

메타데이터의 표준화 사례

글. 이형극 한국전자통신연구원(ETRI) 선임연구원

우리는 이전 기고 내용을 통해 방송 콘텐츠의 메타데이터가 가지는 중요성에 대해서 다루었다. 방송 콘텐츠의 부가정보로서 메타데이터를 보유하게 되면 활용성의 증가는 당연히 기대할 수 있지만 메타데이터를 구축하여 활용하는 데에 어려움과 제한사항도 만만치 않다는 사실도 함께 언급했다. 그 이유 중 한 가지는 방송 콘텐츠를 만드는 주체가 다양하기도 하고, 이에 관한 하나의 표준화된 체계를 만드는 데에도 어렵다는 것이었다. 그래서, 미디어 콘텐츠의 메타데이터를 하나의 체계화된 규격으로 만들어 내고자 노력한 표준화의 실제 사례를 살펴보고 미디어 콘텐츠의 메타데이터를 보유하는 데에 제약점을 극복하는 방법을 논의해 보려고 한다.

표준화의 목적

‘국가표준인증 통합정보시스템’에 따르면 표준화(Standardization)란 ‘실제적이거나 잠재적인 문제들에 대하여 주어진 범위 내에서 최적 수준을 성취할 목적으로 공통적이고 반복적인 사용을 위한 규정을 만드는 활동’이라고 정의하고 있다. 즉, 표준화는 우리가 일상적으로 사용하는 생활용품에서부터 자동차, 비행기 등 모든 제품 및 부품의 치수, 성능, 재질, 시험방법 등을 통일화 및 단순화시켜 기준에 따르도록 하는 것으로 이해관계자의 합의, 공개원칙, 자발성 존중, 통일성과 일관성 유지, 시장 적합성, 경제성, 공공이익의 추구 등 여러 원리를 기초로 만들어지고 있다. 표준화의 목적은 제품, 프로세스 또는 서비스 등 표준화의 대상이 본래의 의도된 목적으로 작용하도록 개선하여 모든 이해관계자의 원활한 의사소통, 공공이익 및 무역 장벽의 제거 등을 추구하는 데 있다. 또한, 생산, 소비, 유통 등 여러 분야에 있어서 능률증진 및 경제성 향상을 통해, 제품의 품질 개선과 생산 능력의 향상, 상거래의 단순화 및 공정화의 형태로 그 효과가 나타나고 있다.¹⁾ 그렇다면 메타데이터 표준화의 목적은 미디어 콘텐츠의 부가정보를 누구나 쉽게 사용할 수 있도록 표준을 만들어 메타데이터의 활용성을 높이려는 목적을 가진다고 할 수 있다.

국의 표준화 단체 중심의 대표적인 표준화 사례

더블린 코어

인터넷을 통해 접근할 수 있는 모든 미디어 콘텐츠를 ICT 기술의 발전으로 메타데이터로 기술되는 원본 콘텐츠의 유형과 이런 콘텐츠에 접근하는 방법이 다양해졌으나, 각각 표준의 메타데이터마다 그 요소를 달리하고 있다. 따라서 데이터의 호환성을 유지하고 네트워크로 연결된 콘텐츠를 설명하는 데에 필요한 일련의 데이터 요소를 정의하고 이들 콘텐츠를 신속히 검색하기 위해 만들어진 표준이 바로 더블린 코어이다. 더블린 코어는 1995년 OCLC(Online Computer Library Center)와 NCSA(National Center for Supercomputer Application)가 더블린에서 개최한 워크숍에서 합의한 메타데이터 표준이다. 즉, 더블린 코어의 메타데이터는 여러 환경 속에서 데이터의 호환성을 유지하고 네트워크 자원의 기술에 필요한 데이터 요소를 구성하며, 자원의 신속한 검색을 목적으로 기존에 사용하던 표준이나 프로토콜을 수정하고 네트워크 자원을 정의하고 접근하는 방안을 제시하고 있다.

1. 국가표준인증 통합정보시스템 ‘표준화의 정의’ 내용 발췌, <https://standard.go.kr/KSCI/standardIntro/standardView.do?menuId=505&topMenuId=502#none>

 DCMI 2022 Conference Login Q	
Specifications ▾	Events ▾
Community ▾	News
Resources ▾	About DCMI ▾
/ Specifications / Dublin Core™ / DCMI Metadata Terms DCMI Metadata Terms	
Title:	DCMI Metadata Terms
Creator:	DCMI Usage Board
Identifier:	http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/
Date issued:	2020-01-20
Latest Version:	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
Version History:	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/release_history/
Document Status:	This is a DCMI Recommendation.
Description:	This document is an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes.

더블린 코어 표준 예제

MPEG-7

MPEG-1, MPEG-2, MPEG-3 및 MPEG-4 등 다양한 멀티미디어 콘텐츠의 코딩 방식 및 데이터 압축에 대한 표준을 보유하고 있는 MPEG 그룹에서 추진한 ISO/IEC(International Organization for Standardization/International Electrotechnical Commission) 표준인 MPEG-7은 ‘멀티미디어 콘텐츠를 설명하기 위한 인터페이스(Multimedia Content Description Interface)’라는 명칭을 가지고 있다. 다양한 멀티미디어 콘텐츠의 활용성을 높이기 위해, 검색 및 관리 기능에 활용할 수 있도록 부가적인 메타데이터를 지원하는 표준이라고 할 수 있다. MPEG 그룹이 가지고 있는 다양한 멀티미디어에 대한 정보를 표준적으로 기술하여 콘텐츠의 제작, 전송, 저장, 유통, 검색, 아카이빙, 데이터 방송, 전자상거래 등에 응용하려는 것을 목표로 하고 있다고 명시하고 있다. 이러한 다양한 응용 분야를 고려하여 고안되었기 때문에 다른 메타데이터 표준 모델을 형성하는 데에 직접적으로 영향을 준 표준이기도 하지만, 표준 체계를 초기 정립 시기의 원대한 기획 의도와는 다르게, 메타데이터 모델 크기가 비대하여 실제로는 활발하게 활용되지 못했다.

TV-Anytime

디지털 방송이 시작되고 방송 매체가 다양해지면서 방송 채널과 프로그램 수가 급격하게 증가하였는데 기존 프로그램 가이드만으로는 수많은 방송 채널 중에서 시청자가 자신이 원하는 프로그램을 찾는 데에는 한계가 있었다. 민간 국제표준단체인 TV-Anytime 포럼에서는 메타데이터를 기반으로 시청자가 원하는 미디어 콘텐츠를 원하는 시간에 소비할 수 있는 맞춤형 방송 서비스를 제공하는 방법에 대한 표준 및 규격을 제공하였다. 시청자가 PDR(Personal Digital Recorder) 또는 PVR(Personal Video Recorder) 등 디지털 방송 단말을 이용하여 미디어 콘텐츠를 언제 어디서나 자유롭게 빠르게 검색하여, 수집하고 시청할 수 있도록 환경을 조성하는 것이 TV-Anytime의 목표인 것이다. TV-Anytime 메타데이터는 XML(eXtensible Markup Language) 형식으로 표현되며, 방송 환경과 제공하고자 하는 기능에 따라 1단계(Phase 1)와 2단계(Phase 2)의 표준 규격을 제정했다. 1단계 표준 규격에는 사용자 단말을 고려하여 콘텐츠 검색 및 선택, 소비하는 환경에 필요한 콘텐츠 참조, 메타데이터 등에 대한 표준을 제정했고, 2단계 표준에서는 양방향 통신 네트워크를 통한 향상된 콘텐츠 소비를 위해 홈 네트워크 환경에서 단말 간의 콘텐츠 공유 및 재분배, 다양한 사용자 환경에 맞는 콘텐츠를 제공하기 위한 기술 규격 등을 제정하였다.

P_META

P_META는 콘텐츠를 제작 및 제공하는 방송사, 제작사, 배급사 등이 모여 만든 조직인 EBU(European Broadcasting Union)에서 만든 메타데이터 규격 체계이다. 구성원의 특징에서 알 수 있듯이 표준 규격을 만드는 목적이 콘텐츠 제작 단계와 시청 및 추가적인 비즈니스 활용 단계까지 사용자 간에 콘텐츠를 주고받는 데에 효율성을 증가하는 데에 있는 것이다.

특히, P_META 메타데이터를 구성하는 요소는 TV-Anytime 메타데이터의 요구사항을 반영하였으며, 콘텐츠 제공자 입장의 P_META와 콘텐츠 사용자(시청자) 입장의 두 단체 간의 긴밀한 협력을 통해 부가적인 비즈니스 모델을 발굴해내도록 유도하고 있다. 또한 방송 프로그램 제작사와 배급사가 서로 거래를 할 때 사용되는 메타데이터 표준으로 활용되어 계약과 관련한 필요정보를 포함하고 있어, 비즈니스 영역에서도 미디어 자산을 활발하게 이용할 수 있는 기반을 마련해 놓은 것이다.

SMPTE

조명, 설비, 필름 부문의 기술적 요소와 연관된 영화, 텔레비전 기술 전문가 집단인 미국 영화·텔레비전 기술인협회(The Society of Motion Pictures and TV Engineers, SMPTE)에서도 메타데이터 표준을 추진하였는데, EBU와 공동으로 작업하여, 방송 콘텐츠를 관리하고 검색하는 데에 필요한 모든 메타데이터를 일차적으로 취합하였다. 그리고 이중 공통되는 메타데이터를 선별하여 활용성을 높이고 또한 메타데이터 항목을 16Byte 길이로 코드화하여 표준화하였다. 여타의 표준이 XML 기반 텍스트로 작성된 데에 반해, SMPTE에서 제안하는 메타데이터 항목은 전체가 코드로 표현되는 차별점이 있다. 물론, 코드가 의미하는 내용을 해석하기 위해 코드북이 별도로 필요하다는 단점이 있지만, 비트스트림 형태로 모든 데이터 항목을 쉽게 표현할 수 있어서 데이터의 저장, 관리나 전송에 용이하다는 장점 또한 동시에 갖고 있다.

SMEF

BBC 기술위원회에서는 콘텐츠의 생성(제작)에서부터 송출, 시청(활용), 자산화 및 관리 등 미디어 콘텐츠의 전 주기 활용에 필요한 데이터 명세 등을 정의하여 미디어 콘텐츠의 대용량 데이터 저장소(아카이브)에 활용 가능한 표준을 추진하였는데, SMEF(Standard Media Exchange Framework) 데이터 모델이 그것이다. 특히 BBC 내부에서만 활용할 수 있는 것이 아니라, 미디어 자산을 개발하거나 관리와 관련된 어떤 그룹이나 조직에서도 활용할 수 있도록 고안되었다.

PBCore

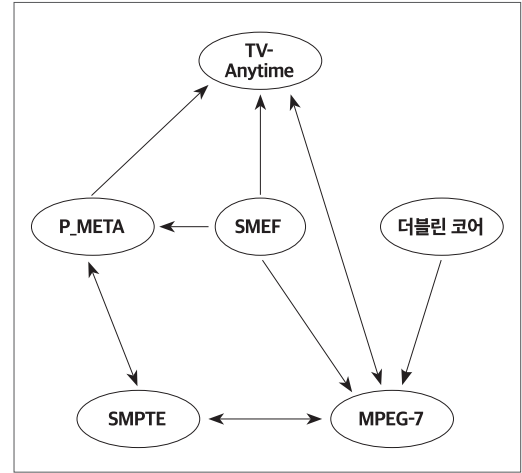
PBCore는 미국 공영방송지원재단에서 라디오와 TV 방송의 메타데이터 개발을 위해 설립한 PBMI(Public Broadcasting Metadata Initiative)에 의해 제안되었으며 더블린 코어를 기반으로 하는 멀티미디어 메타데이터 표준이다. 라디오와 TV 방송국에서 만들어진 콘텐츠에 대한 내용을 설명하고 목록화하는 작업에 필요한 용어와 내용을 가장 함축적으로 표현할 수 있는 체계 및 환경을 구축하는 것에 초점을 맞추었다. 특히, 방송국 간 정보를 교환하고 공유하는 데에 장점이 있으며, 콘텐츠를 자산으로 관리하는 데에 특화되어 각종 미디어 기관 등에서 사용한다고 알려져 있다.

ITU-T SG16

ITU-T(International Telecommunication Union-Telecommunication) SG16의 IPTV 메타데이터 표준 개발 파트에서는 서비스 및 콘텐츠 메타데이터(Service and Content Metadata), 사용자 메타데이터(User Metadata), 콘텐츠 제공 및 관리(Content Provisioning and Management Metadata) 등과 같은 IPTV 서비스를 위한 메타데이터에 대하여 기술하고 있으며, 메타데이터 요소 및 전달 프로토콜에 대한 내용도 포함하고 있다. 2016년 ITU-T SG16/Q13에 장면 단위 시맨틱 클러스터링을 이용한 주문형 비디오 클립 모음 및 부가적인 상품 정보와의 결합이 가능하도록 하는 스마트방송 기술과 관련된 메타데이터 정의 및 서비스 시나리오에 대한 내용을 표준으로 다룬 바가 있다.

앞서 언급한 표준화 내용은 다음과 같이 정리할 수 있다.

명칭	표준 제정 목적	특징
더블린 코어	네트워크 환경에서 콘텐츠의 검색 환경 지원	인터넷으로 연결된 콘텐츠를 사용하는 데에 데이터의 호환성을 유지하고 자원의 신속한 검색을 목적으로 함
MPEG-7	각종 티미디어 콘텐츠의 검색 지원	코덱 전문가 집단에서 추진하여 적용 가능한 콘텐츠 종류가 많은 반면, 메타데이터 구조체계가 비대함
TV-Anytime	시청자 맞춤형 서비스 환경 구축	맞춤형 방송 서비스를 위해 시청자 입장에서의 필요한 내용을 주로 다룸
P_META	콘텐츠 유통 및 전자상거래 서비스 환경 지원	방송사, 제작사, 배급사 주도로 표준화가 진행되어 콘텐츠 제공자의 입장의 내용을 주로 다룸
SMPTE	콘텐츠 전송, 교환 등 전달 방법의 효율화	전문기술 전문가 입장에서 전송 효율화를 위해 최우선 요건으로 고려함
SMEF	콘텐츠 편성, 제작, 저장 및 아카이빙 등 전주기 관리 환경 구축	콘텐츠를 자산(Asset) 관리 측면으로 간주하여, 대용량 아카이브 환경 구축에 용이함
PBCore	방송국 내 콘텐츠 관리 체계 구축	더블린 코어를 기반으로 라디오 및 TV 방송 프로그램을 위주로 고려함
ITU-T SG16	양방향 IPTV 서비스 지원 환경 구축	사용자 중심의 양방향 IPTV 서비스 및 통신융합 미디어플랫폼 환경에 필요한 메타데이터를 발굴함



표준 내용 간 상관관계도

국내 표준화 사례

국내 표준화 기구인 TTA(Telecommunications Technology Association, 한국정보통신기술협회)에서도 PG606 메타데이터 프로젝트 그룹을 통해서, 학술정보와 관련한 메타데이터 및 식별자 등에 관한 표준을 제정하고 있다. 2014년에 방송콘텐츠 유통의 효율성 및 메타데이터 상호운용성 확보를 위해 ‘방송 콘텐츠 유통 메타데이터 구성 요소 및 형식’ 기술 규격을 제정하였으며, 방송콘텐츠 유통 워크플로우에서 각 시스템이 공유해야 하는 메타데이터 공통요소와 이를 정의하는 방법을 규정한 바가 있다. 또한 2015년에는 TV-Anytime 국제표준을 기반으로 국내 맞춤형 방송 서비스를 위한 기술적 정보를 제공하는 것을 목적으로 한 메타데이터 서비스 표준인 ‘맞춤형 방송 서비스’ 기술 규격을 개정하였다.

2008년 ‘KBS 방송 메타데이터 표준’ 체계를 제정해, 방송 콘텐츠 자체의 메타데이터뿐만 아니라, 편성, 제작, 편집, 저장, 송출 및 관리 등 방송사 내부의 독립 시스템에 사용되는 각종 메타데이터를 총정리하여, KBS 내 각종 시스템에 연동 및 사용이 가능한 체계를 구축하였다고 한다. MBC의 경우도 프로그램별로 ID 체계가 다르기도 하고, 이기종 시스템 간에 콘텐츠를 대표하는 ID 체계를 여러 형태로 병행하여 사용하다 보니, 효율적으로 콘텐츠 및 메타데이터를 관리하기 어려웠다. 그래서 실시간 송출정보를 기준으로 통일된 프로그램 ID 체계를 만들어 편집 및 추후 유통에도 연계할 수 있는 메타데이터 체계를 정리한 바가 있다.

정리하머

앞서 언급한 내용을 종합해 보면, 국내외를 통틀어 표준화 주체가 누구인가에 따라 관심을 갖는 영역이 다르기 때문에 서로 상이한 메타데이터 형식, 요소 및 구조를 이용하고 있다. 공통된 내용이 분명히 있음에도 불구하고 서로 다른 명칭들로 인해 정보의 원활한 공유, 교환 및 검색이 이루어지지 못하고 있다. 그렇다고 모든 내용을 통합하여 또 다른 표준화 내용을 추진한다고 하더라도, 사용성이 떨어지는 또 하나의 표준이 될 가능성이 크다.

따라서, 미디어 콘텐츠를 다루는 데에 있어 필수적인 메타데이터 내용을 공통 규격으로 권고하고 선택적으로 사용할 수도 있는 내용 등에는 제한을 두지 않아 메타데이터 체계를 늘리는 데에 자유도를 부여해도 효과적일 수 있다. 또한 각 표준 간 서로 다른 명칭이지만 동일한 내용을 다루는 메타데이터의 스키마를 정리하여 각 정보 값을 변환 또는 번역할 수 있는 인터페이스 기술을 만들어 공유하는 것도 효과적일 수 있다. 