분석을 제안하고 있다.

9) 디지털인문학 프로젝트 기획, 디지털아카이브와 협력, 디지털플랫폼 평가, 데이터 큐레이션, 데이터 분석과 기술, 데이터 공유와 저작권 등의 교육프로그램이 필요하다.

5. 나아가며

역사는 시대의 기록물들이 선별되고 남겨지는 것으로 후대에 해석된다. 구술자료는 문자 이상의 맥락과 음성과 영상이 지니고 있는 감정과 내용을 전달하는 콘텐츠이다. 본 연구는 국내 채록기관에서 수집, 생산, 서비스하고 있는 구술아카이브 검색의 한계점을 국내외 웹 서비스를 통해 분석하고 개선 방안을 제시하였다.

구술자료는 문자 이상의 맥락과 음성과 영상이 가져다주는 감정은 현재 우리가 살아가고 있는 디지털 AI 시대의 인간이 남길 수 있는 유일한 원자료 중 하나이다. 구술자료가 디지털 시대에 구술성이 원본 그대로 표현되어 정보기술을 통한 디지털 아카이빙으로 축적되고 지속적인 이용이 이루어지는 시스템으로 개발되지 않는다면 후속세대의 역사적 기록물과 문화유산으로 활용되지 못할 것이고 역사에서 사라질 것이다.

【참고문헌】

- 국가기록원, 2007, 『구술기록물 종합관리체계 구축 방안 및 구술채록을 위한 연구』 결과보고서.
- 국립중앙도서관, 2023, 『구술자료 메타데이터 작성 지침』.
- _____, 2019, 『국립중앙도서관의 구술자료수집 타당성 및 제도개선 방안 연구』 결과보고서.
- 국사편찬위원회, 2018, 『2018년 구술자료 연구용역: 전국 각 기관·공공단체 등의 구술자료 수집 및 출간 현황 기초 조사연구』 결과보고서.
- _____, 2019, 『2019년 구술자료 연구용역: 전국학회·연구소·민간기관·출판사·개인등의 구술자료 수집 및 출간, 구술콘텐츠 구축 현황 기초 조사』 결과보고서.
- 배은희, 2018, “구술기록의 생산과 관리 현황 및 개선방안: 부산지역을 중심으로,” 『한국기록관리학회지』 18(2): 85-108.
- 선영란, 2019, “정부기관과 공공단체가 수집한 구술자료의 현주소,” 『한국구술사학회 창립10주년 국제학술대회 자료집』, pp. 125-135.
- 이정연, 2009, “구술사기록물 아카이브 구축을 위한 메타데이터 모델링 및 표준 요소 개발에 관한 연구,” 『정보관리학회지』 26(1): 163-184.
- 정연경, 2022, “구술 자료와 국내 도서관,” 『국립중앙도서관 월드라이브러리 포커스』, 2022.2.19.
- 정유진, 2022, “그 많던 ‘여공’들은 어디로 갔을까: 대구 섬유산업 역사에 ‘여성 서사’의 자리 마련하기,” 『구술사연구』 13(1): 7-47.
- 최은주, 이정연, 2009, “웹 구술사료(口述史料) 아카이브 정보시스템의 설계 및 평가용 가이드라인개발에 관한 연구,” 『정보관리학회지』 26(4): 277-298.
- Douglas A. B., M. Larson, 2014, *Oral History and Digital Humanities: Voice, Access, and Engagement*, Palgrave Macmillan.
- Mackay, Nancy, 2015, *Curating oral histories: from interview to archive*, London and New York, Routledge.
- Nelson, Cyns, 2018, *Oral history in your library: create shelf space for community voice*, Oral History Association, 2020, *Oral history metadata and description: A survey of*

practices.

Ritchie, D.A., 2015, *Doing Oral History 3rd*, New York, Oxford University Press.

Rutgers University, 2010, *RU-CORE and NJDH Standards Analysis for Audio Objects*.

The Society of American Archivists, 1995, *Oral History Cataloging Manual*.

White, Krista and C. Radick., 2017, "Legacies and lifecycles: digitizing oral histories at Rutgers University libraries," *Digital Library Perspectives* 33(3): 203-220.

투고일: 2024. 02. 13. 심사일: 2024. 02. 15. 게재확정일: 2024. 02. 29

Abstract

Study on the Limits and Areas of Improvement for Data Search and Sharing in Oral History Archives

Jungyeoun Lee
(Hanyang University)

Despite the quantitative growth and expanded research interest in oral history over the past three decades, accessing and utilizing domestic oral history materials remain challenging. This study conducts a comparative analysis of domestic and international oral archive platforms to delineate the underlying causes of these challenges. The findings underscore significant disparities in database construction and search functionality between domestic and foreign platforms. While overseas platforms have effectively structured oral materials into databases and provided integrated search services, domestic archives primarily rely on linear record technology and institution-centric approaches. To facilitate the search and utilization of domestic oral history materials, several recommendations are proposed. First, the development of standardized data structuring models and database construction based on the life cycle of oral materials is advocated. Secondly, specialized indexing and the development of search systems based on integrated metadata databases are deemed necessary. Thirdly, the establishment of a comprehensive list of oral materials in Korea, coupled with mutual updating and accessibility, is recommended. This approach

enables topic-based searches, reducing institutional duplication and identifying undiscovered oral recording topics. Fourthly, the provision of an Online Use Agreement and academic citation methodologies is suggested to guide responsible usage. Finally, recognizing that oral materials require digitization for preservation and utilization, interdisciplinary research within the realm of digital humanities is emphasized, along with convergent discussions on policy and practical cooperation. In the contemporary digital AI era, they represent one of the few raw materials that encapsulate the essence of human existence. However, without the development of systems ensuring the continuous utilization of oral history materials in their authentic form, they risk fading away from the heritage of future generations.

Key words : Oral history, Archive platforms, Data structuring, Search systems, Digital humanities, Interdisciplinary research