

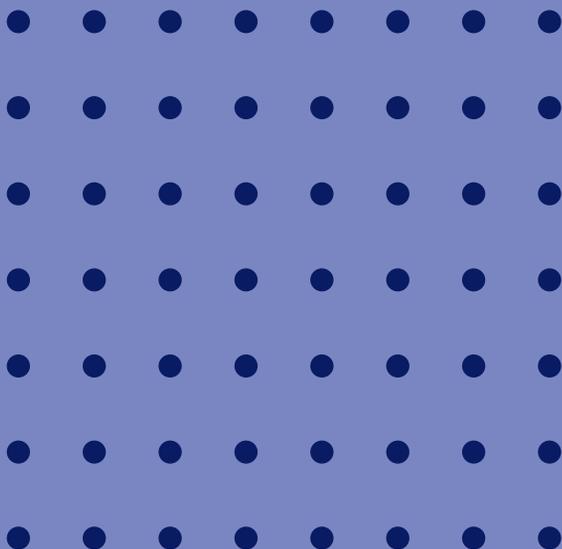
2019 KCA Media Issue & Trend

02.1

전문가리포트

메모리뱅크(Memory bank) 시대에 살다 : BBC 디지털 아카이브

1. 들어가며
2. 디지털 제작 라이브러리
 - 2.1. 상업적 가치 실현
 - 2.2. 공영 방송 서비스의 목표 구현
3. 디지털 아카이브; 미래지향적 사업
4. 마치며



메모리 뱅크(Memory bank) 시대에 살다 : BBC 디지털 아카이브

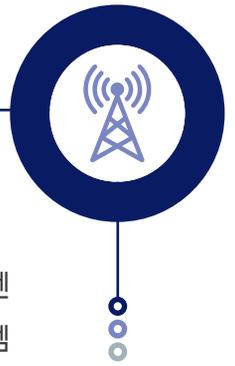
밀너 존 (John Millner, 전 BBC 온라인 콘텐츠 책임 프로듀서)
정현숙 (EBS PD/ 교육사회학 박사)

요약문

디지털 미디어 자산 관리가 관심사로 등장한 21세기, 세계 공영방송사의 대명사로 통용되고 있는 BBC의 디지털 아카이브 운용 현황을 살펴본다. BBC 디지털 아카이브의 취지 및 목표, 특히 공영방송으로서의 BBC가 추구하는 방향을 탐구해보았다. BBC 디지털 아카이브는 BBC 제작자뿐 아니라, 미래 제작에 참여할 인력은 누구든지 거대한 분량의 기존 자료 라이브러리에 쉽게 접근할 수 있도록 체계화하고, 아카이브에 내재되어 있는 잠재적 상업 가치를 실현할 수 있도록 한다. 또한 교사, 학생은 물론 BBC 수신료를 지불하는 모든 국민들에게 아카이브에 있는 풍부한 교육적 자료를 개방하여 배움을 확장할 수 있는 기회를 제공하는 공적 책무를 수행하고자 한다.

1. 들어가며

디지털 미디어 아카이브 및 자산 관리 시스템은, 방송사·제작사 등 전 세계 모든 메이저 미디어 생산자들의 주요 이슈이다. 매일 매일 수백 개의 TV, 라디오 프로그램과 온라인 콘텐츠를 생산하는 영국 공영방송사 BBC 역시 디지털 아카이브를 주요 당면 과제로 여겨 왔다. 거의 100년에 걸쳐 콘텐츠 자산을 축적, 보유해 온 만큼 관리 문제가 중대했기 때문이다. BBC는 비디오와 오디오, 온라인 콘텐츠에서 출판물에 이르기까지, 방대한 분량의 콘텐츠를 지난 5년간 차근차근 디지털화 해 오고 있다. 2012년 런던 올림픽 준비는 이런 막대한 인덱싱(indexing) 프로젝트가 가속화되는 계기가 되었다. BBC 스포츠가 총 2,500시간에 달하는

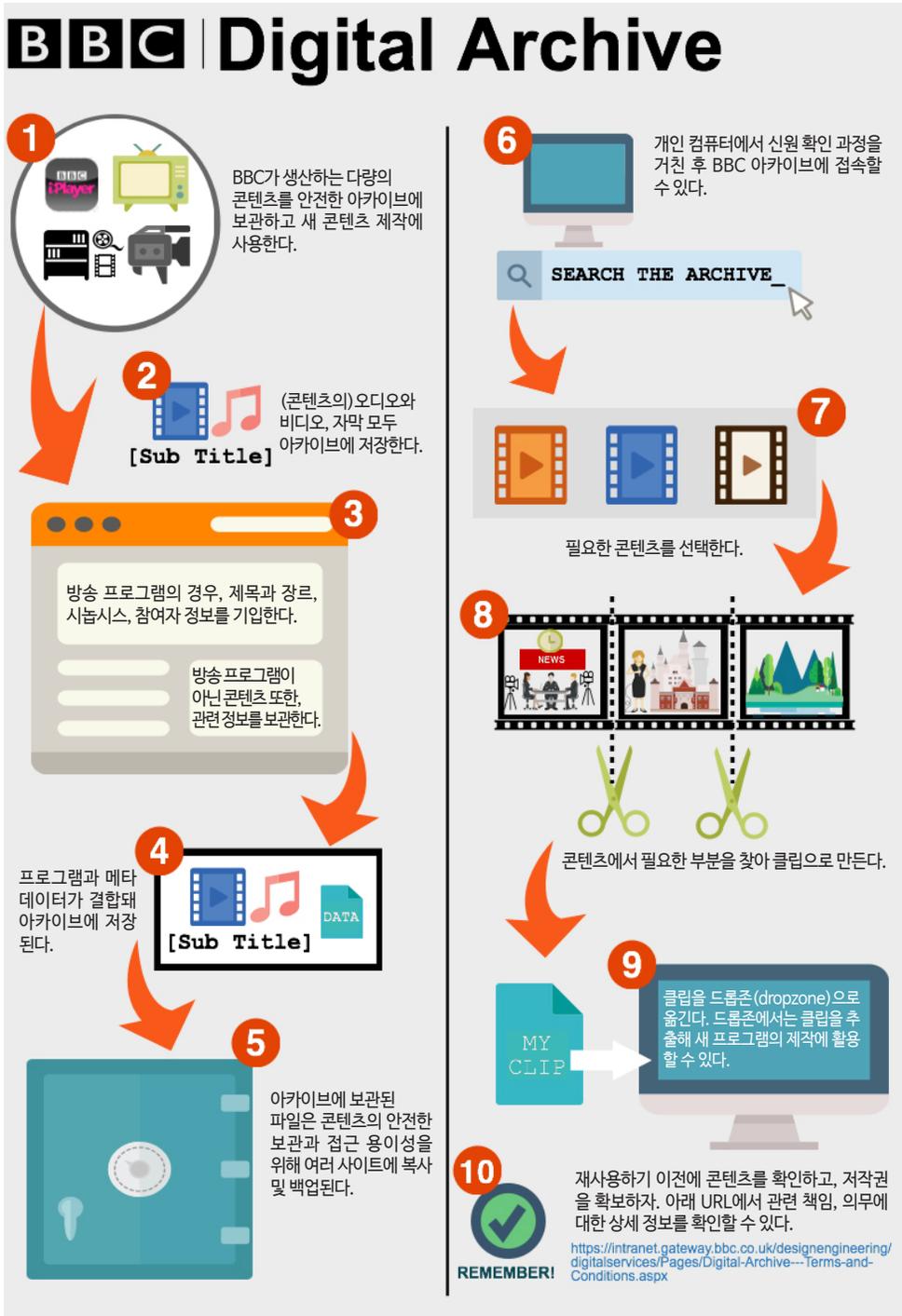


방송프로그램과 국가·종목·선수별로 검색 가능한 수천 페이지에 달하는 정보를 온라인 콘텐츠로 제공하고자 했기 때문이다. 뉴스와 스포츠 콘텐츠를 시작으로, 디지털 아카이브 시스템을 막대한 온라인 콘텐츠 정비에 활용했다. BBC 아카이브팀은 디지털 파일로 전송되기 시작한 2014년 이후의 프로그램에 대한 아카이브 작업을 우선으로 하였다. 이후 테이프로 전송된 그 이전 콘텐츠에 대한 디지털 변환을 거쳤으며, 20세기 초 제작된 BBC 초창기 콘텐츠까지 아카이빙 프로젝트에 포함시켰다. 아카이브 작업의 목표는 기존 콘텐츠에 대한 접근성을 높이는 것에 한정되지 않았다. 현재 제작되고 미래에 만들어질 모든 콘텐츠를 조직적으로 정렬하고자 하며, 회사 내부는 물론, 수신료를 지불하는 국민을 포함하여 제 3자에게까지 제공하고자 하는 목적을 가지고 있다.

2. 디지털 제작 라이브러리

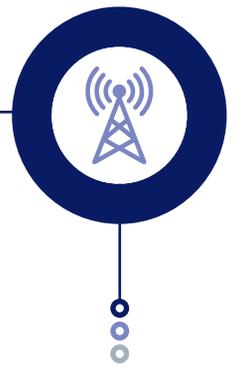
BBC 디지털 아카이브는 BBC의 모든 방송 프로그램과 클립, 비방송용 콘텐츠를 저장하고 제공하는 온라인 시스템이다. 아카이브에 저장되는 콘텐츠는 시놉시스, 참여자 정보, 자막, 캡션, 저작권 정보 등 제작 관련 데이터뿐만 아니라 카탈로그 설명, 비디오 키 프레임(key frames), 관련 외부 링크 등 전담 아카이브팀이 추가하는 메타데이터까지 포함한다. 즉, 콘텐츠 아카이빙 과정을 거치면서 더욱 방대한 메타데이터가 자동적으로 축적될 수 있다. 아카이브는 또한 기본적으로 BBC 콘텐츠 제작자들의 제작 리소스로도 활용될 수 있다. 제작자들은 아카이브에서 필요한 콘텐츠를 찾아서 원하는 특정 클립을 제작하고, 편집실에서 이를 추출해 낼 수 있다. 다음 인포그래픽에서 이 과정을 확인해 보자.

그림 1 BBC 디지털 아카이브 과정 및 사용



출처 : BBC Academy 2017





2.1. 상업적 가치 실현

디지털 아카이브 프로젝트의 주된 효과로, BBC 월드와이드(BBC Worldwide)의 아카이브 사업을 들 수 있다. 디지털 아카이브 프로젝트를 통해 BBC 월드와이드는 보다 효과적이고 원활한 비즈니스 아카이브 사업을 수행하게 되었으며, BBC가 오랫동안 확보해 온 문화적·사회적·역사적·정치적 자료들의 재원적 가치를 제고할 수 있게 되었다.

BBC 모션 갤러리(BBC Motion Gallery)는 현재 스톡(stock) 콘텐츠 서비스 기업인 게티 이미지(Getty Images)와의 파트너십을 구축하고 있다. 다운로드 저작권이 확보된 클립 125,000개의 온라인 데이터베이스를 보유하고 있으며, 사용자 요청에 따라 특별 전담팀이 플랫폼과 무관하게 사용 가능한 오프라인 자료 수백만 시간 분량을 제공하는 전 세계적 자료 라이선싱 서비스도 운영하고 있다.

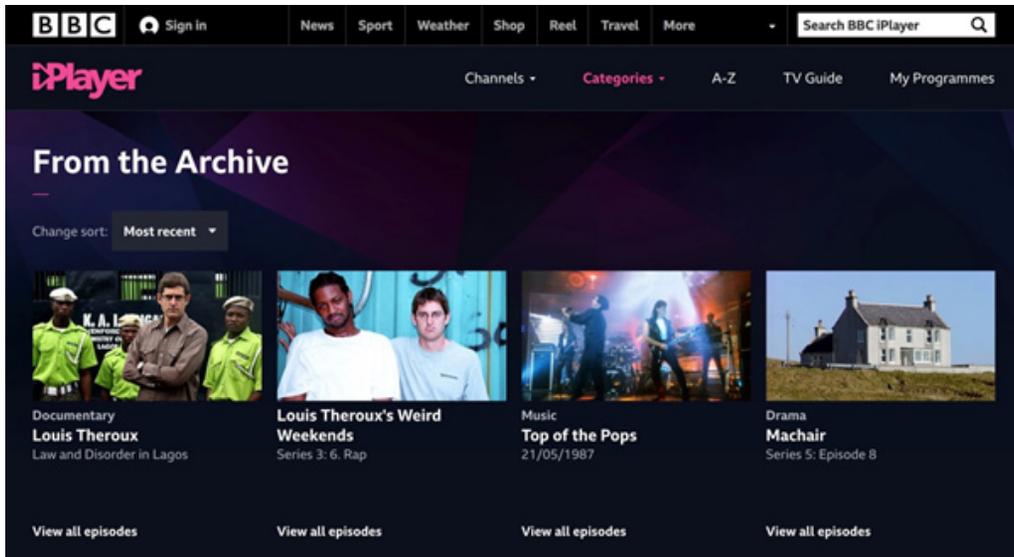
전적으로 이윤 추구를 목표로 하는 모션 갤러리와 함께 BBC 월드와이드 러닝(BBC Worldwide Learning)은 BBC 콘텐츠를 주축으로 하여 영국 개방대학(Open University), 일본 NHK, 미국 CBS 등의 콘텐츠까지 포함한 수천 개의 교육용 클립과 맞춤형 비디오를 제공하고 있다. 이러한 콘텐츠들은 다양한 교육 커리큘럼에 맞추어 색인되어 있으며, 영국 외 지역의 학생 및 교육자들에게 저렴한 가격으로 제공하고 있다.

2.2. 공영 방송 서비스의 목표 구현

디지털 아카이브는 상업적 이윤을 추구하는 것 이상의 목표가 있다. 바로 공적 서비스의 제공이다. ‘교육적’, ‘창의적’, ‘보편적’이라는 공공적 목표를 이루기 위해 BBC는 디지털 아카이브를 더 많은 국민들에게 폭넓게 개방하고자 노력하고 있다. 그 노력 중 하나는, BBC iPlayer 사례이다. BBC iPlayer는 시청자나 청취자가 놓쳤던 BBC의 TV, 라디오 콘텐츠를 디지털 기기로 무상 제공하는 인터넷 스트리밍 플랫폼이다. 최신 콘텐츠를 주로 다루고 있는 듯 보이나, 2017년 말에서 1946년까지 거슬러 올라가 기존에 제작된 수백 개의 우수 프로그램들을 새 아카이브 카테고리¹⁾에 추가하면서 콘텐츠의 양이 풍성해졌다.

1) <https://www.bbc.co.uk/iplayer/categories/archive/featured>

그림 2 BBC iPlayer



출처 : BBC 아카이브, BBC iPlayer

BBC iPlayer는 정보, 오락 제공 목적과 더불어, 교육 제공이라는 BBC의 공익적 목적에 부합하고자 영국의 학교와 학생들에게 무상으로 ‘BBC Teach’s Class Clips’ 서비스 등 학습 자료를 제공한다. BBC 아카이브를 통해 제공되는 본 서비스는 영국 학교의 교육 과정에 맞게 주제 별로 분류되어 있으며, 구체적 학습 목표를 전달하는 데 도움이 되는 짧은 비디오 클립들로 구성되어 있다.²

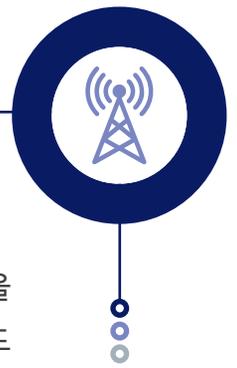
교육용 콘텐츠 아카이브의 활용 사례를 살펴보면, 우선 전 세계 문학인들의 사랑을 받는 셰익스피어(Shakespeare) 관련 자료들을 모아 놓은 컬렉션이 있다. BBC Teach를 통해 학교 사용자에게 제공되는 이 자료들은 Learning on Screen의 Box of Broadcasts 와의³ 파트너십으로 고등교육 학습자에게도 제공되고 있다.

둘째, Research and Education Space (이하 RES)라 불리는 파일럿 파트너십 프로젝트이다. RES는 서비스 제공 잠재력이 있는 업체들에게 전문 아카이브 검색엔진을 사용하도록 하

2) <https://www.bbc.com/teach/about-teach/znhhbmd>
BBC Schools에서 바뀐 사이트로, 교사들을 위한 학습 자료들이 탑재되어 있다.

3) <https://learningonscreen.ac.uk/ondemand>





며, BBC 아카이브뿐만 아니라 박물관, 도서관, 갤러리의 컬렉션에 이르는 넓은 범주의 검색을 제공하여 고등교육 현장에서 교사, 학생, 연구자들의 교육과 학습을 다양하고 내실 있게 만드는데 기여하고 있다. RES는 BBC가 만든 오픈 플랫폼으로 이루어져 있으며, 이 플랫폼은 주로 BBC 문서, 데이터, 이미지, TV와 라디오 프로그램을 다루지만 영국 및 그 외 지역의 공공기관에서 제공하는 풍부한 해설 자료들까지 색인하고 체계적으로 정리하는 프로젝트로 정의된다.⁴ 교육용 콘텐츠 아카이브보다 더 대중적 소구력을 가진 ‘게놈 프로젝트(Genome Project)’도 있다. BBC TV, 라디오 프로그램을 소개하는 잡지 <라디오 타임즈(Radio Times)>에 85년간 실린 모든 프로그램 정보를 디지털화하여 분류한 프로젝트이다. 게놈 프로젝트는 온라인 커뮤니티가 직접 편집할 수 있는 형식으로, 약 5백만 개에 달하는 방송 프로그램의 정보를 쉽게 찾을 수 있도록 아카이빙하고 있다. 현재 정식 서비스를 시작하지는 않았으나 웹사이트에서 내용을 확인할 수 있다.⁵

그림 3 BBC <라디오 타임즈(Radio Times)>



출처: BBC 아카이브

4) <https://bbcarchdev.github.io/res/>

5) <https://genome.ch.bbc.co.uk/>

3. 미래지향적 사업으로서의 디지털아카이브

BBC의 디지털 아카이브 프로젝트는 공영방송 창립 이래 축적된 방대한 자료들의 접근성과 유용성을 높이는 것 이상의 의미가 있다. 본 프로젝트는 본질적으로 미래지향적인 사업이자, BBC 전체를 아우르는 대규모 디지털 분류 체계이다. 어떤 아카이브 프로젝트든 메타데이터를 기반으로 시작, 완성되며, 어느 시기에 제작된 콘텐츠든 디지털 메타데이터를 생성·추가하기 위해 유연한 미래적 시스템(future proof system)에 의존하기 마련이다. 그래야만 다양한 방법으로 분류·개방되어, 접근성을 높일 수 있다. 디지털 시대에 걸맞은 커다란 도전 과제인 디지털 아카이브 프로젝트는, BBC처럼 규모가 크고 역사가 깊은 조직에서는 더욱 그 의미가 크다. 가령, 뉴스와 TV, 라디오 프로그램, BBC 온라인 (BBC Online) 콘텐츠와 더불어 자체적인 제작 문화를 지켜온 BBC 지역 방송사들의 콘텐츠까지 아카이빙하는 작업은 간단할 리 없다. (BBC의 지역 방송사들은 독자적인 콘텐츠 제작, 관리 시스템을 보유하고 있으며, 관련 데이터 역시 다른 방식으로 생성, 조직화한다.) 과연 이처럼 다양하고 거대하며, 지속적으로 증가하는 콘텐츠를 모두 관리할 수 있는 통합 메타데이터 시스템을 개발할 수 있을까? BBC는 그 해답을 링크드 데이터(Linked Data)에서 찾고 있다.

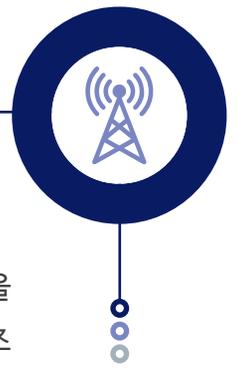
링크드 데이터⁶란 웹상의 내용, 데이터들을 상호 연결해, 그 의미를 컴퓨터가 이해하고 소통할 수 있도록 하는 모형이다. 궁극적으로 ‘기계적’ 혹은 ‘자동적’으로 연결된 웹을 지향하는 모델이라 볼 수 있다. 컴퓨터가 정보를 이해해 추론할 수 있도록 한다는 측면에서 시맨틱 웹 기술(semantic web technology)⁷에 속한다.

BBC는 2011년경 온라인 뉴스, 스포츠 콘텐츠를 필두로 하여, 콘텐츠 메타데이터에 링크드 데이터 요소를 사용했다. 지속적인 개발 결과, 현재는 교육용과 논픽션 온라인 콘텐츠의 체계적 조직화를 위해서 완전히 개발된 링크드 데이터 플랫폼을 사용하고자 노력하고 있다. 링

6) 디지털 메타데이터를 구성하고 게재하는 표준 방식으로 웹상에서 자동적으로 공유되고, 쉽게 업데이트되며 사람은 물론 기계가 식별할 수 있게끔 되어 있는 시맨틱 웹의 상호 연결 집합체이다.

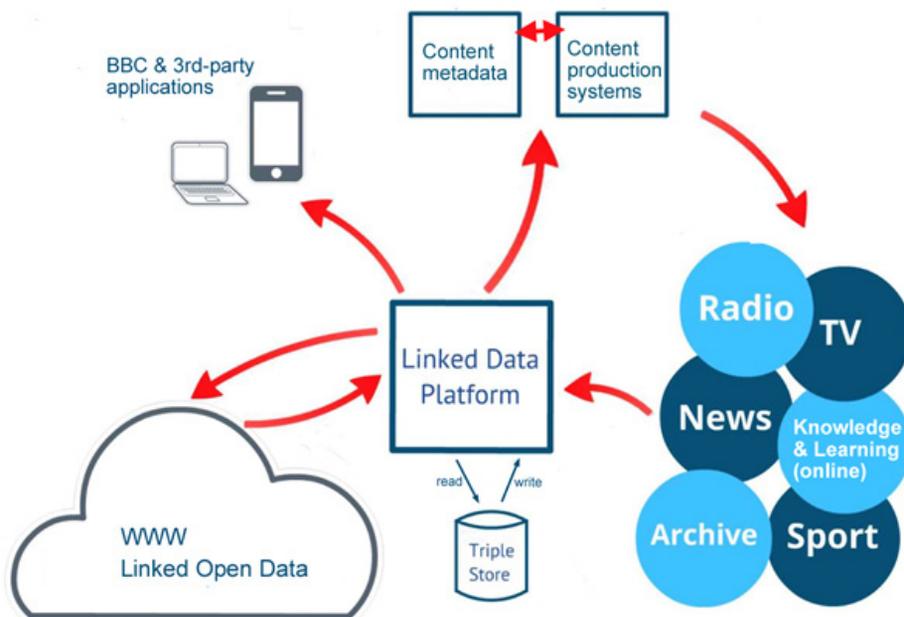
7) 시맨틱 웹은 WWW의 다음 단계로 볼 수 있는데, Web 3.0으로도 알려져 있다. “현재의 인터넷과 같은 분산 환경에서 리소스(웹 문서, 각종 화일, 서비스 등)에 대한 정보와 자원 사이의 관계-의미 정보 (semantics)를 기계(컴퓨터)가 처리할 수 있는 형태로 표현하고, 이를 자동화된 기계(컴퓨터)가 처리하도록 하는 프레임워크이자 기술이다. 시맨틱은 기계끼리 대화를 주고받는 것으로, 웹을 지능형으로 만들어, 컴퓨터들이 웹상의 모든 데이터와 콘텐츠, 링크들을 분석해 인간과 컴퓨터 간의 대화를 처리하는 것이다.” (출처: 매경닷컴)





크드 데이터 플랫폼은 디비피디아(DBpedia)⁸와 같은 오픈 웹 링크드 데이터 처리 기술을 활용하여 BBC 콘텐츠를 탐색하는데, 리소스 메타데이터를 3가지 시멘틱 구조 - ‘트리플즈 (triples)⁹’로 알려져 있으며 트리플 저장소 데이터베이스에 저장됨- 의 형태로 구성하는 논리적 방법을 사용하여 BBC 콘텐츠 자산을 조회하고 공개 웹 데이터 링크드의 데이터 집합을 사용하여 태그를 지정한다. BBC 콘텐츠 관리 시스템의 새로운 메타데이터인 링크드 데이터는 자체 또는 타사의 플랫폼이나 애플리케이션에서도 BBC 콘텐츠를 추출할 수 있도록 지원한다.

그림 4 링크드 데이터(Linked Data) 활용 방식



출처 : David Rogers(BBC Future Media & Technology Senior Technical Architect)

물론 링크드 데이터에 쓰이는 데이터들은 콘텐츠와 묶여 있지 않지만, ‘공영방송’, ‘세익스피어’, ‘정보 아키텍처’와 같이 특정 지식 도메인이나 활동 영역에 연동되어 있다. 사람들이 이해하기 쉽도록 링크드 데이터 시스템은 도메인 특정 개념 컬렉션과 카테고리는 물론이고 온톨로

8) 위키피디아로부터 구조화된 정보를 추출, 링크드 데이터 형식으로 저장해 월드 와이드 웹에서 이용할 수 있도록 만들기 위한 커뮤니티이다.
 9) 문장의 문법 구조와 같이 3가지 구성요소, 주제어(subject), 목적어(object), 술어(predicate)를 활용해 정보를 조직화하는 시멘틱 웹 포맷이다.

지(Ontology)¹⁰로 알려진 그것들의 속성과 관계를 참조해야 한다. BBC는 2012년부터 뉴스·스포츠를 시작하여 음악·라디오·교육 등의 카테고리를 확장하며, 디지털 콘텐츠 아카이브에 적용할 온톨로지를 개발해왔다. 이러한 온톨로지는 여전히 개발 중이고, 진화하고 있다. BBC는 자사 웹사이트에서 현재 진행 중인 작업에 관련한 정보를 계속 공유하고 있으며, 다음과 같이 세 가지 유형으로 모델링하여 온톨로지를 분류하였다:

- 1) 창의적 작업(creative works) - 프로그램, 뉴스 기사, 웹페이지 등
- 2) 사물(things) - 핵심 개념, 도메인, 인물, 장소, 기관, BBC가 만드는 콘텐츠의 소재로 쓰이면서 시청자가 관심 있어 할 만한 이벤트 등
- 3) 데이터 관리 - 링크드 데이터의 소유권이나 운영을 지원하는 온톨로지를 그룹화

BBC는 온톨로지의 활용에 대해 다음과 같이 밝히고 있다: “BBC는 시청자들에게 중요하고 유일한 내용들을 매우 풍부하고 다양한 콘텐츠로 제공하고 있으며, 링크드 데이터는 그 내용들과 관련한 콘텐츠에 (사용자들이 찾아갈 수 있는) 연결고리 기반이 되고 있다. BBC는 온톨로지를 사용해 해외 소식이나 BBC의 콘텐츠는 물론, 링크드 데이터 플랫폼 내에서 데이터들을 운영, 저장, 공유하고 있다.”

사물(things) 온톨로지의 한 예로, BBC가 교사나 학생에게 콘텐츠를 지원하는 영국 학교교육 커리큘럼을 살펴보자. 아래 그림은 잉글랜드와 웨일스의 11~14세 학생들을 대상으로 한 Key Stage3(KS) 과정인 ‘에너지’ 주제에 관한 것이다. 이 모델은 에너지가 하나의 학문 영역뿐만 아니라 2개의 학문 영역에서 역할을 수행한다는 것을 보여줌으로써 도메인 내의 객체간의 주요 관계를 보여준다. 에너지가 하나의 도메인에 나타나는 데 그치는 것이 아니라, KS3 물리학과 KS3 지리학 프로그램은 커리큘럼 수준 KS3의 과학 분야에도 포함된다.¹¹

10) 특정 지식 도메인 내에서 사용되는 개념이나 용어들의 정의를 내려주고 관계를 파악하여 서로 연결시켜주는 형태의 일련 세트라 할 수 있다.

11) <https://www.bbc.co.uk/blogs/internet/entries/31966f38-2164-31ac-bbce-59257bf97c79>



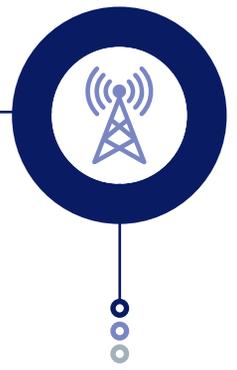
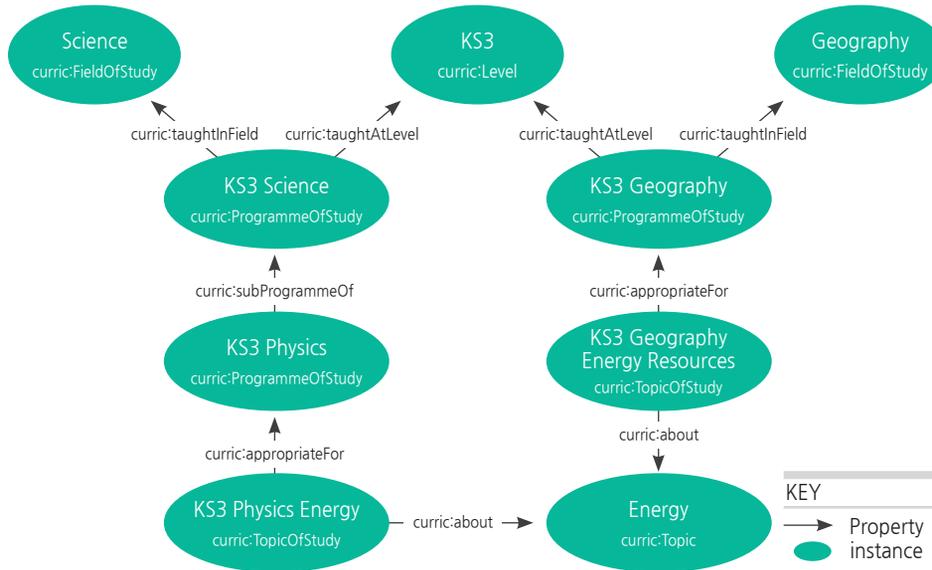


그림 5 온톨로지(ontology) 커리큘럼 예시

Describing content



The curriculum ontology

출처 : BBC

콘텐츠 온톨로지는 이렇게 링크드 데이터와 결합해 메타데이터를 생성하고, 정확하게 업그레이드된다. 이로써 더욱 세분화된 미디어, 콘텐츠의 개인화(personalisation)가 가능해지는데, 이 개인화(맞춤형 서비스의 제공)는 BBC가 전략적으로 우선순위에 놓는 중요 과제이다. 지난해 BBC 디자인 엔지니어링 부서 내에 신설된 BBC 데이터랩(BBC Datalab)팀은 개인화 서비스 강화를 위해 다음 세대의 미디어 소비자를 위한 온톨로지와 빅 데이터, 그리고 머신러닝, 지능 머신의 출현에 밑받침이 될 시멘틱 웹 기술 등 새로운 기술을 다루는 업무를 수행하고 있다. 데이터랩은 BBC 콘텐츠와 소비자에 관한 BBC 데이터를 한 데 모을 수 있는 클라우드 기반의 플랫폼을 구축하였다. 이 팀의 비즈니스 분석가 스베틀라나 비데노바(Svetlana Videnova)는 “새 플랫폼은 데이터 과학자들이 BBC 콘텐츠에 대한 대중들의 요구를 충족시키는 다양한 방법을 찾아나가는 기반이 될 것이다. BBC 사용자들은 이를 통해 새롭게 선보일 BBC+ 애플리케이션에서 보다 고도화된 개인화 서비스를 체험해 볼 수 있을 것이다” 라고 강조한다.

BBC+는 뉴스 콘텐츠에 집중돼 큰 성과를 거두지 못했던 기존의 모바일 클립 애플리케이션을 실험적 형태로 업데이트한 것이다. 새로운 버전의 BBC+ 애플리케이션은 BBC 디지털 아카이브와 데이터 저장소를 활용함으로써 사용자들에게 더욱 세분화된 추천 서비스를 제공한다. 사용자들의 활동 이력(activity history)을 학습하여 선호도를 이해·분석함으로써 보다 광범위하면서도 개개인에 맞는 내용의 맞춤형 콘텐츠를 서비스할 수 있게 되었다. 현재 BBC+ 애플리케이션은 안드로이드¹²와 iOS 앱스토어¹³에서 찾아볼 수 있다.

4. 마치며

지금까지 BBC 디지털 아카이브 개발의 추진 동력들을 분석해 보았다. 효율적인 디지털 아카이브 구축은 아카이브의 상업적 가치 실현은 물론 교육적 가치를 극대화하고, 시멘틱 웹에 걸맞게 콘텐츠 메타데이터를 합리적으로 조직화하기 위해서 대두되었다. 디지털 아카이브는 장차 BBC 수신료를 책임지게 될 미래세대와 함께 하는 공영방송사로 건승하기 위한 시기적절한 미래 전략이면서, 또한 사용자가 원하는 맞춤형 콘텐츠 제공함으로써 폭 넓은 대중에게 어필하는 공영방송의 이상을 실현하는 길이라 생각된다.

BBC의 매튜 포스트게이트(Matthew Postgate) 기술본부장은 최근 한 연설에서 “머지않은 장래에 모든 미디어를 온라인상에서 만날 수 있을 것이다. 인터넷 기술의 발달로 시청자들은 경험의 폭을 더욱 넓히며 자신에게 맞는 콘텐츠를 소비할 수 있게 될 것이다”라고 하였다. 메타데이터, 지식 기반 온톨로지, 개방형 링크드 데이터 등의 기술로 거대한 규모의 글로벌 미디어가 더욱 빠르게 확장되고 있으며, 그 확장은 사용자들이 접근하기 쉽도록 조직화 및 의미 구축의 과정으로 이루어지고 있다. 디지털 기술의 도움으로 미디어는 새로운 국면을 맞이하고 있다. 모든 콘텐츠는 잊힐 위험 없이 계속해서 저장될 뿐만 아니라 보다 확장되면서 풍성하게 공유될 것이다.

12) https://play.google.com/store/apps/details?id=uk.co.bbc.bbc_plus

13) <https://itunes.apple.com/gb/app/bbc-the-bbc-just-for-you/id1110317391?mt=8>