

고고학술정보 데이터 아카이빙 전략과 구현 방안

-한국고고학사전의 디지털 전환을 위한 시맨틱 데이터 아카이브 제안-

김현*, 강혜원**

1. 무엇을 하자는 것인가?
2. 지식백과사전 디지털화의 사례
3. 디지털 전환의 시대, 새로운 지식백과의 요건
4. 『한국고고학사전』 시맨틱 데이터 예시
5. 『한국고고학사전』 시맨틱 데이터 모델 제안
6. 고고학술정보 시맨틱 데이터 아카이브의 활용

1. 무엇을 하자는 것인가?

2022년 ‘고고학술정보 디지털 대전환’ 국제학술 심포지엄은 ‘『한국고고학사전』 등 한국고고학 학술정보의 디지털 체계전환을 위한 과제와 전략’을 논의하기 위한 것이다. 이러한 논의가 시작될 때, 그 배경에 대한 이해가 깊지 않은 이들이 먼저 제기하는 질문은 ‘이미 많은 것이 디지털화 되어 있지 않은가?’라는 것이다. 국립문화재연구원에서 간행한 『한국고고학사전』, 『한국고고학전문사전』만 하더라도 온라인상에서 내용 검색이 가능한 수준으로 서비스되고 있고, 전국 각처에 산재한 고고 유적지, 유물산포지의 발굴조사보고서는 문화재청의 관련 사이트에서 PDF 파일로 다운받을 수 있으며, 고고학 분야의 학술지 게재 논문이나 학위논문, 전문연구서에 관한 정보는 한국연구재단, 교육학술정보원, 국립중앙도서관 등의 데이터베이스에서 찾아볼 수 있다. ‘디지털 전환’(Digital Transformation)을 디지털화(Digitalization)나 전산화(Digitization) 수준으로 이해하는 사고에서는 ‘고고 학술정보’는 이미 활용하기에 벽찰 정도로 디지털화 되어 있다. 여기에 무엇을 더하자는 이야기인가?

전산화(Digitization)
객체의 물리적인 형상이나 속성이 디지털 신호로 표현될 수 있도록 하는 것. ☞ 백과사전 텍스트 전산 입력 / 아카이브 물품의 디지털 사본 제작
디지털화(Digitalization)
전산화(Digitization)의 결과를 활용하여 예전부터 해 오던 일의 프로세스를 디지털 기반으로 효율화 하는 것 ☞ 백과사전/아카이브 콘텐츠의 온라인 서비스 / 데이터 관리 시스템 운영 ※ 디지털화는 기존의 프로세스를 효율화할 뿐, 새로운 것으로 변화시키지는 않는다.
디지털 전환(Digital Transformation)
디지털 기술에 의해 확장된 지평 위에서 프로세스 자체를 변화시키는 것. 아울러 그 새로운 프로세스가 작동하도록 옛것에 매이는 고정관념을 깨뜨리고, 조직과 문화를 개선하는 것 ☞ 백과사전과 아카이브의 구획을 넘어서는 ‘백과사전적 아카이브’의 구현

* 한국학중앙연구원 인문정보학 교수, 디지털인문학연구소장

** 한국학중앙연구원 디지털인문학연구소 연구원, 문화유산 영문해설문 편찬연구팀장

이 세 가지 용어가 각각 무엇을 뜻하는지는 상론하지 않기로 한다. 발표자가 이 발표의 모두에서 강조하고 싶은 것은 이것들을 혼동하는 사고 위에서는 ‘디지털 전환’(Digital Transformation)에 대한 발전적인 논의가 가능하지 않다는 것이다.

2. 지식백과사전 디지털화의 사례

『한국민족문화대백과사전』은 발표자가 재직하고 있는 한국학중앙연구원의 대표적인 저작물이다. 1980년에 편찬사업을 시작하였고 1991년에 초판 27권(본책 25권, 부록 2권)을 책자로 간행하였으며, 2001년에 오프라인 디지털 저작물(『디지털 한국민족문화대백과사전 EncyKorea』: CD-ROM 5장, DVD 1장)의 형태로 1차 개정판을 간행하였다. 이를 위해 종이 매체에 실렸던 『한국민족문화대백과사전』의 내용을 디지털 매체에 옮겨 담은 과정은 ‘전산화(Digitization)’ 단계의 일이었다고 할 수 있다.

『한국민족문화대백과사전』 콘텐츠가 ‘전산화(Digitization)’를 통해 디지털 매체에 실을 수 있게 됨에 따라 그 활용성을 증진시키는 여러 가지 후속 사업이 가능해졌다. 인터넷의 확산에 따라 온라인 서비스 콘텐츠를 확보하는 데 관심이 있었던 인터넷 포털을 대상으로 데이터를 판매할 수 있었고(2007~현재), 이를 기반으로 자체 온라인 서비스 홈페이지를 구축하였으며(2012), 데이터의 지속적 유지/관리를 위해 백과사전 데이터 관리 시스템을 개발하고 운영하기 시작했다. 이것은 백과사전의 출판·보급과 지속적 내용 증보 등의 업무를 컴퓨터 시스템상에서 이루어지게 한, 이른바 디지털화(Digitalization) 단계의 일이다.¹⁾ 『한국민족문화대백과사전』의 편찬·보급을 담당하는 한국학중앙연구원 한국학사전편찬부의 디지털 관련 업무는 현재까지 이러한 성격의 일에 치중하고 있다.²⁾

‘전산화’(Digitization)와 ‘디지털화’(Digitalization)가 ‘디지털 전환’(Digital Transformation)의 선행과정인 것은 분명하지만, ‘디지털화’(Digitalization)를 열심히 한다고 해서 저절로 ‘디지털 전환’(Digital Transformation)의 단계에 올라서게 되는 것은 아니라는 사실도 인식해야 한다. 한국학중앙연구원의 담당 부서는 40여 년 전에 만들어진 고전적인 지식백과사전의 체제

1) 『한국민족문화대백과사전』 편찬·보급에 관한 제반 업무의 디지털화를 가능하게 한 재원은 한국학중앙연구원이 2007년말부터 2017년말까지 수행한 ‘한국민족문화대백과사전 개정증보’ 사업(한국학진흥사업)을 통해 조달되었다. 이 사업의 중점과제는 백과사전 항목을 증보하고, 내용을 개정하는 것이지만, 이러한 업무를 체계화·효율화 하는 취지에서 디지털 기반의 편찬 시스템 및 서비스 시스템 개발 업무를 개정증보 사업 과제로 편입하여 수행하였다.

※ 한국민족문화대백과사전 개정증보(2007~2017) 사업을 통해 수행한 백과사전 편찬 디지털화 주요 과제

연도	디지털화 과제	내용
2008	‘원고 관리 시스템’	백과사전 항목 원고를 집필자로부터 수합, 검수하는 과정의 관리
2008	‘멀티미디어 통합 관리 시스템’	백과사전의 삽도로 쓰일 사진 자료의 관리
2009	‘편찬 시스템’	전자문서화 되어 있는 항목 원고의 편집 관리
2010~2013	‘민백 온라인 서비스 시스템’	백과사전 콘텐츠를 온라인상에서 서비스 (http://encykorea.aks.ac.kr)
2014	‘텍스트 콘텐츠 관리기 개발’	원고 수합·검수 관리 및 전자문서 수정 등의 일관 처리
2015~2016	‘온라인 집필 시스템’	집필자가 수탁받은 항목 원고를 온라인상에서 집필, 수정
2016	‘초판본 종이 사전 검색 서비스 시스템’	초판본 종이 사전 e-Book 서비스

※ 한국민족문화대백과사전 개정증보(2007~2017) 사업의 디지털화 관련 예산 금액

와 운영 프로세스가 디지털 세상에서 여전히 작동되도록 하는 수많은 과업을 안고 있으며, 그 틀에서 벗어나는 일에 신경쓸 여유를 갖지 못하고 있다.

업무 수행체계가 제도적으로 잘 정립되고 그 일에 능숙한 운영 인력을 충분히 보유하고 있는 곳에서는 디지털화(Digitalization)도 용이하다. 기술적인 지식이 없다고 해도, 무엇을 어떻게 하고 있는지를 알려주는 현업의 모델이 있으니, 용역회사의 프로그래머들이 현업의 일거리가 컴퓨터상에서 처리되도록 하는 소프트웨어를 개발하는 것이 가능하다. 하지만 이러한 조직 일수록 그 다음 단계인 ‘디지털 전환’(Digital Transformation)으로의 이행이 어려운 것도 사실이다. 아직도 하지 못한 디지털화(Digitalization)의 과제가 너무 많은 데다가 그 일을 용역의 형태로만 수행해 왔기 때문에 스스로 디지털 기술의 새로운 가능성을 모색하는 것을 주저하기 때문이다.³⁾

3. 디지털 전환의 시대, 새로운 지식백과의 요건

‘디지털 전환’(Digital Transformation)은 옛것을 옛 모습 그대로 디지털 세계에 옮겨 놓는 것이 아니다. 디지털이 아니던 시절에는 불가능했고 그래서 생각해 본 적도 없지만, 이제는 디지털이기 때문에 가능해진 새로운 일들이 실제로 이루어지도록 하는 것이다.

『한국민족문화대백과사전』과 같은 지식백과의 편찬·서비스 영역에서 추구할 수 있는 ‘디지털 전환’(Digital Transformation)은 여러 가지 각도에서 논의될 수 있겠으나, 발표자는 그것의 우선적인 과제를 ‘아카이브의 원천 자료’와 소통하는 ‘백과사전’의 구현이라는 목표를 상정하였고⁴⁾ 이 새로운 개념의 백과사전적 데이터 아카이브의 요건을 다음과 같이 제시하였다.

연도	시스템 개발	데이터베이스 구축	멀티미디어 데이터 제작	전산장비 (HW/SW) 도입	연간사업비 (5.64억)내 비율
2008	83,000,000	0	0	0	15%
2009	34,000,000	0	0	0	6%
2010	0	83,000,000	0	10,000,000	16%
2011	20,000,000	99,980,000	61,300,000	5,500,000	33%
2012	16,702,000	91,900,000	58,978,000	15,907,000	33%
2013	10,000,000	67,700,000	60,024,000	1,200,000	25%
2014	65,000,000	0	40,000,000	0	19%
2015	62,500,000	0	34,000,000	1,000,000	17%
2016	47,000,000	0	34,020,000	3,644,000	15%
2017	55,000,000	40,500,000	24,000,000	11,000,000	23%

출처: 한국학중앙연구원, ‘한국민족문화대백과사전 개정증보’ 사업 보고서 (2010, 2013, 2017)

- 예컨대, 『한국민족문화대백과사전』 전체 항목(73,000여 항목)을 전수 검토하는 업무(2018~2020)를 위해 ‘민백 전수 검토 시스템’을 개발·운용하는 것 등
- 발표자는 2004년부터 2011년까지 한국학중앙연구원의 인문정보학연구실장과 한국학정보센터소장을 역임하면서 『한국민족문화대백과사전』의 디지털화를 현장에서 이끌었지만, 한국학정보센터(현 한국학사전편찬부)와 같이 예전부터 해 오던 일을 유지해야 하는 책임이 크고, 그에 따르는 대부분의 편찬 실무를 외주용역에 의존하는 사업부서에서는 ‘디지털 전환’(Digital Transformation) 수준의 변화를 일으키는 것이 어렵다는 사실도 알게 되었다. 이후 발표자는 백과사전적 아카이브 구축을 위한 실험의 무대를 한국학대학원의 인문정보학 랩(현 디지털인문학연구소)으로 옮겨서, 구체적인 데이터 모델 개발과 함께 그 일을 수행할 수 있는 인력의 양성에 주력했다.
- 어느 동일한 문헌자료 한 권을 두고 볼 때, 그 책에 대한 해제와 관련 인물에 대한 정보가 한국학중앙연구원 『한국민족문화대백과사전』 데이터베이스에 있고, 그 책 실물의 서지사항과 원문 이미지 데이터는 한국학중앙연구원의 장서각 디지털 아카이브에 있는데, 이 두 조직의 데이터베이스는 서로에 대해 어떠한 연결고리도 가지고 있지 않다. 책자 형태의 저작물을 만들어내던 시절부터 그 두 가지

§ 데이터 기반 지식백과사전의 요건⁵⁾

①	데이터 네트워크 형태의 백과사전	분과학문적 전문지식의 내용들이 그것이 다른 사실과 문맥의 연관성에 따라 다방면으로 이어지고 확장될 수 있는 다학문적 지식의 네트워크
②	'지식'과 '자료'를 넘나드는 융합 콘텐츠	지식의 사실적 증거 (문화유적, 박물관 유물, 문헌기록, 현장조사 기록 등)와 그 증거에서 파생한 이론과 해설 사이를 자유롭게 넘나들 수 있게 하는 융합 콘텐츠
③	가상현실에서 지식의 현장을 느끼는 감성적 체험 공간	탐구의 대상과 관련이 있는 현장의 장소감을 공유하고, 그와 관련이 있는 사실을 시각적으로 체험할 수 있게 하는 가상현실 기반의 하이퍼미디어 콘텐츠
④	집단 지성의 기여로 확장되는 개방적 데이터	디지털 자원을 접하는 지식의 소비자가 곧바로 자원에 관한 새로운 지식의 생산자 역할을 할 수 있게 하는 개방적인 소통 및 재이용가능(reusable) 디지털 자원

4. 『한국고고학사전』 시맨틱 데이터 예시

발표자는 앞에서 제시한 ‘백과사전적 데이터 아카이브’의 구상이 국립문화재연구원에서 계획하는 『한국고고학사전』의 디지털 전환 모델로서 검토될 필요가 있다고 제안한다.

이 데이터 모델이 현재의 『한국고고학사전』이나 『한국민족문화대백과사전』 데이터베이스와 다른 점은 지식의 조각들을 모아놓는 데 그치는 것이 아니라, 그것들이 서로 어떤 관계가 있는지를 명시적으로 보여줄 수 있다는 점이다. 이러한 기능을 갖도록 만들어진 데이터베이스를 ‘시맨틱 데이터베이스’ 또는 ‘시맨틱 데이터 아카이브’라고 한다. 여기서 시맨틱(Semantic, 의미 기반)이라는 말은 개개의 데이터가 서로 내용과 의미 면에서 어떠한 관계에 있는지를 알게 한다는 뜻이다.

일은 별도의 조직에서 다른 연구진들에 의해 수행되어 온 일이었기 때문에 디지털 기반으로 일을 하는 현재에도 서로에 대해 무관한 듯한 예전의 체계를 그대로 답습하고 있는 것이다. ‘백과사전 편찬’과 ‘아카이브 운영’을 각기 다른 업무로 취급하는 기존 조직체계의 제약을 넘어서서 유관한 지식자들이 온라인상에서 정밀하게 연결될 수 있는 새로운 개념의 지식정보 아카이브를 구현하자는 것이 ‘백과사전적 아카이브’의 취지이다.

- 5) 김현, ‘데이터 기반 인문지식 백과사전: 기존의 백과사전과 무엇이 다른가?’, http://dh.aks.ac.kr/Encyves/wiki/index.php/데이터_기반_인문지식_백과사전

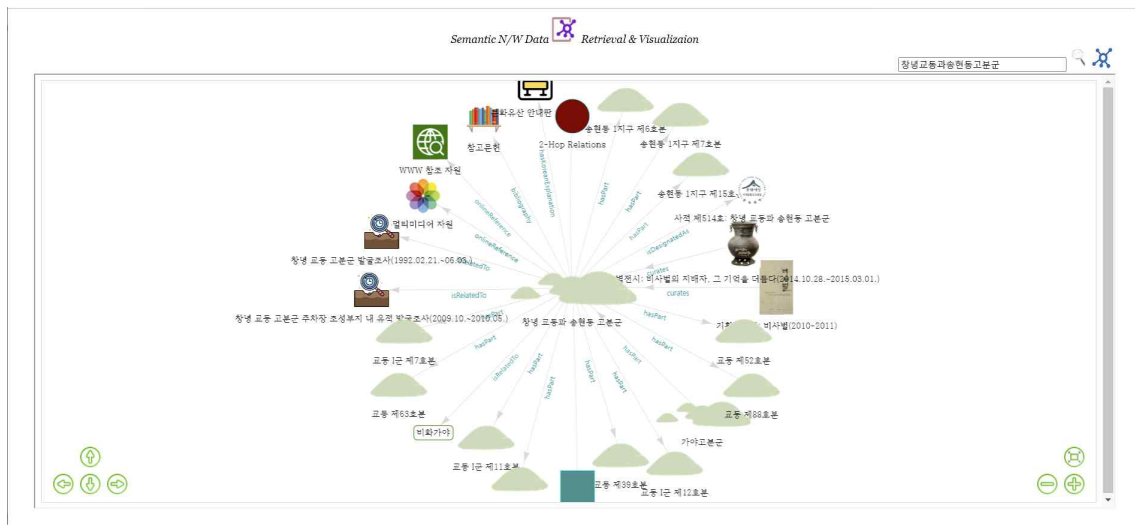
김현, ‘백과사전적 아카이브’, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016, 204~208쪽.

시맨틱 데이터 아카이브(Semantic Data Archive)

- 대상 객체의 개별적인 속성 정보뿐 아니라 객체와 객체 사이의 관계에 대한 정보를 데이터로 기록함으로써 대상 정보의 의미적 맥락을 파악할 수 있게 하는 지식 데이터 아카이브
- 문화재, 인물, 장소, 사건, 문헌, 개념 등의 데이터를 개별 정보요소로 파악하기보다는 그것이 어떠한 역사적·문화적 맥락 속에 있는지를 명시적으로 보임으로써 그 요소와 맥락정보를 활용한 '디지털 스토리텔링'이 이루어질 수 있게 한다.

§ 『한국고고학사전』 시맨틱 데이터 아카이브 예시

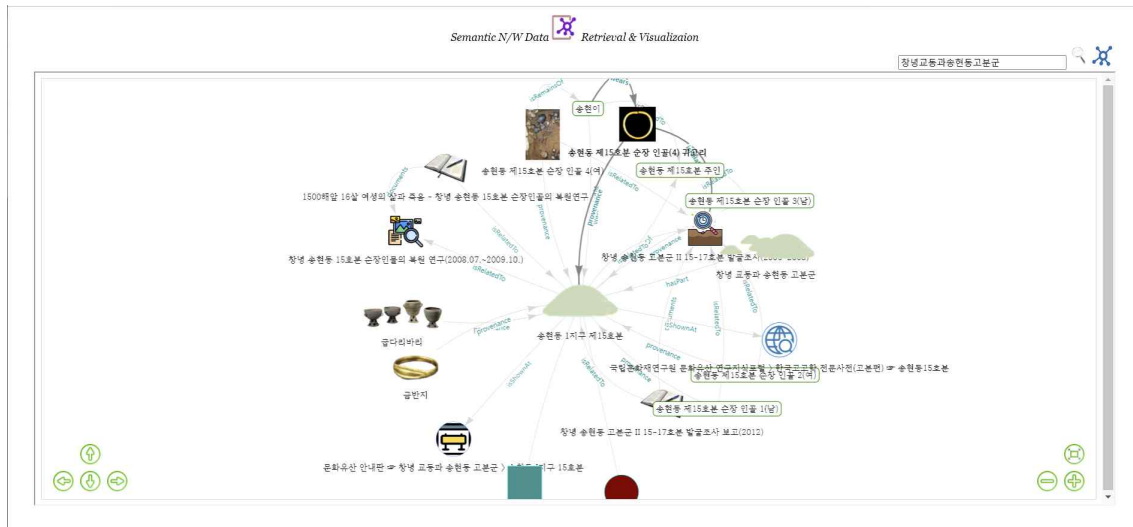
아래의 예시는 ‘창녕 교동과 송현동 고분군’에 관한 여러 가지 유형의 지식정보를 시맨틱 데이터로 구현한 것이다. 고분군과 개별 고분에 대한 사전적 해설 데이터(『한국고고학사전』/『한국고고학전문사전』 데이터)뿐 아니라, 그곳에서 발굴된 유물, 발굴 활동, 발굴조사보고서, 학술논문, 유튜브 영상과 같은 웹 자원 등 내용적으로 관련이 있는 정보들을 연계하였으며, 이 유적과 유물의 역사적 배경과 고분 발굴을 통해 알게 된 사실 등에 관한 정보도 이 데이터 아카이브를 통해 접근할 수 있게 하였다.⁶⁾



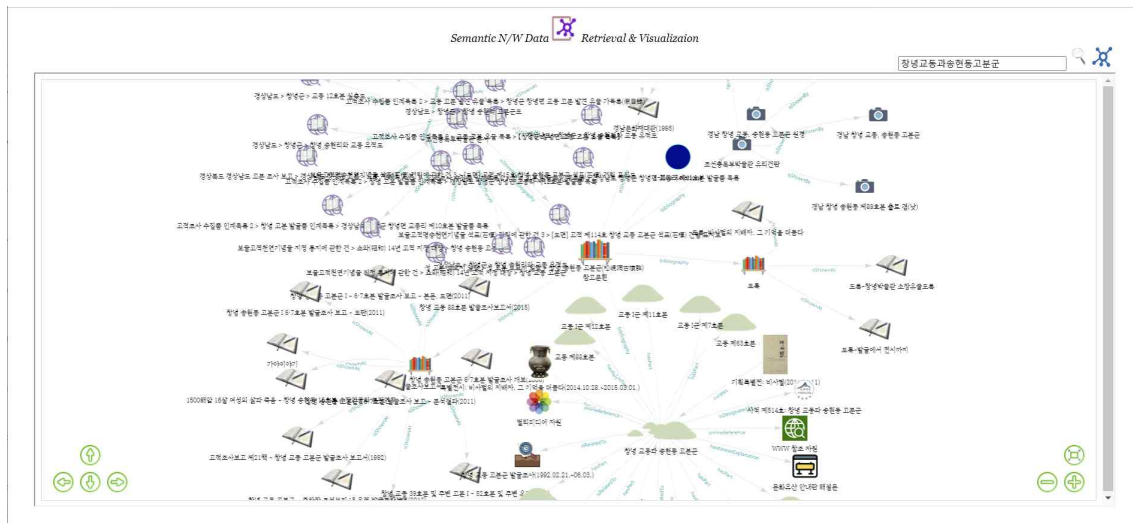
【예시 1】 창녕 교동과 송현동 고분군⁷⁾

6) 이 예시에 보이는 시맨틱 네트워크 데이터는 한국학중앙연구원 디지털인문학연구소에서 개발한 “시맨틱 네트워크 데이터 관리기(Semantic Network Data Management Studio, SN-DMS)”를 활용하여 제작하였다. (자료 조사 및 시맨틱 데이터 편차: 강혜원)

7) <http://dh.aks.ac.kr/~sandbox/cgi-bin/GuestQuery.py?db=tutor&project=GStombs&key=창녕교>
동과송현동고분군



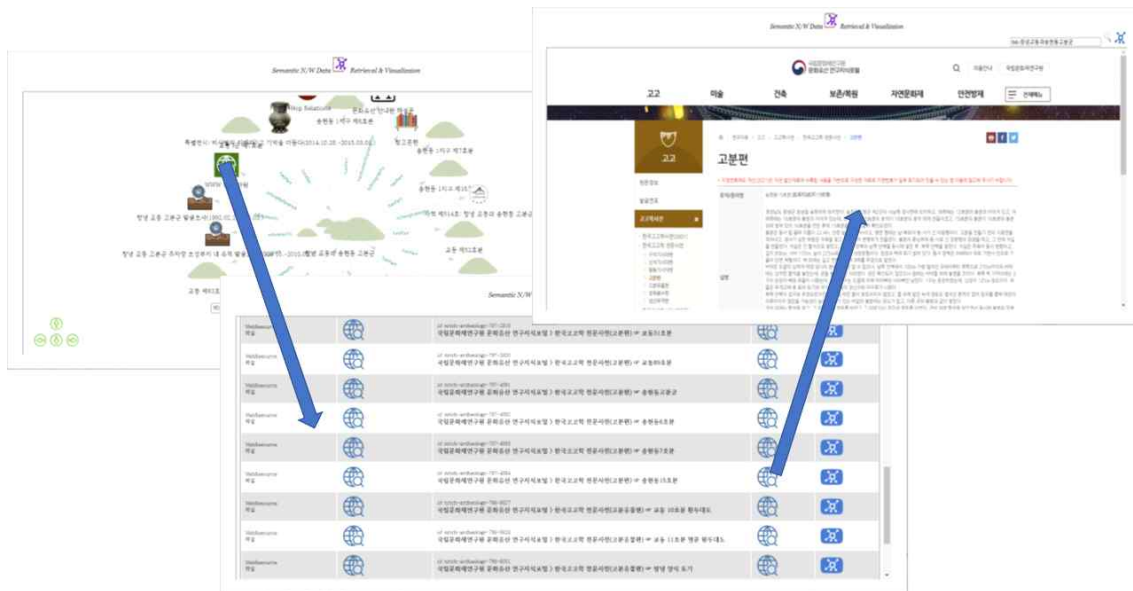
【예시 2】송현동 제15호분⁸⁾



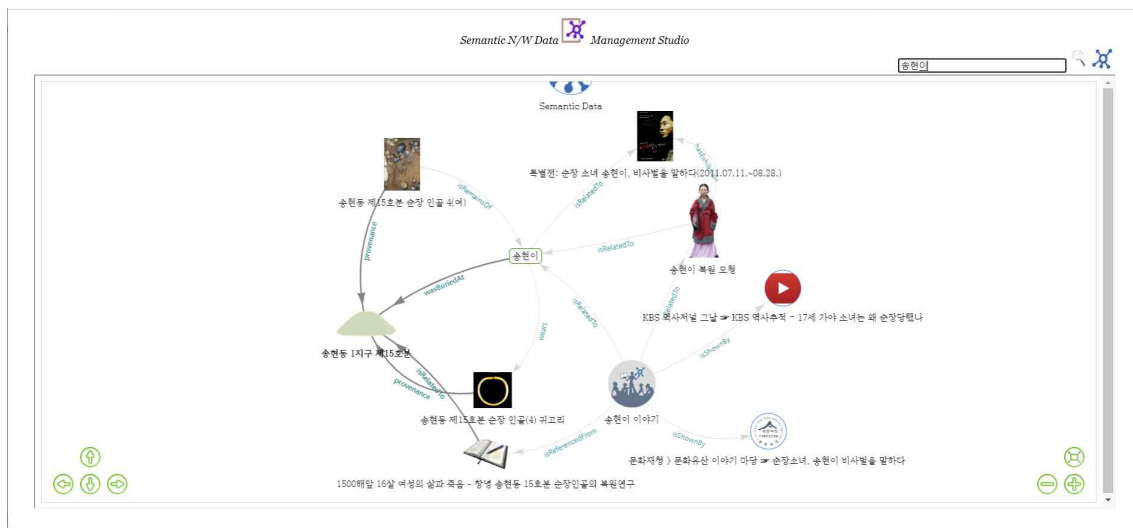
【예시 3】창녕 교동과 송현동 고분군 참고문헌(발굴보고서, 기록사진, 도록, 고고학사전 등)⁹⁾

8) <http://dh.aks.ac.kr/~sandbox/cgi-bin/GuestQuery.py?db=tutor&project=GStombs&key=창녕교동과송현동고분군-송현동1지구15호분>

9) <http://dh.aks.ac.kr/~sandbox/cgi-bin/GuestQuery.py?db=tutor&project=GStombs&key=bib-창녕교동과송현동고분군>



【예시 4】 현행 한국고고학전문사전 데이터 연계¹⁰⁾



【예시 5】 스토리텔링 데이터: 송현이 이야기¹¹⁾

분산된 데이터를 이와 같이 서로 연결 지어서 시맨틱 네트워크를 만드는 이유는 그것이 바로 앞에서 언급한 ‘데이터 기반 지식백과사전’에 기대하는 기능을 구현하는 방법이기 때문이다.

이 예시 데이터에 포함된 모든 개별 데이터는 대부분 국립문화재연구원, 문화재청, 국립중앙박물관, 유튜브 등에서 이미 디지털 콘텐츠로 서비스하고 있는 것이다. 시맨틱 아카이브 구

10) <http://dh.aks.ac.kr/~sandbox/cgi-bin/GuestQuery.py?db=tutor&project=GStombs&key=ref-창녕고동과송현동고분군>

11) http://dh.aks.ac.kr/~sandbox/cgi-bin/GuestQuery.py?db=tutor&project=GStombs&key=송현이_이야기

축 과정에서 편찬자가 한 일은 그 개별 정보들이 서로에 대해 가지고 있는 문맥적 의미를 파악하여 그것을 명시적인 ‘관계 정보 데이터’로 기술한 것이다. 실제로 시맨틱 데이터 아카이브를 구현하는 과정에서는 연구·조사를 통해 새로운 개별 정보 콘텐츠를 만들어내는 일이 우선적으로 요구되는 경우도 있다. 하지만 그 경우에도 이미 온라인상에서 제공되고 있는 유용한 지식 정보가 있다면 그것이 적절한 문맥 속에서 활용될 수 있도록 시맨틱 자원화하는 일이 병행되도록 하고 있다.

5. 『한국고고학사전』 시맨틱 데이터 모델 제안

시맨틱 데이터를 만들기 위해서 가장 먼저 해야 할 일은 그 데이터를 담을 수 있는 틀을 설계하는 일이다. 이것을 온톨로지(Ontology)라고 한다. ‘온톨로지’란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계이다.¹²⁾ 시맨틱 데이터를 생성하는 온톨로지는 정보요소의 ‘유형’(class)과 ‘속성’(attribute)을 정의할 뿐 아니라, 그 요소들이 다른 요소들과 맺는 ‘관계’(relation)를 명시적으로 기술할 수 있도록 하고 있다.¹³⁾

온톨로지는 아카이빙 대상 자원의 성격과 아카이빙의 목적에 따라 만들어져야 하는 것이기 때문에 『한국고고학사전』의 디지털 전환을 위한 온톨로지 설계 역시 고고학술 데이터의 성격과 장기적인 활용 목표 등을 깊이 있게 검토하는 연구개발 과정을 통해 이루어져야 한다. 다

12) 넓은 의미에서는 모든 정보화의 틀이 다 ‘온톨로지’일 수 있겠지만, 정보 요소 사이의 관계를 데이터화 하는 데이터 스키마를 특정하여 ‘온톨로지(ontology)’라고 부르기도 한다. 좁은 의미의 ‘온톨로지(ontology)’에 대한 가장 일반적인 정의는 그루버(Gruber, Thomas. 1959~)가 말한 ‘명시적 명세화의 방법에 의한 개념화’(explicit specification of a conceptualization)이다. (Gruber, ‘A Translation Approach to Portable Ontology Specifications’, *Knowledge Systems Laboratory Technical Report KSL 92-71*, Stanford University, 1992) 여기서 ‘개념화’(conceptualization)라는 것은 정보화하고자 하는 대상 세계를 일정한 체계 속에서 파악하는 것, 예를 들면 그 세계에 무엇이 있고, 그것은 어떤 속성을 품고 있으며, 그것들 사이의 관계는 무엇인가 하는 일정한 질문의 틀 속에서 대상 세계를 이해하는 방식이라고 할 수 있다. ‘명세화’(specification)란 대상 세계에 존재하는 개체, 속성, 관계 등을 일목요연한 목록으로 정리하는 것, 그리고 ‘명시적’(explicit)이라는 그 정리된 목록을 사람뿐 아니라 ‘컴퓨터가 읽을 수 있도록’(machine readable) 한다는 것이다.

13) 온톨로지 설계 용어

온톨로지 구성 요소	용도	Web Ontology Language (OWL)
Class, 클래스	공동의 속성을 가진 개체들을 묶는 범주 a group of individuals that belong together because they share some properties.	owl:Class
Individual, 개체	클래스에 속하는 개체 Instances of classes	owl:NamedIndividual
Relation, 관계	(같거나 다른 클래스에 속하는) 개체들 사이의 관계 relationships between pairs of individuals	owl:ObjectProperty
Attribute, 속성	개체가 속성으로 갖는 데이터 값 relationships from individuals to data values	owl:DatatypeProperty
Relation Attribute, 관계 속성	관계 정보에 부수되는 속성 attributes related to relations	N/A in OWL
Domain, 정의역	특정 Object Property 또는 Datatype Property의 주어가 될 수 있는 클래스를 한정 A domain of a property which limits the individuals to which the property can be applied	rdfs:domain
Range, 치역	특정 Object Property의 목적어가 될 수 있는 클래스를 한정 The range of a property limits the individuals that the property may have as its value	rdfs:range

만 발표자는 향후 한국 고고학술정보 온톨로지 개발 연구를 추진할 관계자들의 이해를 돕고자 하는 취지에서, 인문학 분야의 전문사전 데이터화에 적용할 수 있는 온톨로지는 어떠한 것인지 간략하게 보이고자 한다.

1) 데이터의 기능적 모델

4장에서 보인 ‘창녕 교동과 송현동 고분군’ 데이터는 한국학중앙연구원 디지털인문학연구소에서 개발·운용하고 있는 시맨틱 데이터 아카이브-한국문화 엔사이브(Encyves of Korean Culture)의 온톨로지(EKC Data Model)¹⁴⁾에 고고학 유적·유물 및 발굴 조사에 관한 데이터를 담은 것이다. ‘EKC 데이터 모델’은 한국학중앙연구원 디지털인문학연구소에서 실제적인 아카이브 편찬 실무를 통해 만든 연구결과물로서 유사한 성격의 아카이브 편찬에 활용할 수 있도록 공개하고 있는 온톨로지 설계안이다.¹⁵⁾ ‘EKC 데이터 모델’의 표준적인 구성은 다음과 같은 6개의 섹션(Section)으로 이루어진다. 여기서 섹션의 구분이 의미하는 것은 ‘서비스 데이터로서의 기능’과 그 기능을 충실히 하기 위해 특별히 부가된 속성이 있다고 하는 것이다. 이 6개 섹션의 데이터는 아카이브상에서 하나의 복합적인 네트워크로 엮이지만, 그 네트워크상에서 해당 섹션의 성격에 따르는 고유한 서비스 기능을 수행하게 된다.

14) EKC (Encyves of Korean Culture) Data Model: 한국문화 엔사이브 데이터 모델은 한국의 전통문화에 관한 역사적 사실 관계 및 그 사실의 문헌적 근거에 관한 지식을 시맨틱 데이터로 기술하기 위한 온톨로지이다. 한국학중앙연구원 디지털인문학연구소에서 2016년에 처음 제정하고, 매년 확장해 오고 있다. EKC 온톨로지는 아래와 같은 연구개발 사업의 시맨틱 데이터 아카이브 구축에 적용되었다.

- 한국 기록유산의 디지털 스토리텔링 자원 개발(2016~2017, 한국학중앙연구원)
- 문화유산 속의 인물에 관한 시각적 스토리텔링 자원 개발(2017, 한국학중앙연구원)
- 전통문화 융복합 지원을 위한 지능형 검색 플랫폼 구축(2017~2019, 문화체육관광부/㈜PCN)
- 고양 서삼릉 내 묘역 및 태실 역사성 회복 연구 (2019, 문화재청)
- 목포 근대역사문화공간 근대문화자산 아카이브 구축(2020~2021, 목포시청)
- 문화재명칭 영문 표기 용례 / 문화재 안내판 안내문안 데이터 아카이브(2019~2022, 문화재청/한국문화유산활용단체연합회)
- 한양도성 타임머신 빅데이터 아카이브 구축 (2020~2022, 문화재청/㈜PCN),
- 광주 문화예술 인문 스토리 플랫폼 개발 사업 (2002, 광주광역시/대동문화재단)

15) 한국학중앙연구원 디지털인문학연구소, ‘EKC 데이터 모델’, http://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022

데이터 섹션	Class	성격
① 사실과 문맥 데이터 (Fact & Context Data)	Actor, Event, Place, Architecture, Clothing, Food, Object, Work, Record, Concept, Heritage	지식 정보 네트워크
② 학술문헌/웹자원 데이터 (Reference Data)	Bibliography, WebResource, Text	학술정보/관련자원 연계
③ 스토리텔링 데이터 (Storytelling Data)	Storyline, Topic, Episode	스토리텔링
④ 공간 데이터 (Spatial Data)	Space (SpatialPoint, SpatialPath, SpatialArea)	전자지도
⑤ 시간 데이터 (Temporal Data)	Time (TimeSpan, TimeSpanCirca)	전자연표
⑥ 메타버스 어셋 (Metaverse Asset)	3dModel, 3dScene, VR	메타버스 제작/활용

① 사실과 문맥 데이터 (Fact & Context Data)

‘사실과 문맥 데이터’ 섹션은 백과사전적 아카이브에서 지식의 내용을 전달하는 모든 요소들과 그 요소들의 문맥적 상관 관계로 만들어진 네트워크 데이터 전체를 말한다. ‘사실과 문맥 데이터’의 세부 온톨로지는 아카이빙 대상과 범위, 성격, 아카이브의 활용 목적에 따라 정밀도를 달리하게 된다.

② 학술문헌/웹자원 데이터(Reference Data)

‘사실과 문맥 데이터’가 지식의 내용을 정보화 한 것이라면 ‘학술문헌/웹자원 데이터’는 그 지식의 학술적 근거를 밝히는 참고문헌(원천자료, 2차적 연구자료)과 인터넷상에서 접근할 수 있는 ‘참고할만한 자료’들이다. ‘학술문헌/웹자원 데이터’는 아카이브 콘텐츠의 ‘학술적 근거’를 제시하는 기능과 함께 그 ‘학술적 근거’로 알게 된 사실과 문맥이 어떠한 대중적 문화콘텐츠(웹자원)로 만들어져서 이용자들에게 제공되고 있는지를 알게 하는 기능을 담당한다.

③ 스토리텔링 데이터 (Storytelling Data)

스토리텔링 데이터는 아카이브 속의 특정 데이터를 가지고 큐레이션의 줄거리(Storyline)를 만들어서 고객에게 ‘이야기’(Storytelling)를 제공하는 데이터이다.

‘사실과 문맥 데이터’에는 ‘다양한 관점’에서 서로 관계가 있다고 판단된 많은 정보들이 포함된다. 그 가운데 ‘특정한 관점’에서 일련의 데이터를 엮어서 이론이나 이야기를 표현하고자 할 경우, ‘사실과 문맥 데이터’의 특정 노드들을 선택적으로 재이용(reuse)하는 스토리텔링 데이터를 만들 수 있다.

④ 공간 데이터 (Spatial Data)

‘사실과 문맥 데이터’ 속에 있는 모든 공간 정보를 전자지도상에 표시할 수 있도록 하고, 각 지점들의 공간적 연관 관계¹⁶⁾를 탐색할 수 있게 하는 데이터이다. ‘사실과 문맥 데이터’의 ‘Place’ 클래스 속하는 데이터가 특정 장소의 ‘의미’를 알리는 기능이 있는 반면, ‘공간 데이터’ 속의 정보는 그 장소가 지도상의 어느 위치/영역에 존재하는지를 알리는 기능을 한다. 기

16) 두 공간 사이의 포함, 교차, 인접 등의 관계 및 공간적 거리

본적으로 공간 데이터는 GPS (Global Positioning System) 상에서 지리적 위치를 특정할 수 있도록 위도(latitude), 경도(longitude), 고도(altitude) 속성을 부여한 SpatialPoint(지도상의 지점), 지도상의 점들을 선형으로 연장하여 도로, 수로 등을 표시할 수 있게 하는 SpatialPath, 지도상의 점들로 일정한 영역을 표시할 수 있게 하는 SpatialArea 등의 데이터로 구성된다.

⑤ 시간 데이터 (Temporal Data)

‘사실과 문맥 데이터’ 속에 있는 모든 시간 정보를 전자연표에 표시할 수 있도록 하고, 각 시점들의 시간적 연관 관계¹⁷⁾를 탐색할 수 있게 하는 데이터이다. ‘사실과 문맥 데이터’의 ‘Event’ 클래스에 속하는 데이터가 시간적 사건의 ‘의미’를 알리는 기능이 있는 반면, ‘시간 데이터’ 속의 정보는 그 사건을 전자연표의 특정 위치에 표시될 수 있도록 하는 기능을 한다. 시간 데이터는 사건의 시작과 종료 시점을 연, 월, 일 값으로 알리는 TimeSpan 데이터, 모호한 시간대에 대한 정보를 전하는 TimeSpanCirca 등의 데이터로 구성되며, 양력과 음력 사이의 상호참조 정보를 속성으로 포함한다.

⑥ 메타버스 어셋 (Metaverse Assets)

‘메타버스 어셋’ 섹션은 아카이브상의 지식 정보를 메타버스(Metaverse)¹⁸⁾ 환경에서 활용할 수 있도록 하는 3차원 디지털 어셋(3D Digital Assets)¹⁹⁾을 제공하는 것이다. 이것은 다양한 교육적 활용 가치를 지닌 백과사전적 아카이브가 실제로 디지털 교육 환경으로 활용되는 것을 촉진하고, 이용자 참여를 통해 아카이브의 확장이 이루어질 수 있도록 하는 장치이다. 메타버스 데이터는 아카이브 이용자가 다양한 메타버스 플랫폼상에서 관련 주제의 활동공간을 만드는 데 쓰일 수 있는 3D 모델(건물, 물품, 의복, 음식 등의 3차원 형상) 데이터와 3D VR 영상, 그 데이터를 기반으로 제작한, 이용자 참여에 의해 확장 가능한 3D 공간(3D Scene) 데이터로 구성된다.²⁰⁾

2) 클래스 정의

온톨로지(Ontology)에서 ‘클래스’(Class)란 공동의 속성을 가진 개체들을 묶는 범주를 말한다. 다음은 문화재청의 ‘한양도성 타임머신’(2020~2022) 데이터 아카이브 등 ‘EKC 데이터 모

17) 두 시점/시간 사이의 포함, 교차, 시간적 선후 등의 관계 및 시간적 거리(년, 월, 일수)

18) 메타버스(Metaverse): 다수의 사용자가 동시에 체험하며, 교류하고, 발전시켜 갈 수 있는 가상세계. 메타버스가 가상세계에 관한 기존 개념들(VR: 가상현실, AR: 증강현실, MR: 혼합현실 등)과 차별화되는 부분은 그 세계가 ‘공급자’→‘소비자’의 단방향 세계가 아니라 ‘체험하고 교류하고 생산하는 프로슈머(Prosumer: Producer+Consumer)들에 의해 진화하는 쌍방향 세계’ 또는 ‘그러한 세계가 만들어질 수 있는 열린 가능성의 공간(플랫폼)’이라는 점이다.

19) 디지털 어셋(Digital Asset): 이용자가 재사용(reuse)할 수 있는 디지털 저작물. 디지털 데이터를 일방적으로 보여주는 데 그치지 않고, 그것을 재사용할 수 있는 권리를 함께 제공하거나, 그 권리를 얻을 수 있는 방법을 제공하는 경우에 그것을 디지털 어셋이라고 한다.

20) ‘EKC 데이터 모델’에 ‘메타버스 섹션’이 추가된 것은 ‘한양도성 타임머신’ 프로젝트(2020~2022, 문화재청, (주)PCN)를 통해 연간 500건 이상의 3D 모델 데이터가 재이용가능한 디지털 어셋으로 만들어지게 된 데 기인한다. 현재 한양도성 데이터 아카이브와 ‘목포 근대역사문화공간 근대문화자산 아카이브’(2020~2021, 목포시청)에 3D 모델 어셋을 제공하는 메타버스 섹션이 포함되어 있으며, 2022년에 시작한 ‘광주 문화예술 인문 스토리 플랫폼 개발 사업’(2022, 광주광역시/대동문화재단)에서는 시맨틱 데이터 아카이브에서 메타버스 3D 신(Scene)을 제공하는 시스템을 구현하고 있다.

델'을 준용한 온톨로지의 최상위 클래스 구성이다. 유적과 유물, 관련 인물과 기록물 등에 대한 정보가 중심이 된다는 점에서 한국 고고학술정보 아카이브의 데이터 모델 개발에 참고가 될 수 있을 것이다.

section	class	description
① 사실과 문맥 데이터 (Fact & Context Data)	Actor	역사적인 인물 / 집단적 행위 주체로서의 단체 / 운영주체로서의 기관 등
	Event	아카이빙 대상 공간 및 시간 범위 안에서 일어난 사건과 발생 시점이 있는 의례 / 이를 재현하는 행사, 기념제 등
	Place	역사적 배경(사건, 인물)과 관련이 있는 장소. 유물의 소장처
	Architecture	아카이빙 대상 공간 안에 있는 건축물과 조형물
	Clothing	아카이빙 대상 시간 범위의 역사적 인물들이 착용했던 다양한 성격의 복장과 복식 요소들
	Food	아카이빙 대상 의례·행사에서 차려졌던 음식, 식재료와 상차림
	Object	각종 의례/행사의 도구, 그 시대의 문화를 보이는 물품들
	Work	음악, 연희, 무용 등의 공연예술 작품이나 시, 악장가사 등의 문학작품
	Record	문헌, 사진, 도면, 금석문 등 지식 데이터의 원천 자료가 되는 기록물
	Concept	제도, 의례, 풍속 등을 설명하는 데 필요한 용어와 개념
	Heritage	지정문화재 및 등록문화재 등록 정보
② 학술문헌/웹자원 데이터 (Reference Data)	Bibliography	학술적 연구 자료의 목록, 서지사항과 원문/서지사항 URL
	WebResource	월드 와이드 웹상에서 접근할 수 있는 참고 자료
	Text	설명의 증거 기능을 하는 문헌상의 텍스트
③ 스토리텔링 데이터 (Storytelling Data)	Story	'이야기'(Storytelling) 전달을 목적으로 아카이브 속의 노드들 선택하고 설명한 데이터
④ 공간 데이터 (Spatial Data)	Space	GPS (Global Positioning System) 상에서 지리적 위치를 시각화하는 데이터
⑤ 시간 데이터 (Temporal Data)	TimeSpan	Event의 시점/시간 범위를 전자연표상에서 시각화하는 데이터
⑥ 메타버스 어셋 (Metaverse Asset)	3dModel	건물, 물품, 의복, 음식 등의 3차원 형상
	3dScene	3D 모델 데이터를 기반으로 제작한 3D 공간
	VR	실사 영상 기반으로 제작한 3D 공간

3) 데이터 노드(Node)

각각의 클래스에 속하는 개별 데이터를 객체(Object), 개체(Individual) 또는 네트워크의 노드(Node)라고 한다. 이 노드들은 아카이브상에서 그것을 유일하게 식별할 수 있는 고유한 이름(식별자, id)과 네트워크상에서 독자들에게 읽히는 이름(label)을 갖는다. 또 그것이 속하는 클래스 유형에 따라 여러 가지 속성 정보를 갖는다. 아래의 예시는 4장에서 보인 '창녕 교동과 송현동 고분군' 시맨틱 데이터의 몇 가지 노드들이다.

① 사실과 문맥 데이터 노드 예시

id	class	group	label
비화가가야	Actor		비화가가야

송현이	Actor	인물	송현이
순장	Concept	용어	순장
전시-비사벌의 지배자-2014-2015	Event	전시	특별전시: 비사벌의 지배자, 그 기억을 더듬다(2014.10.28.~2015.03.01.)
조사-송현동15-17호분발굴조사-2006-2008	Event	발굴조사	창녕 송현동 고분군 II 15-17호분 발굴조사(2006~2008)
D00-0514-0000	Heritage	국가지정문화재	사적 제514호: 창녕 교동과 송현동 고분군
송현동15호분-순장인골4	Object	출토유물	송현동 제15호분 순장 인골 4(여)
송현이-복원모형	Object	물품	송현이 복원 모형
창녕교동과송현동고분군-송현동1지구15호분	Place	유적지	송현동 1지구 제15호분
창녕박물관	Place	소장기관	창녕박물관
조선총독부박물관-문서-A091-001-041-001	Record	문헌	성 고분 도요지 일람 > 성 고분 도요지 일람 > 창녕 교동 고분군(校洞古墳群)
조선총독부박물관-유리건판-005952	Record	사진	경남 창녕 송현동 고분군(건판005952)

② 학술문헌/웹자원 데이터 노드 예시

id	class	group	label
RISS:A106971037	Bibliography	논문	한영희, 이상수, 「창녕 교동 11호분 출토 유명원두대도」, 『고고학지』 Vol. 2, 한국고고미술연구소, 1990, 85~97쪽.
도록-창녕박물관-2007	Bibliography	도록	도록-창녕박물관 소장유물도록
보고서-국립가야문화재연구소-2012-창녕송현동고분군	Bibliography	보고서	창녕 송현동 고분군 II 15-17호분 발굴조사 보고(2012)
학술대회-국립가야문화재연구소-2021-창녕교동과송현동고분군	Bibliography	학술회의	2021년 가야사 관련 전문가 학술포럼 - 국립가야문화재연구소 가야할 길을 찾다 - 창녕 교동과 송현동 고분군 묘제 특성 검토
youtube-창녕교동과송현동고분군2	WebResource	동영상	국립문화재연구원 ☞ 창녕교동63호분 발굴조사 현장 공개
Nmuseum-relic-224268	WebResource	유물해설	국립중앙박물관 > 소장품 검색 ☞ 금동 말안장 꾸미개
board-창녕교동과송현동고분군_개별10	WebResource	문화유산 안내판	문화유산 안내판 ☞ 창녕 교동과 송현동 고분군 > 송현동_1지구_15호분

③ 스토리텔링 데이터 노드 예시

id	class	group	label
송현이_이야기	Story		송현이 이야기

④ 메타버스 어셋 데이터 노드 예시

id	class	group	label
vr-창녕교동과송현동고분군	VR		VR-창녕 교동과 송현동 고분군

4) 데이터 릴레이션(Relations)

온톨로지(Ontology)에서 ‘릴레이션’(Relations)이란 개별적인 데이터(Individuals) 사이의 관

계를 약속된 어휘로 정의하는 것을 말한다. 시맨틱 데이터 네트워크는 사실상 두 개의 특정 노드를 적정한 관계어로 연결한 데이터가 모여서 이루어지는 것이다. ‘주어’-‘목적어’-‘서술어’의 문장 형식과 유사한 이 릴레이션 데이터의 형식을 RDF(Resource Description Framework)라고 한다.²¹⁾

RDF 구문에 사용하는 관계성 어휘는 클래스(Class) 정의와 마찬가지로 온톨로지 개발자가 정하는 것이다. 어떤 어휘를 어떻게 정하는가에 대해 기술적인 표준안이나 권장안이 존재하는 것은 아니지만, 여러 전문분야의 국제적인 기구나 단체에서 해당 분야 데이터의 교환·공유를 위해 제정한 온톨로지의²²⁾ 어휘들을 참조할 수 있다.

아래의 목록은 4장에서 소개한 ‘창녕 교동과 송현동 고분군’ 시맨틱 데이터의 릴레이션이 어떤 관계어를 사용하였는지를 보이기 위한 예시이다. 여기에서 사용된 관계어들은 EKC 데이터 모델에서 정의된 것들을 사용하였다.²³⁾

① 사실과 문맥 데이터 릴레이션 예시

source	target	relation	domain	range
가야연-송현동1지구15호분-굽다리바리	국립가야문화재연구소	currentLocation	Object	Place
도록-국립김해박물관-2014	전시-비사벌의지배자-2014-2015	documents	Bibliography	Event
가야연맹	금관가야	hasPart	Actor	Actor
가야고분군	고령지산동고분군	hasPart	Place	Place
고령지산동고분군	D00-0079-0000	isDesignatedAs	Place	Heritage
송현동15호분-순장인골4	송현이	isRemainsOf	Object	Actor
송현이	순장	isRelatedTo	Actor	Concept
창녕교동과송현동고분군-송현동1지구15호분	송현동15호분주인	isTombOf	Place	Actor
가톨릭의과대학가톨릭응용해부연구소	연구-창녕송현동15호분순장인물의복원연구-2008-2009	participatesIn	Actor	Event
가야연-송현동1지구15호분-굽다리바리	창녕교동과송현동고분군-송현동1지구15호분	provenance	Object	Place
송현이	창녕교동과송현동고분군-송현동1지구15호분	wasBuriedAt	Actor	Place
송현이	송현이-귀고리	wears	Actor	Object
전시-순장소녀송현이-2011	송현이	curates	Event	Actor
전시-비사벌-2010-2011	창녕교동과송현동고분군	curates	Event	Place
송현이-복원모형	송현이	isRelatedTo	Object	Actor
송현동15호분-순장인골1	조사-송현동15-17호분발굴조사-2006-2008	isRelatedTo	Object	Event

21) RDF는 W3C(World Wide Web Consortium)가 제안하는, 웹상의 데이터 교환을 위한 표준 모델이다. RDF 구문의 형식은 웹의 하이퍼링크 구조를 확장한 것으로, 연결의 출발점과 도착점뿐 아니라 그것들 사이의 관계도 URI로 명명한다. (Resource Description Framework (RDF), <https://www.w3.org/RDF/>)

22) Dublin Core (DC) ontology (<http://dublincore.org/> 문헌정보), Friend Of A Friend (FOAF) ontology (<http://www.foaf-project.org/> 인적/사회적관계), Europeana Data Model (<https://pro.europeana.eu/share-your-data/metadata> 문화유산 메타데이터) 등

23) EKC 데이터 모델의 관계어 전체 리스트: http://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022:Relation

source	target	relation	domain	range
가야고분군	가야연맹	isRelatedTo	Place	Actor
송현이-복원모형	전시-순장소녀송현이-2011	hasExhibition	Object	Event

② 학술문헌/웹자원 데이터 릴레이션 예시

source	target	relation	domain	range
가야연-송현동1지구15호분-굽다리바리	도록-국립김해박물관-2014	isMentionedIn	Object	Bibliography
board-창녕교동과송현동고분군_개별8	보고서-국립가야문화재연구소-2011-창녕송현동고분군-1	isReferencedFrom	WebResource	Bibliography
창녕교동과송현동고분군-송현동1지구15호분	보고서-국립가야문화재연구소-2009-창녕송현동고분군	isShownAt	Place	Bibliography
송현이_이야기	보고서-국립가야문화재연구소-2009-창녕송현동고분군	isReferencedFrom	Story	Bibliography
순장견	cha-story-155703119	isShownAt	Actor	WebResource
국박-교동1군11호분-은제관 꾸미개	Nmuseum-relic-212818	isShownAt	Object	WebResource
창녕교동과송현동고분군-교동1군12호분	board-창녕교동과송현동고분군_개별2	isShownAt	Place	WebResource
송현이_이야기	cha-story-1333805140000	isShownBy	Story	WebResource

③ 스토리텔링 데이터 릴레이션 예시

source	target	relation	domain	range
송현이_이야기	송현이	hasContextualElement	Story	Actor
송현이_이야기	창녕교동과송현동고분군-송현동1지구15호분	hasContextualElement	Story	Place
송현이_이야기	송현동15호분-순장인골4	hasContextualElement	Story	Object
송현이_이야기	송현이-귀고리	hasContextualElement	Story	Object
송현이_이야기	송현이-복원모형	hasContextualElement	Story	Object
송현이_이야기	전시-순장소녀송현이-2011	hasContextualElement	Story	Event
송현이_이야기	보고서-국립가야문화재연구소-2009-창녕송현동고분군	hasContextualElement	Story	Bibliography
송현이_이야기	youtube-창녕교동과송현동고분군3	hasContextualElement	Story	WebResource
송현이_이야기	cha-story-1333805140000	hasContextualElement	Story	WebResource

데이터 아카이브를 편찬할 때 사용할 관계어를 미리 약속된 어휘로 정의하는 이유는 데이터 사이의 관계가 편찬자에 따라 다르게 정의될 경우, 신뢰할 수 있는 네트워크 데이터의 구현이 어렵기 때문이다. 온톨로지의 관계어는 인간의 언어에서 가져오는 경우가 대부분이지만, 자연어에서처럼 다의적이거나 모호하게 사용되지 않도록 명확하게 한정된 의미로만 쓰이게 해야 한다. 처음부터 많은 수의 어휘를 정의하기보다는 아카이브 대상 정보의 성격을 살펴서 명확하게 데이터화 할 수 있는 기본적인 관계들을 먼저 정의하고, 데이터를 확장할 때 새로운 관계어들을 추가하는 방법이 권장된다.²⁴⁾

24) 한국학중앙연구원 디지털인문학연구소의 온톨로지 운용 및 확장 프로세스

① 연구팀은 새 프로젝트의 기초 연구자료와 스토리 주제가 선정된 시점에 그 프로젝트의 데이터 큐레

6. 고고학술정보 시맨틱 데이터 아카이브의 활용

한국 고고학술정보를 ‘디지털 전환’의 이름으로 새롭게 변신시키는 노력을 기울이고자 한다면, 데이터 편찬과 시스템 개발의 과정에 착수하기 전에 먼저 짚어 보아야 할 문제가 있다. “그 새로운 시스템의 고객은 누구일까?”하는 문제이다.

데이터를 시각적으로 보여줄 수 있다는 점에 주목하는 이들은 이것이 ‘대중’들의 접근을 용이하게 할 것으로 기대할 수 있다. 그러나 아카이브 속의 데이터가 큰 규모로 축적되면, 그래프의 복잡도는 폭발적으로 증가하게 되고, 사람의 눈으로 그 관계성을 추적하는 일이 어려워지게 된다.²⁵⁾ 그러면 시맨틱 아카이브의 고객은 해당 분야의 교수나 연구원 같은 전문가 그룹에 한정되는 것일까? 세분화된 전공영역에서 깊이 파고드는 연구를 하는 이들은 자신의 주의력이 다방면으로 확산되는 것을 원하지 않을 수도 있다.

저작물의 고객을 ‘대중’과 ‘전문가’로 나누는 종래의 이분법적 사고로는 시맨틱 데이터 아카이브의 고객을 이야기하기 힘들다. 오히려, 지식의 소비자가 동시에 새로운 지식의 생산자 되는 디지털 세계 안에는 ‘전문가’의 학술과 ‘대중’의 즐길거리 사이에서 중재적 역할을 하는 매개적 커뮤니케이터들이 다양한 형태로 존재한다는 점에 주목할 필요가 있다. 유튜버, 블로거, 위키백과의 기고자들, 디지털 콘텐츠 제작자들..... 시맨틱 데이터 아카이브를 가장 먼저 반기고 이용할 사람들은 디지털 지식 콘텐츠 시장의 프로슈머들일 것이다.

고등연구기관의 학술활동이 전공이라는 이름으로 극단적으로 세분화되어 있기 때문에 전문 연구자들의 저작물만을 가지고 분야를 넘나드는 횡단적 지식을 탐구하는 것은 쉽지 않다. 만일 그 정도 수준의 심도 있는 지식 데이터가 전공의 벽을 넘어서 꼬리를 물고 이어져서 그 지식 세계의 다채로운 면모에 접근할 수 있는 길을 열어 준다면, 대중이 관심을 갖는 2차적 콘텐츠를 만드는 사람들에 큰 도움이 될 것을 기대할 수 있다. 3D 가상현실을 무대로 하는 콘텐츠 유통 플랫폼-메타버스의 세계에서는 오락물로서의 게임 콘텐츠뿐 아니라, 다양한 수준의 교양, 교육 콘텐츠가 개발·운용될 것이다. 시맨틱 데이터 아카이브는 그러한 오락물과 교양물, 교육 프로그램의 개발이 필요로 하는 스토리라인과 에피소드의 디테일을 제공할 수 있다.

여기서 한 걸음 더 나아가 생각할 것은, 지금은 위에서 언급한 특징인들이 프로슈머의 역할을 수행하지만 디지털 환경에서 태어난 미래 세대들은 누구나 디지털 프로슈머로 자라날 것이

이전에 적용할 온톨로지 초안을 제정. (* 기존의 EKC 데이터 모델을 기반으로 하되, 당해 연도의 기초 연구자료 처리에 필요한 신규 어휘를 추가)

- ② 이 온톨로지에 입각하여 데이터 큐레이션을 수행하면서, 온톨로지 어휘의 사용 현황을 모니터링.
 - ③ 온톨로지 어휘의 관리를 담당하는 태스크 포스를 구성하여, 새로운 어휘 제정의 요구가 있을 때 이를 판단, 제정, 공시하는 일을 수행.
- 25) 궁극적으로 네트워크 데이터에서 유용한 정보를 찾는 것은 사람의 눈으로 할 일이 아니라 소프트웨어의 도움을 받아서 할 일이다. 네트워크 데이터베이스 상에서 SPARQL (Simple Protocol and RDF (Resource Description Framework) Query Language) 검색 명령을 이용하면, 아카이브 속에 담긴 데이터들의 복잡한 연관관계를 추적해서 의미 있는 결과를 얻을 수 있다. 예를 들어 ‘송현동 고분에서 발굴된 유물 중 마구류에 속하는 것은 어느 박물관에 있고, 그 유물에 대해 연구한 논문의 저자는 누구이며, 또 그들이 쓴 다른 고고학 논문은 무엇인가?’라는 질문의 답을 한 번의 SPARQL 명령으로 얻는 것이 가능하다. 이 일은 누가할까? 정보기술자들도 잘 모르는 SPARQL 문법이 1차적인 장애이지만 그 문제는 조만간 인간의 자연어를 SPARQL 명령어로 전환하는 인공지능적 소프트웨어가 해결해 줄 것이다. 하지만 그렇게 된다고 하더라도 아카이브 속의 정보에 대해 어느 정도 기초적인 지식이 있고, 무엇을 찾겠다는 분명한 개념이 있지 않으면 올바른 질문을 던지는 것 자체가 불가능하다.

라는 점이다. 이들 디지털 원어민들은 복잡하게 얽힌 디지털 클라우드의 세계에서 자신의 호기심이 이끄는대로 다양한 지식의 연결고리들을 추적하고, 거기서 얻은 재료들을 가지고 자신의 이야기를 담은 또 다른 디지털 콘텐츠로 만들어낼 수 있는 능력자들이다. 그런데, 디지털 플랫폼상에서 자유롭게 활동할 능력이 있다고 해서 그곳에 담긴 콘텐츠의 내용을 올바르게 이해하고 활용할 능력까지 저절로 얻게 되는 것은 아닐 것이다. 하지만 역사 교과서는 들여다보려 하지 않은 학생들도 여행지에서 본 고대유적이 3차원 가상현실 게임의 무대가 될 수 있음을 알고, 또 그것을 가능하게 하는 다양한 재료가 어디에 있는지를 알게 되면, 그 지식정보 데이터와 어셋들을 찾아서 자신의 콘텐츠로 엮어내는 일에 열정을 다할 수 있다. 『한국고고학사전』 같은 내용의 시맨틱 데이터 아카이브는 디지털 원어민들이 인문학을 공부하고 인문교양인으로 성장할 수 있게 하는 교육환경으로 활용되어야 한다.

그리고, 우리가 궁극적으로 기대하는 시맨틱 데이터 아카이브의 타겟 오디언스는 인공지능이다. 시맨틱 데이터 아카이브를 구성하는 RDF(Resource Data Framework) 형식의 데이터는 인간의 언어 속에 모호하게 담겨있던 인문학 지식을 컴퓨터가 인식할 수 있는 명시적인 데이터로 변환한 것이다. 이것은 딥러닝 프로세스의 1차적인 학습 데이터로 사용될 수 있으며, 인공지능 스스로 새로운 지식을 찾아서 배울 수 있게 만드는 촉진제 역할을 할 수 있다. 그렇게 되면 자연어 문장의 질의어를 SPARQL 구문으로 바꾸는 일만 기계적으로 할 수 있을 뿐 아니라, SPARQL 질의로는 얻을 수 없는 질문에 대해서도 정답을 찾는 데 도움이 될 여러 가지 후보 답안을 인공지능적 추론으로 얻을 가능성도 열리게 될 것이다. 오해하지 말아야 할 점은 믿을 수 있는 기초 데이터가 제공되지 않았는데 컴퓨터가 스스로 인간을 가르칠 무엇을 찾아낼 일은 결코 없다는 것이다. 인공지능 기술이 남용되는 곳에서는 모호한 데이터로부터 도출된 잘못된 판단을 마치 새로운 발견인 것처럼 오해하거나 과장하는 일도 발생한다. 이러한 오류를 통제하는 역할도 시맨틱 데이터 아카이브 속의 명시적인 데이터에 기대할 수 있는 일이다.

【참고문헌】

§ 문헌자료(학술논문, 단행본, 조사보고서, 박물관간행물)

- 강재광, 「한국민족문화대백과사전 개정증보사업(2007-2017)의 성과와 향후 발전 방향」, 『한국사전학』 34, 2019, 7-37쪽
- 국립가야문화재연구소, 『2021년 가야사 관련 전문가 학술포럼 - 국립가야문화재연구소 가야 할 길을 찾다 - 창녕 교동과 송현동 고분군 묘제 특성 검토』, 2021.
- 국립가야문화재연구소, 『가야이야기』, 2015.
- 국립가야문화재연구소, 『학술연구총서 제67집 - 창녕 교동 88호분 발굴조사보고서』, 2015.
- 국립가야문화재연구소, 『학술연구총서 제81집 - 창녕 교동 39호분 및 주변 고분 I - 52호분 및 주변 유적』, 2021.
- 국립가야문화재연구소, 『학술조사보고 제50집 - 창녕 송현동 고분군 I 6·7호분 발굴조사 보고 - 도판』, 2011.
- 국립가야문화재연구소, 『학술조사보고 제50집 - 창녕 송현동 고분군 I 6·7호분 발굴조사 보고 - 본문, 도면』, 2011.
- 국립가야문화재연구소, 『학술조사보고 제50집 - 창녕 송현동 고분군 I 6·7호분 발굴조사 보고 - 분석결과』, 2011.
- 국립가야문화재연구소, 『학술조사보고 제51집 - 창녕 송현동 고분군 II 15-17호분 발굴조사 보고』, 2012.
- 국립가야문화재연구소, 『학술조사보고 제57집 - 창녕 교동 고분군 주차장 조성부지 내 유적 발굴조사보고』, 2013.
- 국립가야문화재연구소, 『학술총서 제43집 - 1500해앞 16살 여성의 삶과 죽음 - 창녕 송현동 15호분 순장인골의 복원연구』, 2009.
- 국립김해박물관, 창녕군, 우리문화재연구원, 『비사벌의 지배자, 그 기억을 더듬다』, 2014.
- 국립창원문화재연구소, 『학술조사보고 제33집 - 창녕 송현동 고분군 6·7호분 발굴조사 개보』, 2006.
- 김현·임영상·김바로, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016.
- 동아대학교 박물관, 『고적조사보고 제21책 - 창녕 교동 고분군 발굴조사 보고서』, 1992.
- 문화재청, 국립중앙박물관, 『국립중앙박물관 문화재청 공동기획 특별전 - 발굴에서 전시까지』, 2007.
- 문화재청, 「순장소녀, 송현이 비사벌을 말하다 - 창녕 교동과 송현동 고분군」, 『문화유산 활용을 위한 이야기 요소 발굴 연구 - 경남·부산·울산 문화유산 이야기 여행』, 2012, 215~218쪽.
- 이주현, 「고고자료로 본 비사벌의 대외교류」, 『영남고고학』 no. 62, 영남고고학회, 2012, 83~104쪽.
- 창녕군청, 『창녕박물관 소장유물도록』, 2007.
- 한국학중앙연구원, 「한국민족문화대백과사전 개정증보 결과 보고서」, 2010, 2013, 2017
- 한영희, 이상수, 「창녕 교동 11호분 출토 유명원두대도」, 『고고학지』 Vol. 2, 한국고고미술연구소, 1990, 85~97쪽.
- 홍승우, 「창녕 교동 11호분 출토 명문대도 재검토」, 『한국고대사연구』 no.101, 한국고대사학회, 2021, 113~142쪽.

§ 웹사이트

- 국립가야문화재연구소, '교동과 송현동 고분군'에서 순장견 3구 확인', 문화재청 보도자료, 2021.11.30. https://www.cha.go.kr/newsBbz/selectNewsBbzView.do?newsItemId=155703119§ionId=b_sec_1&mn=NS_01_02
- 국립문화재연구원 문화유산 연구지식포털 > 한국고고학 사전(2001). https://portal.nrich.go.kr/kor/archeologyUsrList.do?menuIdx=792&type_gbn=3000
- 국립문화재연구원 문화유산 연구지식포털 > 한국고고학 전문사전(고분편). https://portal.nrich.go.kr/kor/archeologyUsrList.do?menuIdx=797&type_gbn=565
- 국립문화재연구원 문화유산 연구지식포털 > 한국고고학 전문사전(고분유물편). https://portal.nrich.go.kr/kor/archeologyUsrList.do?menuIdx=798&type_gbn=2985
- 국립문화재연구원 문화유산 연구지식포털 > 문화유적총람. <https://portal.nrich.go.kr/kor/overviewUsrList.do?menuIdx=572>
- 국립문화재연구원 문화유산 연구지식포털 > 영상자료관. <https://portal.nrich.go.kr/kor/multistageUsrList.do?menuIdx=1047&total=Y>
- 국립문화재연구원 문화유산 360 VR > 창녕 교동과 송현동 고분군. <https://vrnrich.kr/view-project/62>
- 국립문화재연구원 문화유산 360 VR > 창녕 교동과 송현동 고분군 정비사업부지 내 문화재 발굴조사. <https://vrnrich.kr/view-project/61>
- 국립중앙박물관 소장 조선총독부박물관 문서. <https://www.museum.go.kr/modern-history/main.do>
- 국립중앙박물관 소장 조선총독부박물관 유리건판. <https://www.museum.go.kr/dryplate/main.do>
- 김현, '데이터 기반 인문지식 백과사전: 기존의 백과사전과 무엇이 다른가?', http://dh.aks.ac.kr/Encyves/wiki/index.php/데이터_기반_인문지식_백과사전
- 한국학중앙연구원 디지털인문학연구소, '문화재 안내판 안내문안 데이터 아카이브(Cultural Heritage Wiki)'. http://dh.aks.ac.kr/~heritage/wiki/index.php/Cultural_Heritage_Wiki
- 한국학중앙연구원 디지털인문학연구소, 'EKC 데이터 모델', http://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022
- YouTube, 가야고분군, "창녕 교동과 송현동고분군 애니메이션", 2021.09.30. https://www.youtube.com/watch?v=9SpJVx_RxF8
- YouTube, 국립문화재연구원 NRICH, "창녕교동63호분 발굴조사 현장 공개", 2020.11.05. <https://www.youtube.com/watch?v=D6cuTORvWGc>
- YouTube, KBS 역사저널 그날, "KBS 역사추적 - 17세 가야 소녀는 왜 순장당했나 / KBS 2009.04.27. 방송", 2021.02.19. https://www.youtube.com/watch?v=_s-M46KWpe0