

# 고고학술정보 데이터 아카이빙 전략과 구현 방안

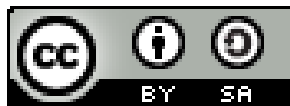
한국고고학사전의 디지털 전환을 위한  
시맨틱 데이터 아카이브 제안

김현, 강혜원

한국학중앙연구원 한국학대학원 인문정보학교실

한국학중앙연구원 디지털 인문학 연구소

[hyeon@aks.ac.kr](mailto:hyeon@aks.ac.kr)



이 저작물(PPT)의 인용 표시 방법:

김현, 강혜원, "고고학술정보 데이터 아카이빙 전략과 구현 방안",  
「고고학술정보 디지털 대전환」 국제학술심포지움 발표자료,  
2022. 9. 28. 국립문화재연구원

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

# '디지털 전환'이란?

❖ 전산화(Digitization) → 디지털화/정보화(Digitalization) → 디지털 전환(Digital Transformation)

- 전산화(Digitization): 객체의 물리적인 형상이나 속성이 디지털 신호로 표현될 수 있도록 하는 것. 🖱 백과사전 텍스트 전산 입력 / 아카이브 물품의 디지털 사본 제작
- 디지털화(Digitalization): 전산화의 결과를 활용하여 해 오던 일의 프로세스를 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

## 한국민족문화대백과사전의 전산화

### ❖ 한국민족문화대백과사전

- 1980년에 편찬사업을 시작하여 1991년에 초판 27권(본책 25권, 부록 2권)을 책자로 간행

### ❖ 한국민족문화대백과사전 CD-ROM

- 2001년에 오프라인 디지털 저작물(『디지털 한국민족문화대백과사전 EncyKorea』: CD-ROM 5장, DVD 1장)의 형태로 1차 개정판을 간행.
- 이를 위해 종이매체에 실렸던 『한국민족문화대백과사전』의 내용을 디지털 매체에 옮겨 담은 과정은 '전산화(Digitization)' 업무 수행

# 한국민족문화대백과사전의 전산화/디지털화

## ❖ 한국민족문화대백과사전 개정증보(2007~2017) 사업을 통해 수행한 디지털화

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 ( <a href="http://encykorea.aks.ac.kr">http://encykorea.aks.ac.kr</a> )
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

# 데이터 기반 지식백과사전

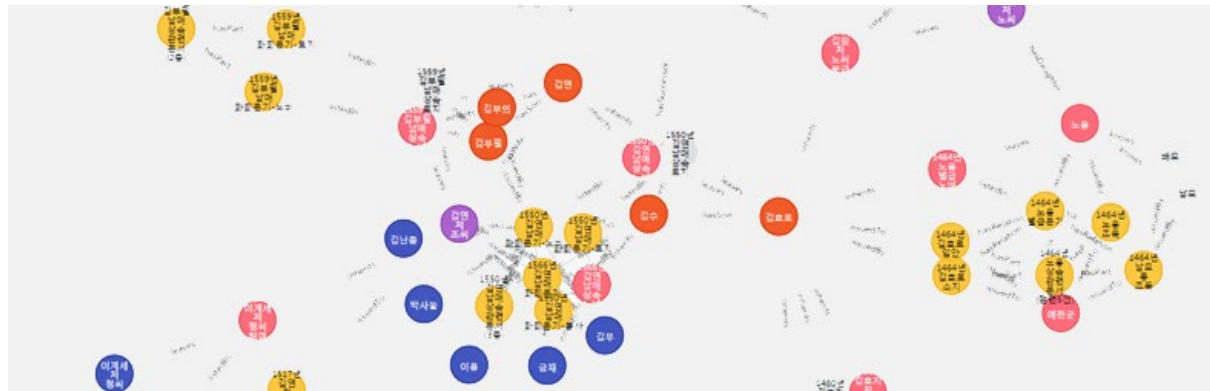
## ❖ 데이터 기반 지식백과사전의 조건

①	데이터 네트워크 형태의 백과사전	분과학문적 전문지식의 내용들이 그것이 다룬 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

## 네트워크 형태의 백과사전

### ❖ 네트워크 형태의 백과사전

- 다양한 주제의 정보가 독립적인 '항목'으로 나열되어 있는 것이 종래의 백과사전이었다고 한다면, 데이터 기반 백과사전은 항목과 항목, 또는 개별 항목 속의 지식 요소들이 서로 어떠한 의미의 연관이 있는지 그 '관계성'을 보이는 데 주력.
- 데이터 기반 백과사전은 '가나다' 순으로 정리된 목록 형태가 아니라, 방대한 규모의 데이터 네트워크의 형태로 편찬.
- 그 네트워크의 노드 하나 하나가 백과사전의 항목이고, 노드와 노드 사이를 연결하는 연결선은 항목과 항목 사이의 다양한 연관관계를 대변.



## ‘지식’과 ‘자료’의 융합

### ❖ ‘지식’과 ‘자료’를 넘나드는 융합 콘텐츠

- 디지털 인문학의 주제 가운데 하나는 대학의 강의실에서 다루는 이론적 지식과 아카이브에 소장된 그 지식의 증거 자료 사이의 통섭.
- 아날로그 세계에서는 ‘연구실’과 ‘아카이브’가 서로 분리된 영역에 속했지만, 한 번의 클릭으로 그 두 세계를 넘나들 수 있는 디지털 환경에서는 ‘지식’과 ‘자료’가 훨씬 가깝게 묶일 수 있다.
- 데이터 기반 백과사전은 서로 관련한 ‘지식’과 ‘자료’를 함께 탐구할 수 있는 길을 제공.





## 감성적 체험 공간

### ❖ 가상현실에서 지식의 현장을 느끼는 감성적 체험 공간

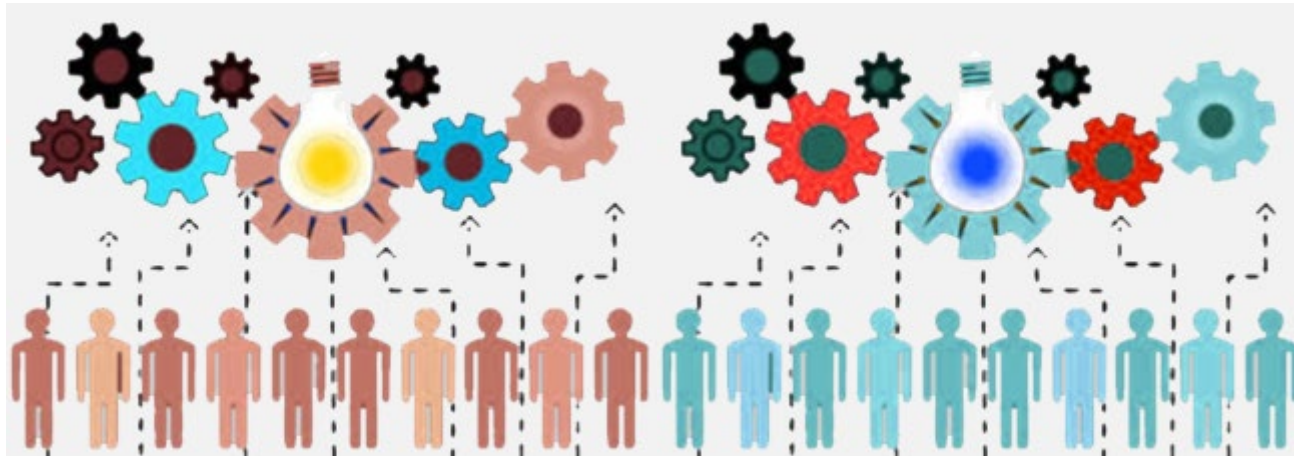
- 새로운 백과사전의 데이터는 학술적 이론과 아카이브의 유물, 현장의 활동을 하나의 가상 공간에서 융합하고 감성적으로 체험하게 할 수 있게 하는 하이퍼미디어 콘텐츠.
- 역사와 문화의 자취를 담은 지리적 공간을 3차원 가상현실의 속에서 체험하면서, 그 구석구석에서 호기심에 답하는 백과사전적 지식을 열어 볼 수 있는 유저 인터페이스를 제공.
- 백과사전 데이터 아카이브 속의 3차원 가상현실 어셋을 기반으로 교육/문화 활동 **메타버스(Metaverse)**의 구현 → 이용자들이 체험하고 학습하며 자신의 콘텐츠를 만들어 가는 참여와 소통의 공간



## 개방적 데이터

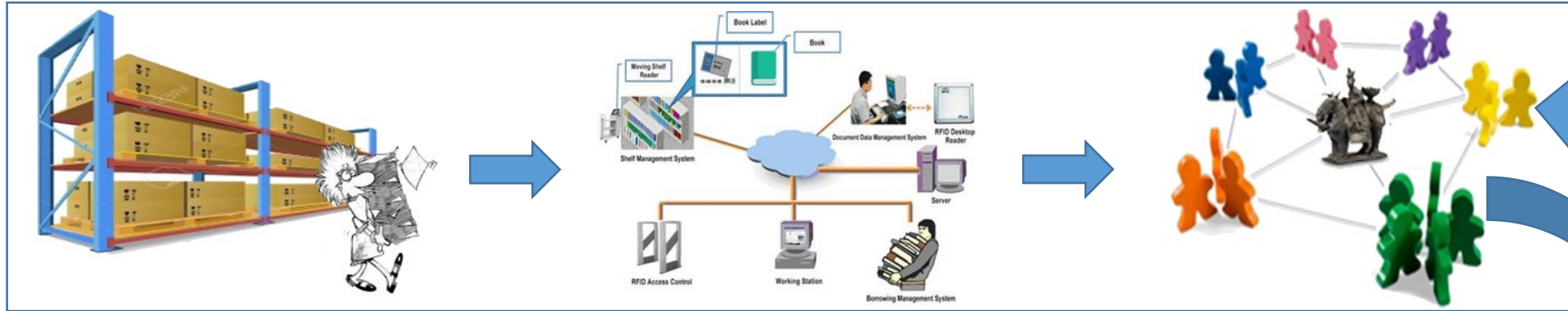
### ❖ 집단 지성의 기여로 확장되는 개방적 데이터

- 데이터 기반 백과사전의 지식 관계 네트워크는 학자들이 생산한 전문 지식뿐 아니라, 다양한 층위의 '유용한 지식'에 접근할 수 있는 통로를 제공.
- 지식의 이용자가 자신의 학습과 경험 속에서 새로운 지식의 연결고리를 발견하고, 자기 관점의 디지털 스토리텔링으로 지식 관계 네트워크를 확장.
- 이 방법으로 사회과학이나 산업 분야의 빅데이터에 대한 연결고리가 만들어질 수 있으며, 초고도 연결 ([Hyperconnectivity](#)) 사회의 빅데이터 속에서 고전 인문지식이 유효한 역할을 수행.



## 디지털 전환, 아카이브와 백과사전

❖ 아카이브의 디지털 전환: “아카이브의 정보화”에서 “디지털 아카이브”로



❖ 백과사전의 디지털 전환: “백과사전의 디지털 서비스”에서 “디지털 백과사전”으로



백과사전적아카이브

## 데이터 기반 지식백과사전: 백과사전적 아카이브

### ❖ 백과사전적 아카이브란?

- 백과사적 아카이브(Encyclopedic Archives), 약칭 엔사이브(Encyves)
- 아카이브 컬렉션의 디지털 사본과 그 유물들의 사회적·문화적 문맥을 알려 주는 백과사전적 지식 정보를 긴밀하게 연계시키는 융합적 시스템
- (인문학 교육·연구 활동으로서의) 인문지식 디지털 큐레이션의 산출물이자, 새로운 교육과 연구의 도구로 쓰일 수 있는 지식 공유 환경

## 시맨틱 데이터 아카이브

### ❖ ‘시맨틱(Semantic, 의미 기반)’

개개의 데이터가 서로 내용과 의미 면에서 어떠한 관계에 있는지를 알게 한다는 뜻.

### ❖ 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.

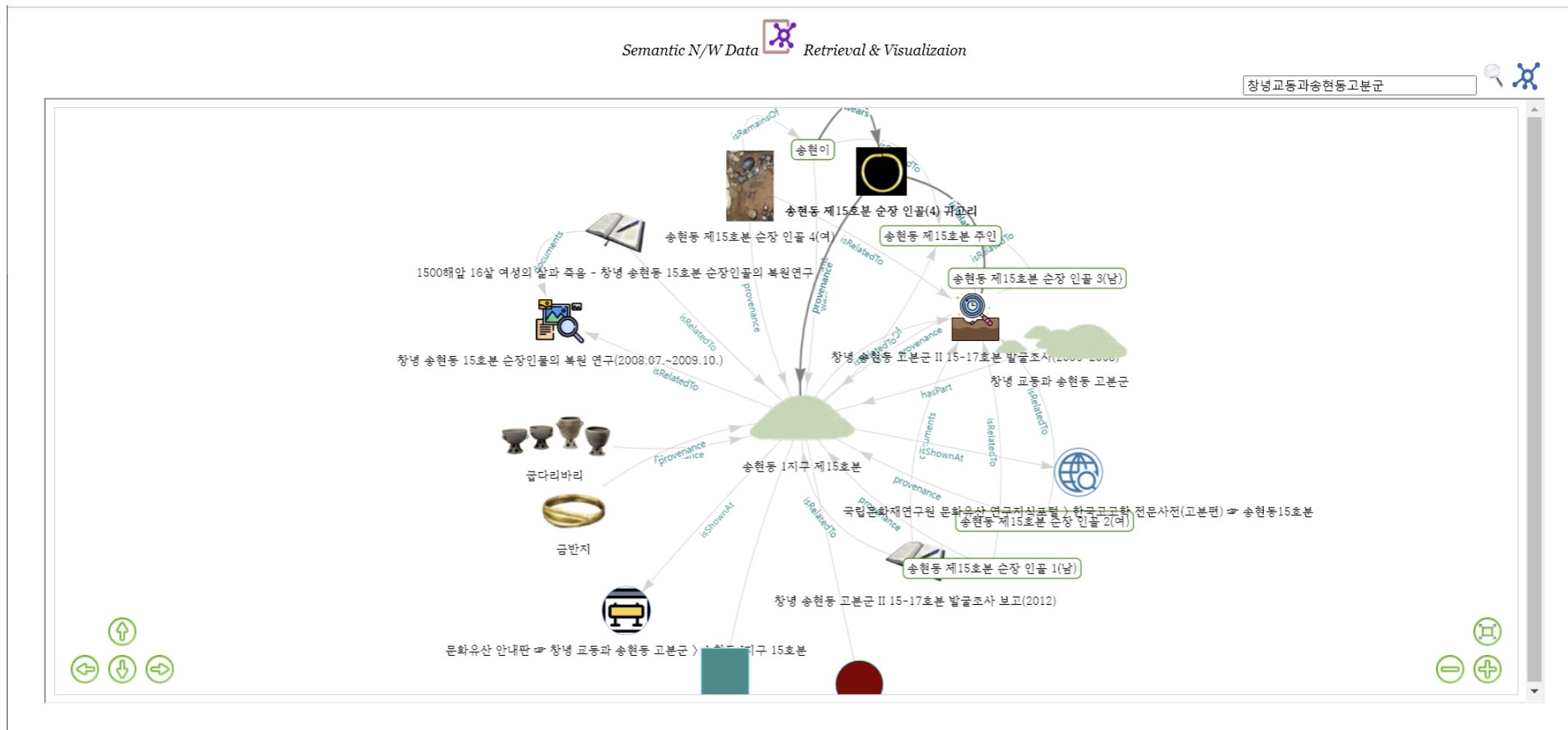
# 『한국고고학사전』 시맨틱 데이터 아카이브 예시

## ❖ 【예시 1】 창녕 교동과 송현동 고분군



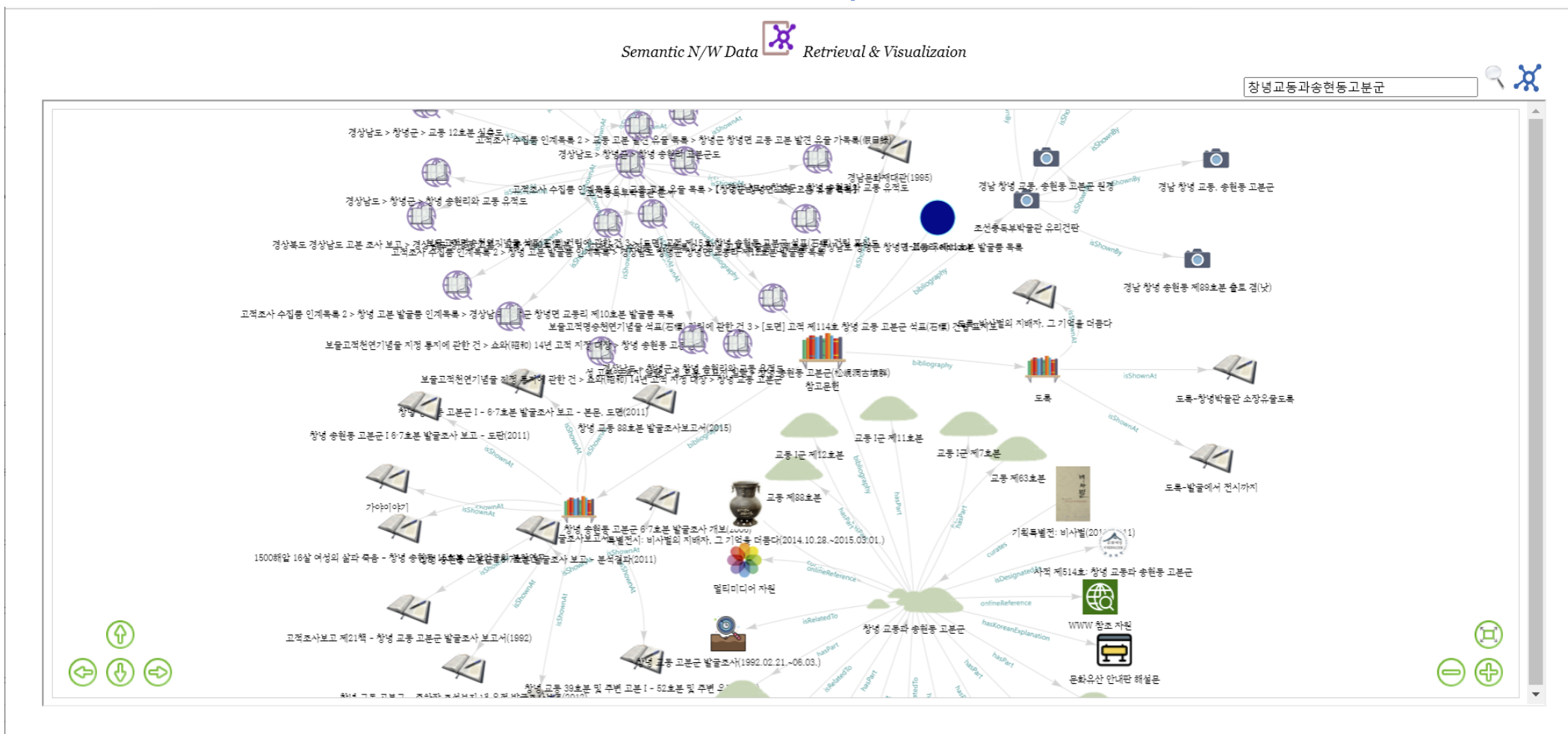
## 『한국고고학사전』 시맨틱 데이터 아카이브 예시

## ❖ 【예시 2】 송현동 제15호분





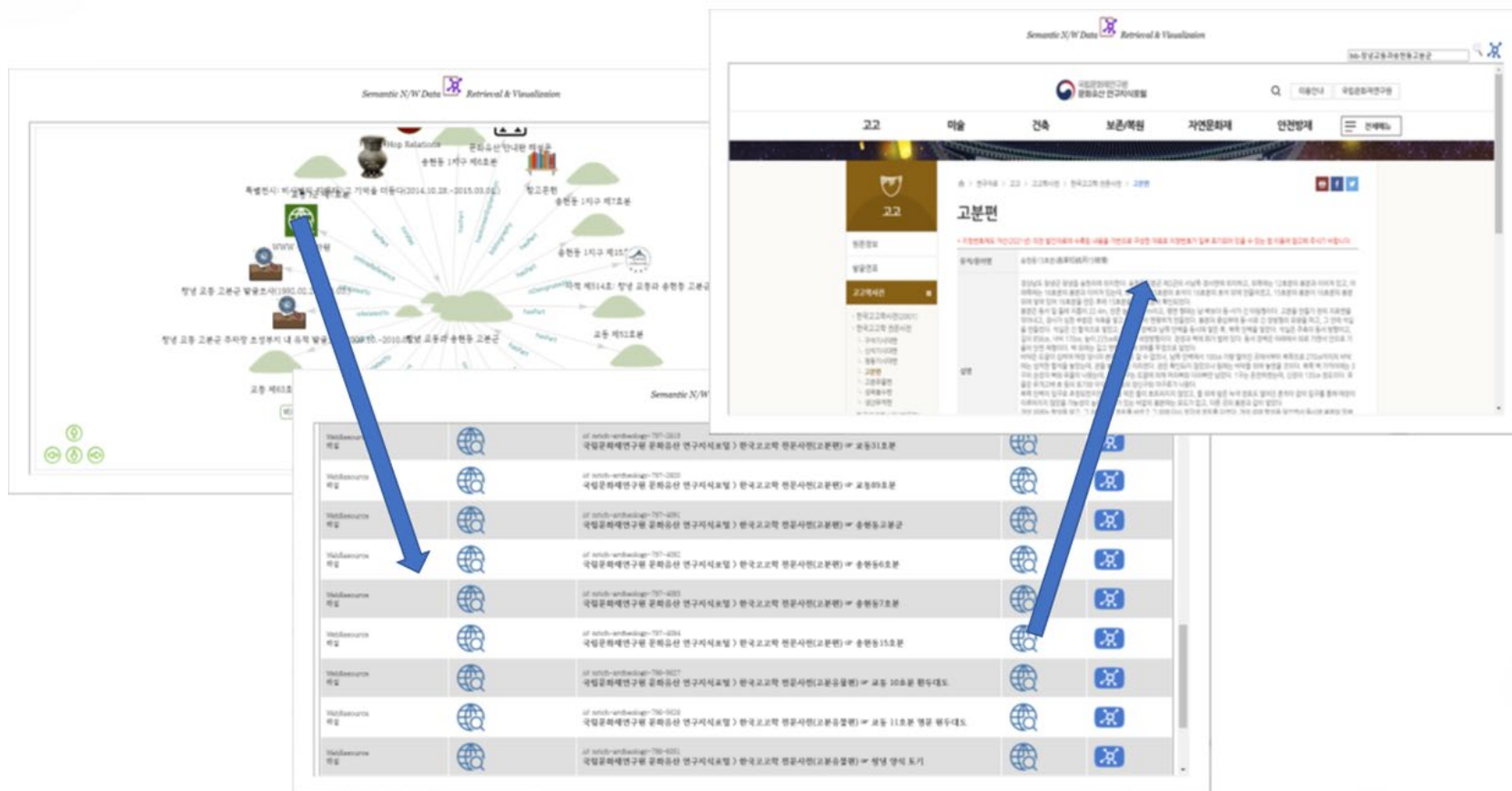
❖ 【예시 3】 창녕 교동과 송현동 고분군 참고문헌(발굴보고서, 기록사진, 도록, 고고학사전 등)





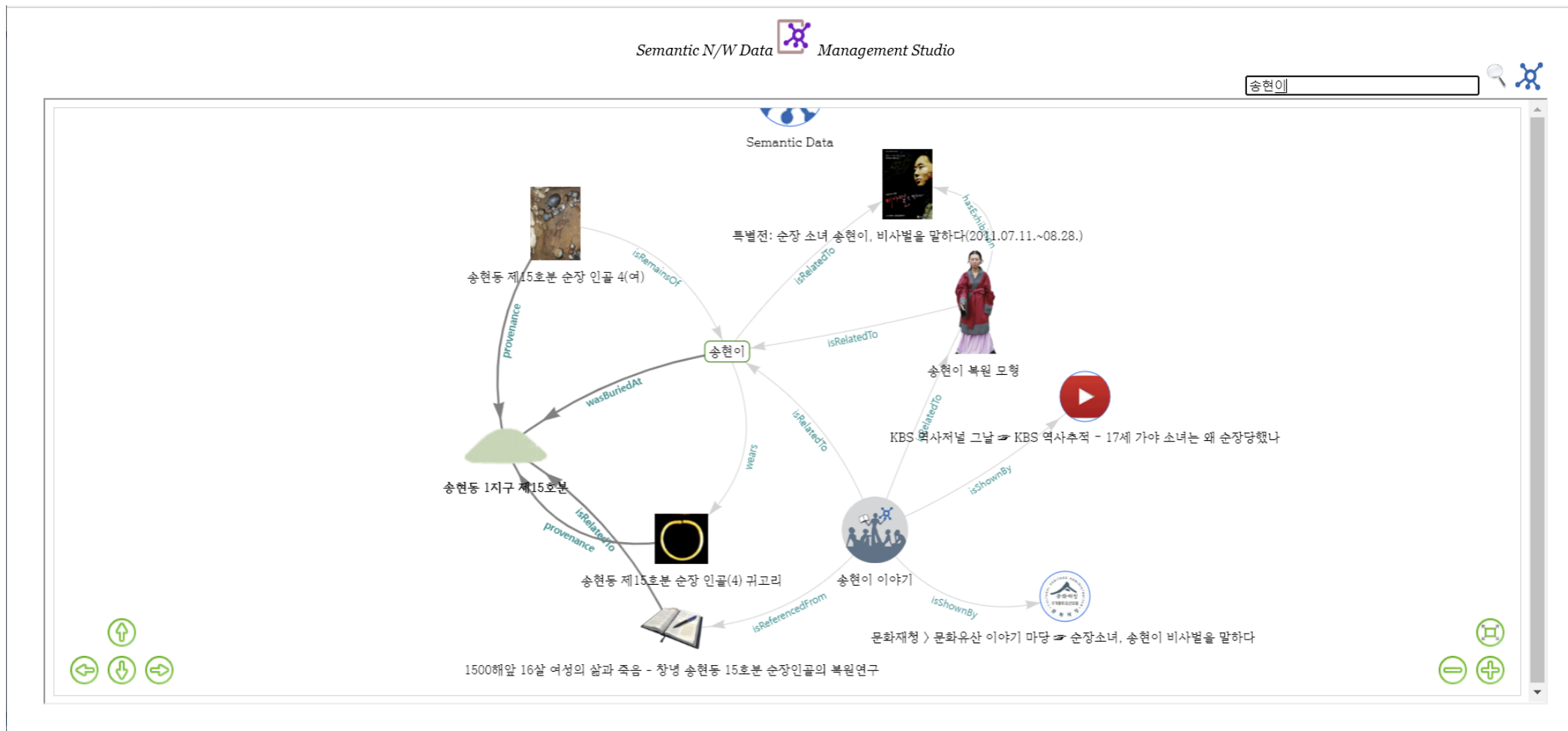
## 『한국고고학사전』 시맨틱 데이터 아카이브 예시

## ❖ 【예시 4】 현행 한국고고학전문사전 데이터 연계



## 『한국고고학사전』 시맨틱 데이터 아카이브 예시

## ❖ 【예시 5】 스토리텔링 데이터: 송현이 이야기



## 온톨로지(Ontology)

### ❖ 온톨로지(Ontology)

- 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계. 넓은 의미에서는 모든 정보화의 틀이 다 '온톨로지'일 수 있겠지만, 정보 요소 사이의 관계를 데이터화 하는 데이터 스키마를 특정하여 '온톨로지(ontology)'라고 부르기도 한다.
- 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.
- 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.

### ❖ 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 Data Model

## ❖ 데이터 섹션 (데이터의 기능적 역할에 의한 구분)

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

# EKC Data Model

## ❖ 클래스(Class) 정의(1)

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 복장과 복식
	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

# EKC Data Model

## ❖ 클래스(Class) 정의(2)

section	class	description
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

# EKC Data Model

## ❖ 데이터 노드(Node) 정의 예시

id	class	group	label
비화가야	Actor		비화가야
송현이	Actor	인물	송현이
순장	Concept	용어	순장
전시-비사벌의지배자-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)

# EKC Data Model

## ❖ 관계성(Relation) 정의 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



## 온톨로지 운용 및 확장 프로세스

### ❖ 온톨로지 어휘 개발

- 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 때문.
- 처음부터 많은 수의 어휘를 정의하기보다는 아카이브 대상 정보의 성격을 살펴서 명확하게 데이터화할 수 있는 기본적인 관계들을 먼저 정의하고, 데이터를 확장할 때 새로운 관계어들을 추가하는 방법을 권장.

### ❖ 한국학중앙연구원 디지털인문학연구소의 온톨로지 운용 및 확장 프로세스

- ① 연구팀은 새 프로젝트의 기초 연구자료와 스토리 주제가 선정된 시점에 그 프로젝트의 데이터 큐레이션에 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.

## 디지털 전환, '유물'과 '지식'의 새로운 관계

### ❖ 아카이브의 세계 vs 백과사전의 세계

- 아날로그 시대에는 박물관의 큐레이션과 학자들의 강의·저술을 다른 일로 취급할 그만한 이유가 있었다. 박물관의 수장고와 전시실, 대학의 연구실과 강의실은 서로 다른 독립된 공간이었고, 그 각각의 공간에서 일하는 사람들은 서로 다른 일을 하는 전문가로 취급되었다. 지식의 성격과 내용은 유관한 것일지라도, 들여다보는 대상물이 다르면 연구방법이 달라지고, 그래서 그 일은 서로 다른 일이었다. 고객들도 그들의 분업 체계를 당연한 것으로 받아들였을 것이다. 유물을 보기 위해서는 박물관으로, 문서를 확인하기 위해서는 기록관으로, 그 증거들을 학술적으로 설명하는 지식을 얻기 위해서는 대학의 강의실로 가야 했다.
- 하지만 그 모든 지적 감성적 경험의 세계에 '디지털'이라고 하는 새로운 변수가 등장하면서 상황은 달라지기 시작했다. 현대의 지식 고객들은 박물관의 유물이든, 아카이브의 문헌기록이든, 그것들을 가지고 역사를 탐구한 인문학자들의 논술이든, 이 모든 것들을 사이버 공간에서 디지털화된 콘텐츠로 접하는 데 점점 익숙해져 가고 있다. 이 세계에서는 박물관과 강의실이 분리된 공간일 필요가 없다. 유물의 형상과 색체를 보이는 이미지 자료와 그것의 제작기법이나 역사적 배경을 설명하는 연구논문이 다 같이 디지털 데이터로 존재하는 곳에서 두 세계를 넘나드는 행위는 한 번의 마우스 클릭이면 족하다.

## 디지털 전환, 고객은 누구?

### 1. 프로슈머(Prosumer)

- '전문가'의 학술과 '대중'의 즐길거리 사이에서 중재적 역할을 하는 매개적 커뮤니케이터
- 유튜버, 블로거, 위키백과의 기고자들, 디지털 콘텐츠 제작자들.....
- 시맨틱 데이터 아카이브를 가장 먼저 반기고 이용할 사람들

### 2. 디지털 원어민(Digital Natives)

- 환경에서 태어나, 누구나 디지털 프로슈머로 자라날 미래 세대들
- 디지털 클라우드의 세계에서 다양한 지식의 연결고리들을 추적하여 자신의 이야기를 펼칠 수 있는 기술적 능력자
- 시맨틱 데이터 아카이브는 디지털 원어민들이 인문학을 공부하고 인문교양인으로 성장할 수 있게 하는 교육환경으로 활용되어야 한다.

## 디지털 전환, 고객은 누구?

### 3. 인공지능(Artificial intelligence)

- 우리가 궁극적으로 기대하는 시맨틱 데이터 아카이브의 타겟 오디언스는 인공지능
- 시맨틱 데이터 아카이브를 구성하는 RDF(Resource Data Framework)\* 형식의 데이터는 인간의 언어 속에 모호하게 담겨있던 인문학 지식을 컴퓨터가 인식할 수 있는 명시적인 데이터로 변환한 것.
- 이것은 딥러닝 프로세스의 1차적인 학습 데이터로 사용될 수 있으며, 인공지능 스스로 새로운 지식을 찾아서 배울 수 있게 만드는 촉진제 역할을 할 수 있다.

---

\* 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/>)