

석사학위논문

공공기관 유튜브 동영상 아카이빙을 위한
메타데이터 핵심 요소 연구

2022년

한 성 대 학 교 대 학 원

문 헌 정 보 학 과

기 록 관 리 학 전 공

김 락 근

석사학위논문
지도교수 박진호

공공기관 유튜브 동영상 아카이빙을 위한
메타데이터 핵심 요소 연구

2022년 6월 일

한성대학교 대학원

문헌정보학과

기록관리학전공

김 락 근

석사학위논문
지도교수 박진호

공공기관 유튜브 동영상 아카이빙을 위한
메타데이터 핵심 요소 연구

위 논문을 기록관리학 석사학위 논문으로 제출함

2022년 6월 일

한 성 대 학 교 대 학 원

문 헌 정 보 학 과

기 록 관 리 학 전 공

김 락 근

김락근의 기록관리학 석사학위 논문을 인준함

2022년 6월 일

심사위원장 박 희 진 (인)

심 사 위 원 이 호 신 (인)

심 사 위 원 박 진 호 (인)

국 문 초 록

공공기관 유튜브 동영상 아카이빙을 위한 메타데이터 핵심 요소 연구

한 성 대 학 교 대 학 원
문 헌 정 보 학 과
기 록 관 리 학 전 공
김 락 근

유튜브는 영상자료 공유를 위한 콘텐츠 유통 플랫폼의 역할뿐만 아니라 다양한 정보를 입수하는 경로로 활용되고 있다. 국내의 경우 유튜브 이용시간은 다른 소셜미디어 유형(트위터, 페이스북 등)에 비해 월등히 높다. 또한 정보검색에 있어서도 유튜브는 상위권에 위치한다. 대다수의 공공기관은 유튜브를 기관 홍보의 도구로 사용하고 있다. 이처럼 공공기관에서 유튜브를 운영하는 것은 공적인 업무로 바라볼 수 있으며, 유튜브에 업로드된 동영상은 공적 업무 수행의 결과물로 볼 수 있다.

공공기관은 유튜브 동영상을 통해 국민과 소통할 수 있다. 양방향 소통이 가능하다는 점과 이 결과가 기록으로 남는다는 점에서 소셜미디어는 기록으로서 가치가 높다고 할 수 있다. 유튜브에 영상을 업로드하는 행위 자체와 더불어 영상 속에서 사용하는 자막, 유행어를 통해 당시 사회의 트렌드를 파악

할 수 있다. 또한 댓글 속에서 구축된 공론의 장은 사회, 문화적인 특징을 발견할 수 있는 중요한 수단이다. 영상기록으로서 유튜브 동영상은 같은 내용의 텍스트 기록보다 이용자들의 이해가능성을 높여줄 수 있다. 이런 점을 감안하면 공공기관의 유튜브 동영상은 하나의 디지털 공공기록물로서 관리 및 보존해야 한다는 필요성이 제기된다. 즉, 유튜브 동영상을 아카이빙하여 가치 있는 동영상을 보존하고 관리해야 한다. 이 경우 동영상 자체와 동영상을 설명해주는 메타데이터를 함께 아카이빙 해야 한다. 하지만 현재 유튜브 동영상 아카이빙을 위한 메타데이터는 찾기 어렵고, 관련 연구도 부족하다. 이에 본 연구는 공공기관의 유튜브 동영상을 아카이빙 하기 위한 메타데이터를 설계하고, 핵심 요소를 선정하는 것을 목적으로 한다.

이를 위해 국내 기록관리 메타데이터 표준인 NAK 8:2021(v2.2)과 국제 보존 메타데이터 표준인 ISAD(G), PREMIS, 방송, 영상 메타데이터 표준인 PBCore를 살펴보았으며, 유튜브에 영상을 등록하고 이용 시 생성되는 정보(유튜브 메타데이터)들을 조사하였다. 이후 선행연구를 통해 종합한 아카이빙 메타데이터 구조를 토대로 조사한 메타데이터 요소들을 매핑하였다. 그 결과 5개 메타데이터 영역과 10개 하위 영역으로 구성된 유튜브 동영상 아카이빙을 위한 메타데이터를 설계하였다. 5개 메타데이터 영역은 설명 메타데이터, 구조 메타데이터, 관리 메타데이터, 보존 메타데이터, 이용자 참여 메타데이터로 구분된다. 하위 영역은 설명 메타데이터의 설명 영역, 구조 메타데이터의 구조 영역, 관리 메타데이터의 권리, 관리, 기술 영역, 보존 메타데이터의 출처, 참조, 문맥, 인증 영역, 이용자 참여 메타데이터의 이용자 참여 영역으로 세분화하였다.

설계한 메타데이터에 대해 평가를 받기위해 디지털 아카이빙 또는 메타데이터 업무 관련 경험 및 연구 경험이 있는 15명의 전문가를 통해 설문조사를 실시하였다. 설문지는 일반현황과 전체 메타데이터 요소 63개에 대해 5점 리커트 평가 방법, 추가 요소 및 기타의견으로 구성되었다. 설문지는 15부를 배포하여 14부(93%)를 회수하였다. 또한 데이터 정제 작업을 실시하여 설문지 문항 ‘1.1 소속기관 유형’을 공공기관, 기타기관, 대학별로 정제하였다.

본 연구는 기술 통계 분석을 수행하였다. 빈도분석과 평균분석으로 구분하

였는데, 빈도분석을 통해 응답자의 소속기관 유형, 최종학력, 경력 사항을 분석하였고, 평균분석을 통해 5점 리커트 평가 결과에 대한 데이터의 최솟값, 중위값, 최댓값, 평균값, 표준편차에 대해 분석하였다. 기술 통계 분석에 대한 결과는 다음과 같다. 첫째, 빈도 분석을 실시한 결과는 다음과 같다. 응답자 소속기관은 공공기관 9명(64%), 기타기관 4명(29%), 대학 1명(7%)으로 나타났다. 최종학력에 대한 응답은 박사(수료 포함) 6명(42.86%), 석사 4명(28.57%), 학사 4명(28.57%)으로 나타났다. 경력 사항 응답에 대해서는 한 명이 미응답하여, 13명으로 분석한 결과 0~5년이 3명(23%), 6~10년이 4명(31%), 11~15년이 2명(15%), 그 이상이 4명(31%)으로 나타났다.

둘째, 평균값을 중심으로 한 기술 통계 분석의 결과는 다음과 같다. 설명 메타데이터의 설명 영역 요소 중 ‘제목’은 전체 메타데이터 요소에서 모든 응답자들에게 5점(매우 적절)을 받은 유일한 요소로 나타났다. 이에 따라 ‘제목’ 요소는 더 이상의 분석 없이 핵심 요소로써 유의미한 결과가 나타났다. 구조 메타데이터의 구조 영역의 응답 결과에서 ‘재생시간’ 요소의 평균값이 제일 높았으며 ‘URL’, ‘프레임 비율’, ‘비디오 ID’ 순으로 나타났다. 평균값이 제일 낮은 요소는 ‘장르’로 나타났다. 관리 메타데이터의 권리 영역, 관리 영역, 기술 영역의 응답을 종합한 결과는 ‘공개 구분’ 요소의 평균값이 세 영역에 속한 요소들 중 제일 높았으며 다음으로 ‘생산일시’, ‘등록일시’, ‘파일크기’, ‘포맷명’, ‘공공저작물여부’, ‘공공저작물자유이용허락표시유형’ 순으로 나타났다. ‘관리 행위자’ 요소는 전체 메타데이터 요소에서 평균값이 가장 낮은 요소로 나타났다. 보존 메타데이터의 출처 영역, 참조 영역, 문맥 영역, 인증 영역의 응답을 종합한 결과는 ‘소장처’ 요소의 평균값이 네 영역에 속한 요소들 중 제일 높았으며 다음으로 ‘원본의 존재와 위치’, ‘기관명’, ‘소장 위치’ 순으로 나타났다. 이용자 참여 메타데이터의 경우 평균값이 높은 요소는 ‘조회수’, ‘좋아요수’ 순으로 나타났으며, ‘대댓글 ID’ 요소의 평균값이 가장 낮은 것으로 나타났다.

셋째, ‘추가 요소 및 기타 의견’은 7명이 응답하였다. 추가 요소 응답에서는 “채널 이름”, “구독자 수”, “콘텐츠 활용 연령”, “삭제” 요소라고 응답하였으며, 기타 의견 응답에서는 “보존 메타데이터의 추가 요소 필요”, “상, 하위

요소의 불명확성과 필수/선택 조사 여부”, “아카이빙 기준 제시”라고 응답하였다.

다음으로 타당도 검증과 신뢰도 검증을 진행하였다. 타당도 검증은 요인분석 방법으로 이루어졌고, 신뢰도 검증은 크론바흐 알파값을 활용하였다. 핵심 요소 선정방법은 타당도 검증과 신뢰도 검증의 기준값을 모두 충족하는 요소로 선정하였다. 타당도 검증과 신뢰도 검증 결과는 다음과 같다.

첫째, 해당 설문 응답에 대한 5개의 메타데이터의 10개 하위 영역별로 타당성 검증과 신뢰도 검증을 진행하였다. 각 하위 영역별로 요인분석이 적합한지 알아보려고 KMO 및 Bartlett의 구형성 검증을 실시한 결과, $KMO > 0.5$, Bartlett의 $p < 0.05$ 를 만족하지 못한 영역은 관리 메타데이터의 권리 영역이 나타났다. 보존 메타데이터의 출처 영역은 요소가 한 개(‘수집, 기관의 직접적 출처’)만 존재하기 때문에 요인분석을 할 수 없었으며, 참조 영역은 요인분석 결과 추출되는 값이 존재하지 않았다.

둘째, 요인분석 결과는 다음과 같다. 설명 메타데이터의 설명 영역에서는 ‘제목’ 요소를 제외하고 나머지 요소들만 요인분석을 실시한 결과 ‘언어’, ‘설명’, ‘범주’가 추출 공통성 값과 성분 수치의 기준값을 충족하였다. 구조 메타데이터의 구조 영역에서는 ‘프레임 비율’, ‘데이터 전송 속도’, ‘재생시간’, ‘압축’, ‘인코딩 정보’만이 공통성 값과 성분 수치가 기준값을 충족하였다. 관리 메타데이터의 관리 영역에서 추출 기준값을 충족한 요소는 ‘관리일시’, ‘관리행위자’, ‘변경요소’, ‘생산일시’, ‘관리설명’, ‘관리유형’으로 나타났다. 기술 영역에서는 ‘생성 어플리케이션 명’, ‘생성 어플리케이션 버전’, ‘운영시스템’, ‘포맷버전’으로 나타났다. 모든 요소가 추출 기준값을 충족한 영역은 보존메타데이터의 문맥 영역(‘관계대상식별자’, ‘관계유형’, ‘관계설명’)과 인증 영역(‘무결성체크법’, ‘서명’, ‘무결성체크값’), 이용자 참여 메타데이터의 이용자 참여 영역(‘조회수’, ‘좋아요수’, ‘싫어요수’, ‘댓글수’, ‘댓글ID’, ‘댓글내용’, ‘대댓글ID’, ‘대댓글내용’)으로 나타났다.

셋째, 타당성 검증 기준값을 충족한 요소들에 한해서 메타데이터별, 영역별로 신뢰도 검증을 실시하였다. 그 결과 모두 0.7 이상으로 확인되어 매우 바람직한 것으로 나타났으며, 해당 요소들은 핵심 요소로써 유의한 결과가 나

타났다고 할 수 있다.

본 연구에서 설계한 메타데이터는 5개 메타데이터의 10개 영역에 포함되는 63개의 측정요소로 구성되었다. 이 중에서 타당성 검증과 신뢰도 분석을 통해 핵심 요소로써 유의한 결과가 나타난 요소는 5개 메타데이터의 7개 영역에 포함되는 33개 요소로 나타났다. 이에 따른 메타데이터별, 영역별, 측정요소별 채택률은 각각 100%, 70%, 53%로 나타났다.

본 연구에서 메타데이터 요소 도출을 위해 연구 대상으로 선정한 유튜브 동영상은 소셜미디어 유형 전체의 동영상을 대표한다고 보기 어려울 수 있다. 또한 핵심 요소 도출을 위해 실시한 요인분석의 결과를 가지고 핵심 요소의 여부를 판단하는 것은 어렵다는 한계점이 존재한다. 하지만 본 연구는 다양한 메타데이터 연구 중 소셜미디어 동영상을 아카이빙하기 위한 메타데이터의 초기 연구로서 의의가 있으며, 유튜브 동영상을 기록으로 접근했다는 점은 다른 메타데이터 연구와 차별화를 가진다고 볼 수 있다. 향후 다양한 후속 연구를 통해 소셜미디어 메타데이터 대한 요소가 다양화된다면 대규모 집단을 대상으로 한 확인적 요인분석 등 방법이 가능하리라 본다.

기록관리의 궁극적인 목적은 기록의 활용이라고 할 수 있다. 유튜브 기록 역시 아카이빙을 통해 이용자에게 서비스를 제공하고, 수집된 데이터를 통해 유의미한 현상을 파악함으로써 해당 기록의 가치를 높일 수 있다. 본 연구는 유튜브 동영상 기록물의 아카이빙에서 메타데이터에 주목했으며, 후속 연구를 통해 다양한 연구가 진행되기를 바란다.

【주요어】 공공기관, 소셜미디어 기록, 유튜브 기록, 아카이빙, 메타데이터, 핵심 요소

목 차

I. 서 론	1
1.1 연구의 필요성과 목적	1
1.2 연구방법	3
II. 이론적 배경과 선행연구	5
2.1 이론적 배경	5
2.1.1 전자기록물 보존	5
2.1.1.1 전자기록물 개념	5
2.1.1.2 국내 전자기록물 보존을 위한 표준 현황	7
2.1.2 유튜브 기록	10
2.1.2.1 유튜브 기록 개념	10
2.1.2.2 공공기관 유튜브 현황	12
2.1.2.3 유튜브 기록 특징	13
2.1.3 공공기관 유튜브 동영상 아카이빙 필요성	17
2.1.4 메타데이터	19
2.1.4.1 메타데이터 일반	19
2.1.4.2 아카이빙을 위한 메타데이터	20
2.2 선행연구	22

Ⅲ. 메타데이터 요소 설계	26
3.1 메타데이터 요소 조사	27
3.1.1 국내 기록관리 메타데이터 표준 요소	27
3.1.2 국외 메타데이터 요소	31
3.1.3 유튜브 메타데이터 요소	46
3.2 메타데이터 요소 매핑	58
3.2.1 설명 메타데이터	59
3.2.2 구조 메타데이터	60
3.2.3 관리 메타데이터	62
3.2.4 보존 메타데이터	66
3.2.5 이용자 참여 메타데이터	69
3.3 메타데이터 요소 도출	70
Ⅳ. 메타데이터 요소 평가	79
4.1 설문조사 개요	82
4.1.1 전문가 선정	83
4.2 설문조사 데이터 취합 및 정제	84
4.3 설문조사 결과	85
4.3.1 응답자 분석	86
4.3.2 기술 통계 분석	87
4.3.2.1 설명 메타데이터 기술 통계	87
4.3.2.2 구조 메타데이터 기술 통계	87

4.3.2.3	관리 메타데이터 기술 통계	88
4.3.2.4	보존 메타데이터 기술 통계	91
4.3.2.5	이용자 참여 메타데이터 기술 통계	94
4.3.2.6	추가 요소 및 기타 의견	95
4.3.3	타당도 검증	95
4.3.3.1	설명 메타데이터 타당도 검증	96
4.3.3.2	구조 메타데이터 타당도 검증	97
4.3.3.3	관리 메타데이터 타당도 검증	98
4.3.3.4	보존 메타데이터 타당도 검증	100
4.3.3.5	이용자 참여 메타데이터 타당도 검증	102
4.3.4	신뢰도 검증	103
4.4	최종 메타데이터 요소 도출	104
V. 결 론		107
참 고 문 헌		111
부 록 1		118
부 록 2		127
ABSTRACT		130

표 목 차

[표 1] 연구 절차	3
[표 2] 소셜미디어 유형	10
[표 3] 공공기관 유튜브 채널 개설 현황	13
[표 4] 국가기록원 기록관리 메타데이터 표준 요소	28
[표 5] PREMIS 의미 단위와 의미 구성 요소	33
[표 6] ISAD(G) 영역과 기술요소	42
[표 7] PBcore 메타데이터 표준 요소	44
[표 8] 유튜브 메타데이터 요소	55
[표 9] 아카이빙 메타데이터 내용 및 예시	58
[표 10] 설명 메타데이터 요소 비교	59
[표 11] 구조 메타데이터 요소 비교	61
[표 12] 관리 메타데이터 요소 비교	63
[표 13] 보존 메타데이터 요소 비교	67
[표 14] 이용자 참여 메타데이터 요소	69
[표 15] 유튜브 동영상 아카이빙 메타데이터 요소 및 설명	70
[표 16] 메타데이터 요소 평가 상세 절차	79
[표 17] 요인분석 및 신뢰도 검증 방법 및 추출 기준값	81
[표 18] 통계 분석 도구	81
[표 19] 설문지 구성	82
[표 20] 연구자별 메타데이터 전문가 선정방법	83
[표 21] 메타데이터 요소 평가 설문지 결과 데이터 취합 정제 내용	85
[표 22] 소속기관 유형 응답 결과	86
[표 23] 최종 학력 응답 결과	86
[표 24] 경력 사항 응답 결과	86
[표 25] 설명 메타데이터 - 설명 영역 응답 기술 통계표	87
[표 26] 구조 메타데이터 - 구조 영역 응답 기술 통계표	88
[표 27] 관리 메타데이터 - 권리 영역 응답 기술 통계표	89

[표 28] 관리 메타데이터 - 관리 영역 응답 기술 통계표	90
[표 29] 관리 메타데이터 - 기술 영역 응답 기술 통계표	90
[표 30] 보존 메타데이터 - 출처 영역 응답 기술 통계표	91
[표 31] 보존 메타데이터 - 참조 영역 응답 기술 통계표	92
[표 32] 보존 메타데이터 - 문맥 영역 응답 기술 통계표	93
[표 33] 보존 메타데이터 - 인증 영역 응답 기술 통계표	94
[표 34] 이용자 참여 메타데이터 - 이용자 참여 영역 응답 기술 통계표	94
[표 35] 설명 메타데이터 - 설명 영역 타당도 검증 결과	96
[표 36] 구조 메타데이터 - 구조 영역 타당도 검증 결과	97
[표 37] 관리 메타데이터 - 권리 영역 타당도 검증 결과	98
[표 38] 관리 메타데이터 - 관리 영역 타당도 검증 결과	99
[표 39] 관리 메타데이터 - 기술 영역 타당도 검증 결과	99
[표 40] 보존 메타데이터 - 문맥 영역 타당도 검증 결과	101
[표 41] 보존 메타데이터 - 인증 영역 타당도 검증 결과	101
[표 42] 이용자 참여 메타데이터 - 이용자 참여 영역 타당도 검증 결과	102
[표 43] 신뢰도 검증 결과	103
[표 44] 유튜브 동영상 아카이빙을 위한 메타데이터 핵심 요소	104
[표 45] 기존 메타데이터 요소와 핵심 요소 결과 비교	105

그림 목 차

[그림 1] 기록물건 장기보존 메타데이터 구성	9
[그림 2] PREMIS 3.0 모델	32
[그림 3] 유튜브 메타데이터 - 동영상 세부정보 1	47
[그림 4] 유튜브 메타데이터 - 동영상 세부정보 2	48
[그림 5] 유튜브 메타데이터 - 동영상 세부정보 3	49
[그림 6] 유튜브 메타데이터 - 동영상 세부정보 4	50
[그림 7] 유튜브 메타데이터 - 동영상 요소 1	51
[그림 8] 유튜브 메타데이터 - 동영상 요소 2	52
[그림 9] 유튜브 메타데이터 - 동영상 요소 3	52
[그림 10] 유튜브 메타데이터 - 공개 상태	53
[그림 11] 유튜브 메타데이터 - 자동 추출 정보 및 이용자 피드백 정보 ..	54

I. 서론

1.1 연구의 필요성과 목적

2006년 구글이 인수한 유튜브(YouTube)는 동영상을 공유하기 위한 소셜미디어 플랫폼이다. 유튜브는 현재 전 세계 최대 규모의 동영상 공유와 호스팅 서비스로 발전했으며, 개인뿐만 아니라 다양한 기관들이 영상정보 제공자로 참여하고 있다. 유튜버라는 새로운 직업군을 만들어졌다는 점에서 그 규모와 파급력을 짐작해볼 수 있다.

전통적인 미디어에 속하는 신문, 라디오, TV의 경우는 전문가에 의한 정보 생산과 전달 및 일방향적인 정보전달의 특징이 강하다. 반면 유튜브의 경우는 일반 대중 모두가 콘텐츠 생산자가 될 수 있으며, 비동시성과 상호작용이 특징인 새로운 미디어로 볼 수 있다. 유튜브의 영향력이 커지면서 전통적인 미디어의 경우도 유튜브를 활용한 정보 유통에 참여하고 있다. 인터넷 이용 조사업체인 나스미디어(2021)에 따르면 전체 이용자 중 91.8%가 유튜브를 통해 온라인 동영상을 시청하고 있다. 또 다른 인터넷 조사업체인 와이즈앱에서 2021년에 조사한 결과에 따르면 카카오톡, 네이버 등 여러 플랫폼 중 유튜브가 총 701억 분으로 한국인이 가장 오랜 시간 사용하는 앱으로 확인되었다(MoneyS, 2021). 우리나라 정부 및 공공기관 역시 높은 이용률을 자랑하는 유튜브를 통해 기관을 홍보하고, 정보를 전달하는데 힘을 쓰고 있다. 2022년 3월 15일 기준, 공공기관 경영정보공개시스템인 알리오(Alio) 사이트에서 확인한 결과, 기획재정부에서 지정한 350개 공공기관 중 324개(92%) 기관에서 유튜브 채널을 운영하고 있음을 확인하였다. 구독자가 가장 많은 공공기관 유튜브 채널은 인천국제공항공사가 194,000명으로 나타났고, 업로드한 영상의 수가 가장 많은 유튜브 채널은 재단법인 국악방송이 3,657개로 나타났다.

유튜브는 단순히 영상자료를 공유하고 즐기기 위한 콘텐츠 유통 플랫폼임과 동시에 다양한 정보를 입수하는 주된 경로로 활용되고 있다. 영상자료의 공유와 정보검색 모두에 있어서 중요한 점은 이용자가 원하는 정보자원을 잘

찾을 수 있어야 한다는 점이다. 이를 위해서는 해당 정보자원이 갖는 속성들이 체계적으로 정리되어 있어야한다. 정보자원의 형태에 따라 다양한 접근법이 존재하지만, 유튜브의 경우는 동영상이라는 정보자원의 특징을 잘 반영할 수 있는 속성들이 필요하다. 물론 유튜브에 영상을 제작하여 발행하기 위해서는 사전에 설정된 영상 관련 정보들을 입력해야한다. 또 유튜브는 자체 추천 알고리즘을 통해서 이용자가 이전에 열람한 영상과 유사한 자료들을 추천해주고 있다. 구글이 추천 알고리즘에 대해 구체적으로 밝힌 바는 없지만 주로 이용자들이 해당 영상을 이용하면서 남긴 흔적(조회수, 시청시간, 반응 등)과 더불어 제작자가 입력한 키워드와 생중계 콘텐츠의 비중 등에 비중을 두고 있는 것으로 알려져 있다(전자신문, 2020). 즉, 이용자가 적합한 영상자원을 탐색하기 위해서는 제작자, 발행자가 입력하는 키워드를 포함한 메타데이터 정보와 더불어 영상제작 후 생성되는 로그 정보가 핵심임을 알 수 있다. 또한 생중계 즉, 실시간 정보자원이 존재한다는 점도 유튜브의 또 다른 특징임을 알 수 있다.

공공기관이 유튜브를 운영하고 콘텐츠를 제작하는 비용은 공적자금에 해당한다. 이는 개인이나 기업이 제작하는 자료보다 공적 기능이 강하다고 볼 수 있으며, 해당 콘텐츠는 하나의 기록으로 보존될 필요가 있다. 공공기관이 제작하는 영상자료 자체는 목적이 유튜브 발행용일 수도 있고, 기존에 제작된 영상을 재편집하거나 변환하여 발행할 수도 있다. 어떤 목적이든 공적 영역에 속하는 저작물이라는 점에서 보존에 대한 고민이 필요하다. 물론 공공기관이 제작하는 영상자료가 유튜브에 발행되기 전 원본이 존재한다는 점을 감안한다면 보존대상을 원본으로 생각할 수도 있다. 그러나 유튜브 저작물이 갖는 특징과 중요성 중에는 이용자 참여, 스트리밍 기반의 영상자료라는 점을 고려해야한다. 즉 원본이 존재하지만 유튜브에서 생성되는 자료의 경우 새로운 부가정보를 포함하고 있다는 점과 실시간 영상자료의 경우 원본이 유튜브 플랫폼이 될 수도 있다는 점을 감안할 필요가 있다. 유튜브에 등록된 영상자료의 경우 원본과는 다른 가치를 가질 수 있음에 주목한 것이다. 유튜브에 정보자원을 등록하고 관리하는 기관의 경우도 내부에 콘텐츠 관리시스템이 존재하고 아카이브가 존재하는 경우가 아니라면 등록 이후의 관리가 이루어지고 있

다고 단정하기도 어렵다. 박성미(2002)도 이와 관련하여 영상물의 원본을 기록으로 인정하는 1차 영상자료와 달리 2차 영상자료는 영상의 제작 및 편집 과정에서 영상물에 반영된 당시의 사회적 요구와 상황 즉, 제작 당시를 읽어 낼 수 있는 자료로서 가치가 있다고 하였다.

본 연구에서는 이런 점을 고려하여 유튜브 영상자료를 하나의 디지털 기록물로 보고 이를 아카이빙하기 위한 표준 메타데이터를 설계하였다. 특히 영상자료 중 공공기관이 운영하는 자료를 대상으로 하였다. 유튜브 동영상은 소셜미디어 기록의 성격과 디지털 동영상 기록의 성격 모두 가지고 있다. 따라서 유튜브 동영상을 아카이빙하기 위해서는 소셜미디어 기록과 동영상 기록의 특성을 포괄할 수 있는 메타데이터 요소가 필요하다. 본 연구는 소셜미디어 동영상 기록인 유튜브 동영상이 갖는 특성을 파악하여 이용자들의 지속적인 접근과 활용이 가능한 메타데이터를 설계하는 데 있다. 본 연구의 목적을 정리하면 공공기관 유튜브 동영상을 아카이빙하여 보존할 가치 있는 동영상들을 원활하게 이용할 수 있도록 메타데이터 핵심 요소를 개발하는 것이다. 이를 위해 연구 과정에서 국내외의 표준 메타데이터 요소들과 유튜브에 콘텐츠 발행과 이용 시 필요한 메타데이터 요소를 함께 검토하였다.

1.2 연구방법

본 연구는 메타데이터 요소 설계와 요소 평가의 두 단계로 수행하였다. 메타데이터 요소 설계와 메타데이터 요소 평가에 대한 상세 수행내용은 [표 1]과 같이 정리된다.

[표 1] 연구 절차

연구 절차	세부 절차	수행 내용
메타데이터 요소 설계	요소 조사	• 국내 기록관리 표준 조사: NAK 8:2021 (v2.2)
		• 국외 보존기록 표준 조사: ISAD(G), PREMIS
		• 영상 표준 조사: PBCore
		• 유튜브 표준 조사: 콘텐츠 등록 및 이용 시 생성되는 정보

	요소 매핑 및 요소 도출	<ul style="list-style-type: none"> • 아카이빙 메타데이터 구조에 조사한 표준 매핑 • 유튜브 동영상 아카이빙을 위한 메타데이터 요소 설계
메타데이터 요소 평가	요소 평가 및 분석	<ul style="list-style-type: none"> • 설문 조사 • 데이터 정제 • 기술 통계 분석 • 타당도 검증 • 신뢰도 검증
	핵심 요소 선정	<ul style="list-style-type: none"> • 설문결과 분석을 통해 유의한 결과가 나타난 요소 선정

Ⅱ. 이론적 배경과 선행연구

2.1 이론적 배경

2.1.1 전자기록물 보존

2.1.1.1 전자기록물 개념

2022년 7월 12일부터 시행되는 공공기록물 관리에 관한 법률(이하: 공공기록물법) 19조 2항에 따르면 전자기록물이란 정보처리능력을 가진 장치에 의하여 전자적인 형태로 작성하여 송신, 수신 또는 저장되는 기록정보 자료라고 명시되어 있다. 기록학 용어사전(2008)에 따르면 전자기록물은 전산 시스템을 통해 처리되는 기록물로서 아날로그 기록물을 디지털화한 기록물과 전자 형태로 생산된 기록물(born digital) 모두 전자기록물에 포함된다고 하였다. 국가기록원(2005)은 기록관리 측면에서의 전자기록물은 업무 수행을 목적으로 기관에서 컴퓨터 시스템을 통해 전자포맷으로 생성, 유통, 관리하는 모든 기록물이라고 하였다. 이를 통해 업무 수행의 증거적인 역할을 하고 문서간의 관계가 기술된 기록물이라고 하였다.

이상의 정의를 종합하면 전자기록물이란 기록물의 한 유형으로 기관의 전산 시스템을 통해 생성되고, 유통되는 모든 기록물이라고 할 수 있다.

국가 표준인 KS X ISO15489-1:2016(Information and documentation — Records management — Part 1: Concepts and principles)에서는 기록이 업무 수행에 대한 공신력 있는 증거로서 역할을 하고, 업무 과정에서 요구사항을 만족시키기 위해 그 형태나 구조에 상관없이 진본성, 신뢰성, 무결성, 이용가능성을 갖추고 있어야 한다고 명시되어 있다. 각 속성은 아래와 같이 정리하였다.

첫째, 진본성(Authenticity)이다. 기록학 용어사전(2008)에서는 기록이 위조되지 않은 생산 당시의 상태임을 지칭하는 것을 진본성이라고 하였다. 진본성을 보장하기 위해서는 기록의 생산, 수령, 전달 등 기록이 유통되는 전 과정을 통제하는 정책이나 절차 등이 문서화되어야 한다(국가기록원, 2005). 이를 통해 기록이 나타내고자 하는 그대로의 기록인지, 기록을 생산 또는 송신한 행위자가 바로 그 행위자가 맞는지, 기록에 명시된 시점에 생산, 송신되었는지를 입증할 수 있다.

둘째, 신뢰성(Reliability)이다. 신뢰성이 있는 기록은 기록이 입증하고자 하는 처리행위, 활동 또는 사실의 완전하고 정확한 표현물로서 해당 기록의 내용을 신뢰할 수 있고, 이후의 처리행위나 활동 과정에서 의존할 수 있는 것이다. 설문원(2005)은 업무 활동의 결과로 생산된 기록물이 그 활동을 얼마나 정확하고 충분하게 보여주는지를 나타내는 것이라고 하였다. 이로 인해 기록은 관련 사건이 일어난 시점이나 사건이 일어난 직후에 생산되어야 한다.

셋째, 무결성(Integrity)이다. 무결성을 가진 기록이란 완전하며, 변경되지 않은 기록이다. 즉, 업무 수행과정에서 기록이 허가받지 않은 절차 또는 요인에 의해서든 변형이 되지 않았다는 것을 의미한다. 기록의 무결성을 보장하기 위해서는 기록이 생산된 이후 새롭게 내용을 추가해도 되는지, 추가하기 위한 조건은 무엇이며, 그 권한은 누구한테 있는지를 명확하게 기록하여 추적할 수 있도록 해야한다.

넷째, 이용가능성(Usability)이다. 이용가능성이 있는 기록은 이해관계자들이 원하는 시간 안에 해당 기록의 위치를 발견할 수 있고, 검색할 수 있으며, 볼 수 있고, 이를 해석할 수 있는 것이다. 이를 위해 기록을 생산한 업무 과정이나 처리행위와 연계가 되어야하며, 이러한 활동의 과정에서 생산된 기록 사이의 연계도 유지되어야 한다. 이용가능성이 보장되지 않는다면 기록의 진본성, 신뢰성, 무결성을 확인할 수 없다. 또한 기록을 ‘보존’하는 것은 결국 기록이 완전무결한 상태로 관리할 수 있음과 동시에 해석할 수 있는 상태로 유지하는 것이기 때문에 이용가능성을 보장하는 행위라고 할 수 있다(ICA, 1996; 설문원, 2005에서 재인용).

이처럼 기록물이 갖추어야 할 4가지 속성은 국내 공공기록물관리에 관한 법률 시행령(이하: 공공기록물법 시행령) 46조 2항에서도 나타난다. 영구기록물관리기관의 전자기록물 보존 및 관리에 있어서 전자기록물의 진본성, 무결성, 신뢰성, 이용가능성이 보장되도록 관리정보 메타데이터와 행정전자서명 등에 대한 검증을 실시해야하고, 장기보존포맷을 주기적으로 변환해야 한다고 명시되어 있다.

2.1.1.2 국내 전자기록물 보존을 위한 표준 현황

공공기록물법 시행령 36조 1항에 따르면 “기록관 또는 특수기록관의 장은 인수가 종료된 전자기록물 중 보존기간이 10년 이상인 전자기록물에 대해서는 제35조 제5항에 따른 인수 완료 결과를 통보한 후 1년 이내에 중앙기록물관리기관의 장이 정하는 바에 따라 문서보존포맷 및 장기보존포맷으로 변환하여 관리하여야 한다.”라고 명시되어 있다. 이 중 국내의 기록물 보존과 관련된 표준은 각각 국가기록원의 NAK 30:2008(v1.0)(이하: NAK 30)과 NAK 31-1:2020(v2.2)(이하: NAK 31-1), NAK 31-2:2020(v1.0)(이하: NAK 31-2)에서 살펴볼 수 있다.

NAK 30은 전자기록물 문서보존포맷 기술 규격이다. NAK 30에서 정의한 문서보존포맷이란 문서가 생산된 당시의 애플리케이션이 없어도 해당문서의 내용과 외형을 그대로 재현하여 내용보기를 가능하게 하는 포맷이다. 국내 문서보존포맷의 규격은 PDF/A-1(ISO 19005-1:2005 Part 1: Use of PDF 1)으로 규정하고 있다.

PDF/A-1에서 ‘A’는 Archive를 의미하고 ‘1’은 인쇄가 가능한 모든 매체를 대상으로 한다는 것을 의미한다. 따라서 동영상이나 비디오, 오디오와 같은 매체에서는 적용이 불가능하다. NAK 30의 기술규격 범위에서 제외되는 사항은 다음과 같다. 첫째, 종이 또는 전자문서를 변화하는 특정 방법, 둘째, 기술적 디자인 또는 사용자 인터페이스, 렌더링에 관한 구체적인 방법, 셋째, 저장 방법 및 저장 매체, 넷째, 필요한 컴퓨터의 하드웨어 또는 운영체제이다.

다음으로 NAK 31-1, 31-2는 각각 전자기록물 장기보존패키지 기술규격

-제1부 XML로 포맷화된 방식(NEO2), 전자기록물 장기보존패키지 기술규격
-제2부 디렉토리로 구조화된 방식(NEO3)이다.

앞서 살펴본 공공기록물법 시행령 36조 2항의 장기보존포맷은 ISO 14721 OASIS 참조모델의 정보패키지에 해당하는 개념이나 컴퓨터 파일 포맷으로 오해하는 소지를 방지하고 ‘장기보존패키지’라는 용어를 사용하였다. 즉, 공공기록물법 시행령 36조에 언급되어있는 장기보존포맷은 NAK 31-1, 31-2의 장기보존패키지에 해당한다. 장기보존패키지란 디지털객체에 대한 논리적, 물리적 캡슐화 규격을 규정함과 동시에 진본성, 무결성, 이용가능성을 유지하기 위한 디지털 기록물에 대한 장기보존패키지에 대한 기술규격을 기술하는 것으로 정의된다.

해당 표준들은 ISO 14721 OASIS 참조 모형을 참고하여 패키징 정보에 의해 실제적 또는 논리적으로 묶일 수 있다. 실제적 장기보존 패키지란 단일 디지털 객체로 생성하는 방식이며 NAK 31-1이 이에 해당한다. 논리적 장기보존 패키지는 여러 디지털 객체들로 나누어져 있는 것들을 서로 연결하는 방식이며, NAK 31-2가 이에 해당한다.

장기보존패키지의 필요성은 다음과 같다. 전자기록물의 특성상 변형, 훼손, 유실되기가 쉽기때문에 안전하게 보존할 수 있도록 하는 기술적 기반을 필요로 한다. 즉, 오랜 시간이 지나도 생산 당시에 가지고 있던 기록물의 내용을 그대로 재현하여 접근할 수 있도록 보존되어야 하며, 업무 활동의 증거와 업무에 대한 책임 소재를 분명하게 밝혀주는 법적 증거를 확보할 수 있도록 하여야 한다. 기록물이 갖고 있는 법적 증거는 기록물의 진본성과 무결성이 있어야만 확보될 수 있으므로 장기보존패키지의 적용이 필요하다

국내에서 현행되는 방식은 NAK 31-1이다. 즉, XML로 포맷화된 방식인 NEO2를 사용하고 있다. 이에 따라 NAK 31-1의 구조를 살펴보았다. NAK 31-1의 장기보존패키지는 원문, 보존포맷, 장기보존 메타데이터, 장기보존 메타데이터 스키마 정보, 진본확인 정보로 구성되어있다. 그중 보존포맷은 “전자기록물 생산 당시의 애플리케이션이 없이도 해당문서의 내용과 외형을 그대로 재현한 포맷 또는 원본의 디지털객체 특성정보(내용과 문맥정보, 유형별 특성정보 등)를 원본과 동일한 수준으로 보존 가능한 포맷으로 시간과 기술변

화에 상관없이 이용자가 기록물 내용에 접근할 수 있게 한다”라고 명시되어 있다. 장기보존 메타데이터는 기록물 원문과 함께 기록물의 생산부터 보존까지 전 과정을 기술한 정보이다. 또한 국내 기록관리 메타데이터 표준인 NAK 8: 2021(v2.1)(이하: NAK 8)을 반영하여 국내 기록관리에서 호환성을 유지하였다. 장기보존 메타데이터는 기록계층에 따라 기록철 장기보존 메타데이터, 기록물 건 장기보존 메타데이터로 나뉘고 장기보존 패키지가 수정되는 경우에는 수정된 장기보존 메타데이터로 나뉜다. 기록물 철과 건 장기보존 메타데이터 중 유튜브 동영상 기록물을 적용시키기 위해 먼저 소셜미디어 기록의 계층을 알아보았다. 박유경(2014)은 공공기관 소셜미디어 기록의 경우 공공기관에서 등록한 콘텐츠가 대부분이기 때문에 단위과제 범위 안에서 관련된 기록물을 편철한 기록물철의 단위가 있기 어렵다고 하였다. 따라서 본 연구에서도 이를 적용하여 계정(유튜브 채널)을 하나의 기록물 철로 보았고, 계정에 등록된 콘텐츠를 건으로 보았다. 기록물 건 장기보존 메타데이터는 다음 [그림 1]와 같다.



[그림 1] 기록물건 장기보존 메타데이터 구성

2.1.2 유튜브 기록

2.1.2.1 유튜브 기록 개념

공공기록물 관리에 관한 법률 시행령 2조 2항에는 “전자기록물은 정보처리능력을 가진 장치에 의하여 전자적인 형태로 작성하여 송신, 수신 또는 저장되는 전자문서, 웹 기록물 및 행정정보 데이터세트 등의 기록정보자료를 말한다”라고 명시되어 있다. 그중 웹 기록물은 공공기관 기록물관리 지침(2021)에 따르면 “공공기관에서 운영, 활용하는 웹사이트, 블로그, 소셜네트워크서비스(SNS) 등 웹을 기반으로 생산된 기록정보자료와 웹사이트 운영 및 구축과 관련된 관리 정보”라고 명시되어 있다. 여기서 웹 기록물에 포함되는 블로그와 SNS는 소셜미디어에 속한다.

소셜미디어는 제한적인 시스템 내에서 특정 개인이 대중에게 자신을 소개하고, 다른 사람과 정보를 교환하며 다대다 관계를 형성 및 유지함으로써 다른 이용자들이 형성하고 있는 관계망을 보거나 자신과 연결할 수 있는 웹 기반 서비스이다(Boyd & Ellison, 2007). 다음 [표 2]는 최민재, 양승찬, 김강석(2009)이 제시한 소셜미디어 유형을 표로 재구성한 것이다.

[표 2] 소셜미디어 유형

구분	서비스 유형	대표 예시	정의
커뮤니케이션 모델	블로그 (Blog)	네이버 블로그, 티스토리	- 네트워크 형성이 주안점 - 콘텐츠 업로드의 자율성 - 미디어의 자유로운 업/다운로드 - 손쉬운 검색
	마이크로 블로그 (MicroBlog)	트위터, 미투데이	
	소셜 네트워킹 (Social Networking)	페이스북, 네이버밴드	
	이벤트 네트워킹 (Event Networking)	Meetup, 이벤터스	

협업 모델	위키 (Wiki)	위키피디아, 나무위키	<ul style="list-style-type: none"> - 이용자 간의 협동 기반 - 유동성 - ‘협동’에 의한 콘텐츠 생산 - 높은 정보의 신뢰도 - 사회 경향 파악 가능
	소셜 북마킹 (Social Bookmarking)	네이버 북마크, 마가린	
	소셜 뉴스 (Social News)	Reddit, 위키투리	
	리뷰와 의견 사이트 (Review & Opinion Site)	디시인사이드, 뽀뿌	
	커뮤니티 Q&A (Community Q&A)	네이버 지식인	
콘텐츠 공유 모델	사진 공유 (Photo Sharing)	인스타그램, 플리커	<ul style="list-style-type: none"> - 네트워크보다는 “공유”에 집중 - 콘텐츠 유형에 따른 자 신들만의 전문 영역 구축 - 외부서비스와의 연동이 중요하기 때문에 활발한 타 사이트와의 공유
	비디오 공유 (Video Sharing)	유튜브, 카카오tv	
	라이브 방송 (Livecasting)	아프리카tv, 트위치	
	오디오와 음악 공유 (Audio & Music Sharing)	멜론, 지니	
엔터테인먼트 모델	가상 세계 (Virtual World)	SecondLife, The Sims Online	<ul style="list-style-type: none"> - 네트워크를 오락적인 서 비스를 기반으로 하여 제공 - 콘텐츠 자체를 즐기기 위해 네트워크성이 일정 부 분 가미
	게임 공유와 플레이 (Game Sharing & Play)	Kongregate	

출처: 최민재, 양승찬, 김강석. (2009). 인터넷 소셜미디어와 저널리즘. 서울: 한국언론재단. p.22를 기반으로 하여 재구성

[표 5]를 통해 알 수 있듯이 소셜미디어 유형에 유튜브가 속해있는 것을 알 수 있다. 이를 통해 공공기관에서 생산한 유튜브 동영상은 웹 기록물로서의 가치를 가지고 있다고 볼 수 있다.

2.1.2.2 공공기관 유튜브 현황

넓은 개념으로의 공공기관은 국가기관, 지방자치단체, 그 밖에 대통령령으로 정하는 기관을 말한다. 대통령령에서는 공공기관의 운영에 관한 법률과 공기업법 등 다양한 범위의 기관을 공공기관이라고 칭하고 있다(한국. 공공기록물법, 제3조 1항). 본 연구에서 조사하고자 하는 공공기관의 범위는 정부의 투자, 출자 또는 정부의 재정지원 등으로 설립, 운영되는 기관으로서 기획재정부 장관이 지정한 기관으로 한정하였다(한국. 공공기관운영법, 제4조 1항).

2022년 1월 기획재정부는 공기업 36개, 준정부기관 94개, 기타공공기관 210개로 총 350개 기관을 발표하였다(기획재정부, 2022). 이에 해당하는 공공기관은 공기업, 준정부기관, 기타공공기관으로 구분된다. 그중 공기업은 시장형과 준시장형으로 구분되고, 준정부기관은 기금관리형, 위탁집행형으로 구분된다.

우리나라는 2009년 청와대의 ‘뉴미디어 비서관’ 제도를 소셜미디어에 도입하는 것으로 처음 시작하였다. 그 후, 2011년 10월 국무회의를 통하여 온라인 대변인제의 자율적 도입이 이루어지면서 정부에서도 소셜미디어의 기능을 활용하여 국민과의 소통을 더욱 확대하고자 노력하였다(이귀영, 2015). 이러한 노력으로 인해 공공기관 소셜미디어 게시물이 큰 화제를 몰고 오는 경우도 있었는데 그중 유튜브에 등록된 동영상도 큰 화제를 몰고 온 사례는 다음과 같다. 첫째, 2020년 우리나라에서 뮤직비디오를 제외한 가장 많이 시청한 유튜브 영상으로는 질병 관리청의 “코로나 19 국민 행동수칙 꼭 기억해주세요”가 당시 1738만(현재 1740만) 조회수를 기록하였다. 둘째, 같은 해 한국관광공사의 유튜브 채널에 가수 이남치의 판소리 음악과 춤을 가미한 영상이 업로드되었다. 해당 영상은 2020년 12월 기준, 누적 조회수 3억 회를 기록하였다. 이를 통해 가수 이남치는 음악방송이나 음악 사이트가 아닌 곳에서 큰 호응을 얻었으며, 한국관광공사 역시 엄청난 홍보 효과를 얻었다고 볼 수 있다. 공공기관은 수익에 대한 고려는 잘 반영하지 않기 때문에 단순히 높은 노출효과 자체만으로도 충분히 의미 있는 홍보가 될 수 있다. 셋째, 지방자치단

체 경상북도 채널인 ‘보이소TV’의 ‘우리동네 자랑 씨리즈’는 경북 지역 특산품을 발굴하고 실제 판매를 유도하는 영상이다. 해당 기관의 관계자의 말에 따르면 유튜브 콘텐츠 제작으로 인해 매출이 지난해 대비 두 배 늘었다고 하였다(미디어오늘, 2021).

이와 같이 공공기관에서 유튜브를 통해 얻는 효과는 높다고 할 수 있다. 실제로도 대다수의 공공기관이 유튜브 채널을 개설하였다. 다음 [표 3]은 2022년 3월 15일 기준, 공공기관 경영정보공개시스템인 알리오(Alio) 사이트에서 살펴본 기획재정부에서 지정한 공공기관의 유튜브 채널 개설 현황이다.

[표 3] 공공기관 유튜브 채널 개설 현황

기관 구분		유튜브 채널 현황		총 개수
		운영	미운영	
공기업	시장형	15	0	15
	준시장형	20	1	21
준정부기관	기금관리형	13	0	13
	위탁집행형	78	3	81
기타공공기관		198	22	210
계(%)		324(92%)	26(8%)	350(100%)

[표 6]에서 알 수 있듯이 기획재정부에서 지정한 350개 공공기관 중 324개(92%)의 기관에서 유튜브 채널을 개설한 것으로 나타났다.

2.1.2.3 유튜브 기록 특징

소셜미디어 활용의 증가는 기록학 측면에서도 중요한 현상으로 인식하였다. 이에 따라 소셜미디어를 기록으로 바라보고자 하는 연구가 진행되었다. 본 연구는 그 중 유튜브 기록에 주목하였다. 유튜브 기록에는 소셜미디어 기록의 특성뿐만 아니라 영상 기록의 특징도 가지고 있다. 다수 논문에서 살펴본 소셜미디어 기록과 영상 기록으로서의 유튜브 기록의 특징은 3가지 정도로 요약할 수 있다.

첫 번째 특징은 원활한 소통이다. 케빈 알로카(Kevin Aloca, 2018)는 “유튜브의 미학은 사람들이 만드는 것이며 그것을 만드는 사람과 지켜보는 사람들 간의 연결을 방해하는 요소들을 없애는 것이 시청자들이 가장 바라는 것이다”라고 하였다. 이는 의사인간관계(pseudo-relationship)와 비슷한 의미를 갖는다. 의사인간관계란 이용자가 미디어 프로그램 속에서의 인물과 정서적으로 친밀감과 유대감을 형성하는 것이다(Horton & Wohl, 1956). 장운석(2019)은 의사인간관계가 미디어 이용을 초래하고, 프로그램과의 상호작용을 반영하며, 이용의 결과에 영향을 주는 등 미디어와 이용자의 관계에 대한 이해에 중요한 의미를 지니는 변인으로 분석된다고 하였다. 더불어 라디오 DJ와 청취자와의 관계를 예시로 들면서 유튜브 채널과 이용자가 소셜미디어의 특성 중 하나인 의사인간관계를 형성하는 것이 중요하다고 하였다. 이는 소셜미디어에서 콘텐츠가 업로드 되는 동시에 ‘좋아요’나 ‘댓글’ 등 이용자와의 실시간 소통을 통해 기관은 콘텐츠에 대한 호응도를 즉시 파악(정혜정, 2018)할 수 있고, 이를 반영하는 양방향 소통과 같은 맥락이다.

양방향 소통과 관련해서 우선, ‘좋아요수’와 ‘조회수’와 같은 이용자 참여 수치 정보는 다수의 행위와 판단을 집적시켜 추출한 결과로서 사회적 정보 중 하나라고 할 수 있다. 즉, 다양한 이용자들이 표현한 것을 다른 이용자가 보고 판단을 하고, 또 다른 행위를 하는 데 도움이 되는 정보가 될 수 있다는 것이다(임다운, 2017). 또한 이를 통해 영상에 대한 시청자들의 의견을 한눈에 살펴볼 수 있으며, 영상에 대한 영향력과 대중의 관심도를 파악할 수 있는 중요한 정보이다. 다음으로 댓글과 대댓글과 같은 정보를 활용함으로써 기관과 이용자, 이용자와 또 다른 이용자끼리의 소통 역시 가능하다고 할 수 있다. 더 나아가 작성된 댓글들 중 기관에서 몇몇을 선정하여 그에 대한 답을 해주는 Q&A 영상을 제작하는 경우도 있다.

유튜브는 청각이 불편한 이용자, 영상의 언어와 다른 언어를 사용하는 이용자 등 다양한 이용자 집단을 위해 자막을 사용하여 영상의 내용을 이해할 수 있게 한다(한규린, 2021). 사회적 약자들에 대한 유튜브의 이러한 특징은 그들에게 정보습득의 기회를 줄 수 있고, 소통과 참여를 증진시킬 수 있다. 시공간을 초월하여 어떠한 제한 없이 자신의 의견을 피력함으로써 참여의 기

회와 사회 관계성을 높여줄 수 있고, 실제 사회구조에서보다 활발하게 자발적으로 참여할 수 있다(최지혜, 2021).

공공기관은 각기 다른 유형의 이용자 참여로 인한 다양한 피드백 중 부정적인 평가는 지양하고 긍정적인 평가는 지향함으로써 보다 발전된 영상을 업로드할 수 있으므로 이는 소통 측면에서 중요하다고 할 수 있다. 송주형(2012)은 기록관리 거버넌스 실현을 위해서는 개개인의 참여와 공유 및 소통은 핵심적인 가치가 있다고 하였다. 이에 따라 그 가치를 이용자들이 충실히 이행하고 있으므로 소셜미디어는 기록으로서의 가치가 있으며, 사회의 정체성과 집단기억, 아래로부터의 역사, 다양성과 개별성을 인정하고 참여를 강조하는 기록학 패러다임의 변화에서도 소셜미디어를 기록화하는 의미와 일치한다고 하였다.

둘째, 사회, 문화적 특징이다. 19C의 인쇄 매체, 20C의 TV와 라디오를 비롯한 전자 매체가 존재하듯 21C는 소셜미디어의 세기로 기억할 수 있다. 새로운 매체의 출현은 인간의 의식구조 및 문화적 수용방식에도 영향을 미칠 수 있다. 즉, 매체의 역사는 곧 인류의 역사라고 해도 과언이 아니다(천현순, 2011). 그러한 의미에서 현재 전 세계 사람들이 활발하게 사용하고 있는 유튜브의 사회, 문화적인 영향력과 의미는 그 자체로 기록되어야 할 만큼의 역사적 사건이다. 유튜브 영상의 사회, 문화적 가치는 미디어의 공공성과 공익성에도 직결되는 요소이다. 태보라(2020)는 유튜브 채널에 업로드된 영상을 통해 당시 사회에서 주목하고 있는 것을 알 수 있으며, 이용자와의 적극적인 소통과 긍정적인 타문화 수용으로 인한 문화적 가치가 형성된다고 하였다.

영상의 다양한 댓글을 통해서도 사회 모습을 파악할 수 있다. 댓글에는 국민 대부분이 당시에 중요하게 여기는 가치 및 이념 등의 양상과 실태를 보여주는 관점을 확인할 수 있으며(임다운, 2017), 이용자들 간에 비슷한 관심사나 사회적 이슈에 대한 공감 댓글을 통하여 서로 연결되고, 공유할 수 있다(윤승욱, 장준갑, 김건, 2019). 또한 댓글을 통해 익명성에 기반하여 자유롭게 의견을 제시할 수 있으므로 다양한 여론이 형성될 수 있는 공론의 장이 구축될 수 있다. 이는 위르겐 하버마스(Jürgen Habermas)가 언급했던 평등한 미디어가 되기 위한 조건인 “누구나 자유롭게 대화에 참여하고 행동할 수 있으

며, 어떤 참여자든 자신의 의견 피력에 대한 내적, 외적 강압을 받지 않아야 한다”에도 성립한다(송주형, 2012). 박유경(2014)은 소셜미디어 콘텐츠에 작성된 댓글을 통해 지금까지 기성 언론이 외면하고 감추려 하는 부분에 대해 이용자들의 솔직한 생각들을 확인할 수 있다고 하였다. 이에 언론사의 정해진 기준에서 방송되는 뉴스보다 소셜미디어는 훨씬 ‘솔직한 저널리즘’이라고 하였다. 이처럼 소셜미디어의 댓글은 사회의 여론을 주도할 수 있으며, 이른바 21C 공론의 장으로서의 역할을 수행하고 있다(윤승욱, 장준갑, 김건, 2019).

사회, 문화적인 특징을 종합하면 다음과 같다. 현재 공공기관에서 홍보의 주된 수단으로 사용하고 있는 유튜브에 영상을 등록하는 행위와 더불어 영상 속에서 사용하는 자막, 유행어를 통해 당시 사회, 문화의 트렌드를 발견할 수 있다. 또한 영상을 시청한 후 이용자 참여 수치 정보와 수많은 댓글들로 형성되는 공론의 장을 통해 당시의 문화적인 특징을 발견할 수 있다. 유튜브에서 어떠한 이유로 하루아침에 서비스를 중단한다면 저장되지 않은 콘텐츠들과 관련 정보들은 사라지게 된다. 실제로 한국에서 ‘네띠앙’ 기업이 파산하면서 서비스가 중단된 사례가 있는데 이러한 사례가 다시는 없을 것이라는 보장이 없다(송주형, 2012). 이러한 점에 있어 공적 업무 수행의 결과로 제작된 유튜브 영상과 문화적인 가치가 있는 이용자 참여 정보를 체계적으로 관리하고 보존하여 미래의 이용자에게 보여줄 필요가 있다.

셋째, 영상기록으로서 특징이다. 유튜브는 기본적으로 동영상을 기반으로 운영되는 소셜미디어 플랫폼이다. 영상은 텍스트, 이미지, 소리 등의 수단을 입체적으로 사용하기 한다. 이로 인해 커뮤니케이션 과정에서 메시지 전달이 단순한 텍스트, 음성과 같은 한 가지 커뮤니케이션 수단보다 훨씬 이해하기 쉽다(장운석, 2019). 또한 포맷 및 컨테이너에 담긴 움직임 있는 형태로 내용이 표현된다는 특수성을 가지고 있다(심보미, 2018). 즉, 이미지, 녹음본 등과 같은 정적기록물보다 당시의 상황을 생동감 있고 현장감 있게 사실을 전달하는 동적 기록물로서 전달하려는 의미를 쉽게 파악할 수 있다. 따라서 당시 업무 활동의 증거로서 영상 기록은 개인과 사회의 활동과 경험을 나타내는 증거적, 역사적 가치를 갖는다(최진호, 2017). 영상기록의 중요성은 2015년 KBS 특별생방송 ‘이산가족을 찾습니다.’ 기록물이 13번째 세계기록유산으로

등재된 것을 통하여 알 수 있다(심보미, 2018). 이처럼 비언어적인 소통인 표정, 말투 등이 영상에 그대로 표현되기 때문에 실제 움직이는 모습을 담을 수 있다. 사람들은 읽고 내용을 파악하는 데 오래 걸리는 텍스트보다 영상을 보는 것을 선호한다는 증거는 텍스트를 기반으로 한 트위터에서 모바일의 대중화와 함께 사진 및 영상을 기반으로 한 인스타그램과 유튜브로 유행이 넘어 온 것에서 확인할 수 있다(한규린, 2021).

2.1.3 공공기관 유튜브 동영상 아카이빙 필요성

유튜브 기록은 양방향 소통의 증거, 사회, 문화적 특징, 영상 기록으로서의 특징을 가지고 있는 것을 확인하였으며 현재 국내 350개의 공공기관 중 324개(92%)의 기관이 유튜브 채널을 개설한 것도 확인하였다. 이처럼 대다수의 공공기관이 유튜브를 운영하고 있다면, 생산되는 동영상은 웹 기록물로서 공적 업무 수행의 결과물로 인식할 필요가 있다. 즉 공공기록물로서 가치 있는 영상을 선정하여 체계적으로 관리하고 보존할 필요성이 있다.

그러나 유튜브를 포함하는 소셜미디어 기록은 전통적인 기록의 관점에서 살펴보면 기록의 4대 속성을 만족한다고 보기 어렵다(이상현, 2020) 즉, 소셜미디어 기록은 진본성, 신뢰성, 무결성, 이용가능성을 확보하지 못해 기록으로 보기 어렵다. 또한 공공기관에서 소셜미디어에 콘텐츠를 업로드하는 행위는 상부의 결재를 받고 허가된 콘텐츠를 등록하지만 누군가가 해킹을 통해 소셜미디어에 계정에 로그인하여 부적절한 콘텐츠를 업로드할 수 있고, 반대로 기존에 업로드되어 있는 콘텐츠를 쉽게 삭제시킬 수 있다. 이는 웹 기록물의 특징 중 휘발성과 관련이 있다. 손원기(2013)는 종이기록과 같은 정적기록물이 아닌 디지털코드로 구성되어 있는 웹 기록물은 다른 기록 매체보다 쉽게 소멸될 수 있다고 하였다. 이는 유튜브에도 적용될 수 있다.

그럼에도 불구하고 앞서 언급한 것처럼 유튜브 동영상이 갖는 양방향 소통, 사회·문화적 특징은 공적 업무 수행의 결과물이자 다양한 측면에서 중요한 기록으로 볼 수 있다. 예를 들면 소셜미디어를 비롯한 유튜브 동영상의 ‘좋아요’ 또는 ‘댓글’은 기본적으로 익명을 기반으로 작성된다. 하지만 익명으

로 작성되었기 때문에 당시 사회에 대한 국민의 생각이 적나라하게 드러난다. 송주형(2012)은 SNS를 통해 전달된 허위 뉴스 등이 당시의 시대적 상황과 배경을 설명할 수 있는 또 다른 기록이 될 수 있다고 하였다. 즉, 기록 그 자체로 보는 것이 아닌 사회의 반영이라는 의미에서 SNS를 전체적인 맥락에서 살펴보면 이는 중요한 기록이 될 수 있다고 하였다.

앞서 살펴본 국내 기록물 장기보존에 관련된 표준은 NAK 30과 NAK 31-1, NAK 31-2이다. 해당 표준을 사용하여 공공기관 유튜브 동영상을 장기보존 할 수 있는 방법을 살펴보면 다음과 같다. NAK 30의 경우 본 연구 대상인 유튜브 동영상 기록물은 소셜미디어 기록인 웹 기록물과 동영상 기록이 결합된 형태이므로 문서보존포맷에 적합하지 않다. 또한 NAK 30에서도 동영상, 비디오, 오디오 유형 매체는 PDF/A-1의 적용이 불가능하다고 명시되어있다. 한희정, 오효정, 양동민(2020)의 연구에서도 국내에서 선택할 수 있는 보존포맷은 PDF/A-1 밖에 존재하지 않아 문서 이외에 행정정보데이터세트, 시청각기록물, 웹 기록물을 적용하기 어렵다고 하였다.

NAK 31-1의 경우 기록물건 장기보존 메타데이터 요소 중 소셜미디어 기록 측면에서 중요한 정보인 이용자 참여와 관련된 요소는 부재하다. 또한 NAK 8을 기반으로 작성하였기 때문에 기록의 진본성, 무결성, 신뢰성, 이용가능성은 보장할 수 있으나 전자문서에 기반한 메타데이터 요소로 이루어져 있어 동영상 기록물을 기술하기 위한 요소가 적다. 이에 따라 유튜브 동영상 기록물을 보존하기 위한 메타데이터로는 적합하지 않다고 할 수 있다.

결국 유튜브를 포함한 소셜미디어 기록을 보존하기 위한 형식은 부재하다고 볼 수 있다. 다만 소셜미디어 중 문서류 위주의 기록물로 구성되어 있는 트위터나 페이스북 기록물에 대해서는 어느정도 기술 할 수 있으나 해당 플랫폼들도 이용자 참여와 관련된 메타데이터 요소가 부재하여 완전하다고 보기 어렵다. 유튜브 영상의 경우는 문서보다 더 어려운 상황으로 볼 수 있다.

2.1.4 메타데이터

2.1.4.1 메타데이터 일반

메타데이터(Metadata)의 메타(meta)는 위치나 상태의 변화, 또는 ‘더 높은, 초월한’이라는 뜻을 갖는 단어이다. 메타데이터란 데이터보다 더 높은 데이터를 의미하는 “데이터에 대한 데이터”로 구조화된 정보를 분류 및 분석하고 부가적인 정보를 추가하기 위해 그 데이터 뒤에 함께 따라가는 정보이다(곽성호, 2012). 즉, 데이터를 설명하는 데이터 요소 집합을 의미한다. 기록학 용어사전(2008)에서는 정보를 지적으로 통제하고 구조적으로 접근할 수 있도록 정보 유형을 정리한 이차적인 정보라고 정의하였다. 메타데이터의 역할은 방대한 양으로 생산되는 물리적, 논리적 자원을 전자적인 방법을 통하여 관리할 수 있고, 다양한 자원에 대한 관리 할 수 있게 한다(박인우, 2018).

남태우, 이승민(2010)은 메타데이터가 “데이터에 대한 데이터”라는 정의를 충족시키기 위해 필요한 기능 요건을 제시하였다. 첫째, 정보자원에 대한 정보를 기술하고 표현할 수 있어야 한다. 둘째, 모든 유형의 정보자원을 다룰 수 있어야 한다. 셋째, 기계가 이해할 수 있고 처리할 수 있는 구조화된 정보라는 것이 뒷받침되어야 한다고 하였다. 또한 적용되는 상황에 따라 다양한 의미를 포괄하게 되는 복합적인 개념이며, 어떤 측면이 고려되는지 그리고 어떤 목적을 충족시키기 위해 사용되는지에 따라 그 의미와 기능이 다르게 적용되기도 한다고 하였다.

ISO 15489, ISO 23081에서는 메타데이터를 ‘시간과 공간을 초월하여 기록의 생산, 관리와 이용이 가능하도록 하는 구조화된 혹은 반구조화된 정보’로 정의하고 있다. 기록관리의 원칙으로서 기록은 ‘메타데이터와 그 내용으로 구성된다’고 선언함으로써 메타데이터가 기록을 구성하는 중요한 부분임을 명백히 밝히고 있다. 메타데이터의 역할은 업무 맥락, 기록과 기록 시스템관의 의존성과 관련성, 법적, 사회적 맥락에 대한 관련성, 기록을 생산, 관리 및 이용하는 행위자에 대한 관련성을 기술할 수 있어야 한다. 즉, 기록 메타데이터

자체가 기록으로써 관리 되어야하며, 훼손, 허가받지 않은 삭제나 보유, 파괴로부터 보호되어야 한다.

미국의 국가정보표준기구(NISO)는 메타데이터의 유형을 설명적(Descriptive) 메타데이터, 구조적(Structural) 메타데이터, 관리적(Administrative) 메타데이터로 구분하였다(Jenn Riley, 2017). 이 중 관리적 메타데이터는 기술(Technical) 메타데이터와 권리(Rights) 메타데이터, 보존(Preservation) 메타데이터로 구분된다. 디지털 보존에 대한 중요성이 높아짐에 따라 보존 메타데이터가 관리적 메타데이터에서 분리될 수도 있다(서은경, 2005).

2.1.4.2 아카이빙을 위한 메타데이터

디지털 아카이빙이란 지속적으로 보존할 가치가 있는 디지털 객체를 장기간 관리하여 이후의 이용을 보장할 수 있는 활동이며, 기록의 보존에서 넓은 의미로 사용된다(양은정, 2021). 이러한 활동은 디지털 객체의 형태와 구조를 비롯하여 해당 객체의 생산부터 수집, 기술, 저장, 보존 및 접근 등이 포괄적으로 고려되어야 하는 과정이다(Hodge, 2000). 기록관리 분야에서는 아카이빙이라는 용어를 기록의 보존을 처리하는 과정이라는 의미로 사용하고 있다(양은정, 2021). OCLC/RLG Working group(2002)은 보존 메타데이터가 이러한 과정을 지원할 수 있다고 하였으며, 이를 통해 디지털 자원의 손상을 방지하는 장기간 생존가능성(Viability), 0과 1로 이루어진 비트스트림을 이용자가 볼 수 있거나 컴퓨터에서 처리할 수 있는 형식으로 변환할 수 있는 실행가능성(Renderability), 렌더링(rendering)된 내용을 이용자가 해석하고 이해할 수 있는 이해가능성(Understandability)을 유지할 수 있다고 하였다. 이를 위해 보존 메타데이터에 포함되어야 하는 정보는 보존정보와 더불어 디지털 정보 자원 자체에 대한 포괄적인 기술사항이라 할 수 있다. 이승민(2015)은 디지털 아카이빙에서 디지털 객체가 수록하고 있는 콘텐츠의 진본성, 무결성, 신뢰성, 이용가능성을 정보 환경의 변화에 상관없이 지원할 수 있어야 한다고 하였으며, 이를 위해서는 아카이빙 대상이 되는 정보자원에 대한 상세한 기술사항이

뒤따라주는 보존 메타데이터가 필수적이라고 하였다. 이인혁(2017)은 보존 메타데이터 요소 설계를 위해서 아카이빙 대상이 되는 객체의 특성을 파악하는 것이 중요하다고 하였다.

이러한 이유로 보존 메타데이터는 일반적으로 사용되는 메타데이터에 비해 많은 정보들이 포함된다고 할 수 있다. 서은경(2005)은 보존 메타데이터를 구성하는 정보에는 기술적, 관리적, 구조적, 맥락, 이력, 출처, 진본성 인증 등과 같은 정보가 포함된다고 하였다. 김혜영(2010) 역시 보존 메타데이터에 포함되는 다양한 정보 유형을 구조 정보, 테크니컬 정보, 이벤트 정보, 권한 정보, 변경이력 정보, 수집 정보로 나누어 범주화하였다. 디지털 자원의 장기적 보존을 가능하게 하는 CEDARS(CURL Exemplars in Digital Archive European Deposit Library) 프로젝트에서 제안한 보존 메타데이터는 디지털 정보자원에 대해 원활한 접근을 지원하기 위해 설명적, 관리적, 기술적, 법적인 정보를 기술할 수 있도록 하였다(OCLC/RLG Working Group, 2001). 웹 기록물 아카이빙 메타데이터 연구를 진행한 안영희, 박옥화(2009), 오상훈, 최영선(2009)은 아카이빙 메타데이터를 파일과 비트스트림으로 이루어져있는 웹 정보자원을 영속적으로 관리 및 보존하기 위해 설명적, 구조적, 관리적, 보존적 메타데이터로 구분하였다.

설명적 메타데이터에는 디지털 자원에 대한 설명할 수 있는 설명 정보가 들어가 있으며, 구조적 메타데이터에는 디지털 자원을 표현하기 위한 수단과 방법에 따른 자원의 구조 및 구성에 대한 정보를 제공한다. 관리적 메타데이터는 관리 정보, 권리 정보, 기술 정보로 나누어지는데, 관리 정보는 아카이빙에서 입수된 자원을 관리하고 아카이빙을 운영 및 관리하기 위한 정보를 제공하며, 권리정보는 디지털 자원의 지적 재산권에 관련된 개인, 단체, 기관에 대한 정보를 제공한다. 기술 정보는 디지털 자원을 사용하기 위해 요구되는 기술적인 환경에 대한 정보를 제공한다. 보존적 메타데이터는 출처 정보, 참조 정보, 맥락 정보, 고정 정보로 나누어진다. 이들은 각각 자원의 출처, 이력, 맥락 및 관계, 진본성 보장과 관련된 정보를 제공한다(안영희, 박옥화 2009).

2.2 선행연구

본 연구는 소셜미디어 기록 중 유튜브 기록에 주목하여 관리 및 보존의 필요성을 확인하고, 이를 아카이빙하기 위한 메타데이터를 설계하고자 한다. 이를 위해 유튜브 기록, 소셜미디어 메타데이터, 아카이빙 메타데이터로 구분하여 선행연구를 살펴보았다.

소셜미디어의 사용이 활발해짐에 따라 이를 기록학적으로 바라보고자 하는 연구는 송주형(2012), 최두원 외(2018), 김태영 외(2019), 한규린(2021)의 연구가 존재한다. 이 중 본 연구와 같이 유튜브를 기록학적으로 살펴보고 연구결과에 적용시킨 연구는 송주형(2012), 한규린(2021)의 연구만 존재한다. 최두원 외(2018), 김태영 외(2019)의 연구도 유튜브를 소셜미디어 기록으로 인식하고 관리 및 보존의 필요성을 주장하였지만 다른 소셜미디어 유형을 연구결과에 적용시켰다.

송주형(2012)은 웹 2.0의 발전으로 참여와 공개, 소통의 시대가 도래했다고 하였다. 이에 따라 전통적으로 가지고 있었던 기록관리에 대한 문제를 해결할 수 있는 방법으로 SNS 기록을 제시하였다. 다양한 SNS 중 페이스북, 트위터, 블로그, 유튜브를 기록학적으로 살펴보고 SNS 기록관리 방법을 제안하였다. 구체적으로 기록학의 패러다임 속에서 기록 속성과 아키비스트의 역할 변화와 더불어 기록의 재정의 등을 통하여 SNS를 기록으로서 관리해야 하는 근거 및 육하원칙에 따른 SNS 기록관리 원칙을 제시하였다.

한규린(2021)은 공공기관에서 유튜브에 업로드한 코로나19 관련 영상을 대상으로 유튜브 기록의 수집방안을 제안하였다. 이를 위해서 저장공간, 휘발성, 저작권 보호, 유튜브의 특성을 고려한 수집으로 나누어 수집방안을 살펴보았으며, 저작권 문제를 해결하기 위해 설문을 통하여 시사점 및 보완점을 도출하였다.

최두원 외(2018)는 정치적 목적으로 이용되는 SNS 기록에 주목하였다. 대통령의 모든 활동이 트위터, 페이스북, 유튜브 등의 기록으로 생산되는 것에

관하여 SNS가 기록학적으로 독특한 중요성과 높은 가치를 지닌다고 하였다. 하지만 국내 대통령 SNS 기록물의 관리가 제대로 되어있지 않았고, SNS 기록물의 수집, 관리를 위한 마땅한 법의 근거도 미흡하다고 하였다. 이에 따라 역대 국내 대통령들의 SNS 기록물 현황을 기록학적인 관점에서 접근하여 관리방법에 대한 문제점을 제시하였고 해외의 사례연구를 통해 대통령 SNS 기록물 관리방안을 국내에 도입하고자 하였다.

김태영 외(2019)는 다양한 공공기관에서 소셜미디어를 통하여 국민과 소통을 하고 있다고 하였다. 이에 따라 기록관리 영역에서 소셜미디어 기록에 관한 수집과 보존의 논의가 주요 쟁점이 되었다고 하였다. 하지만 아직 국내에서는 소셜미디어 수집과 보존과 관련해서 마땅한 지침이 없는 것을 문제점으로 제기하였다. 이에 국외의 페이스북, 트위터 및 유튜브 동영상 아카이빙 프로젝트 사례를 조사, 분석하여 소셜미디어 유형에 따른 수집과정과 보존 모형을 제안하였다. 또한 제안 모형을 경상남도청 소셜미디어 중 트위터를 대상으로 직접 적용하여 시계열 분석을 통한 수집에 관한 전략 및 수집주기에 대한 다양한 시사점을 도출하였다.

이외에도 장운석(2019), 태보라(2020), 임다운(2017), 윤승욱, 장준갑, 김건(2019)의 연구에서도 소셜미디어 기록과 관련하여 양방향 소통과 사회, 문화적인 측면에 대해서 언급한 바 있다.

둘째, 소셜미디어를 기록학적으로 관리 및 보존해야 한다는 측면에서 이를 기술하기 위한 연구도 진행되었다. 황운영 외(2013)는 OAIS 참조모델을 바탕으로 정부부처의 페이스북 페이지 기록물의 메타데이터 및 장기보존 포맷을 설계하였다. 이를 통해 소셜 네트워크 서비스 기록물의 역사적이고 과학적인 가치를 보존하고자 하였다. 해당 포맷은 국내 기록관리 메타데이터 표준의 장기보존포맷과 같은 구조로 설계하였으며, 메타데이터 요소는 기존의 국내 메타데이터 표준을 확장하여 12개 요소를 설계하였다.

박유경(2014)은 공공기관에서 운영하는 트위터, 페이스북, 블로그의 구조를 분석하고 이를 아우르는 소셜미디어 표준 메타데이터를 설계하고자 하였다. 이를 위해 중앙행정기관 53곳과 지방자치단체인 서울시에서 사용하는 소

설미디어 현황을 살펴보았다. 또한 국내 기록관리 표준 메타데이터 요소를 피츠버그 프로젝트의 메타데이터 모델 범주에 적용하여 소셜미디어 메타데이터 요소를 제안하였다. 해당 연구의 한계점으로는 모든 플랫폼을 포괄할 수 없다는 점과 연구 대상을 서울 소재 지방자치단체로만 한정된 점, 피츠버그 모델 범주 중 ‘조건’을 제외했다는 점이 있다.

이외에 류지연(2019)은 SNS와 교육용 자료로써 효과적인 유튜브가 브라우징 및 카테고리별로 검색할 수 있는 기능이 부재하다고 하였다. 이에 유튜브 메타데이터 요소를 제안하여 유튜브 내에서 검색과 접근을 해결하고자 하였다. 이를 위해 넷플릭스, 비메오(Vimeo), 한국영상자료원의 메타데이터 현황 및 다양한 국외 멀티미디어 메타데이터 표준을 비교, 분석하여 33개의 요소를 도출하였다. 류지연(2019)의 연구는 유튜브를 주제로 메타데이터를 설계했다는 점에서 본 연구와 유사하다고 할 수 있지만, 유튜브 내에서 효과적인 검색을 위한 연구이다. 따라서 유튜브 동영상을 기록학적으로 바라보고 이를 아카이빙하고자 하는 본 연구와 차이점이 있다.

셋째, 아카이빙 메타데이터와 관련해서 유튜브가 동영상을 기반으로 운영되는 소셜미디어임을 고려하여 영상 아카이빙 메타데이터 연구를 살펴보았다. 또한 유튜브 동영상이 디지털 기록물이라는 관점에서 웹 기록물 아카이빙 메타데이터 연구로 구분하여 살펴보았다.

영상 아카이빙을 위한 메타데이터 요소 설계로는 박신애(2010)의 연구가 있다. 박신애(2010)는 디지털 기록의 유형이 다양해짐에 따라 새롭게 생겨난 유형인 TV 뉴스 기록에 주목하였다. TV 뉴스 기록은 정보전달과 사회상이 반영되는 영상 기록으로서 가치 있다고 하였다. 또한 인터넷의 발전으로 디지털화된 방송 아카이브 및 뉴스 제작 시스템을 관리 및 활용하고 있는 상황 속에서 각 방송사의 상호 교류 및 기록의 맥락을 파악할 수 있는 TV 뉴스 기록의 아카이빙을 위한 표준 메타데이터를 설계하였다. 이를 위해 TV 뉴스의 특성 및 제작과정과 TV 뉴스의 기록 계층을 살펴보았다. 다음으로 방송과 관련된 메타데이터 표준들을 비교, 분석한 후, 국제 보존기록 기술규칙인 ISAD(G)의 기술요소들을 반영하였다. 그 결과 TV 뉴스의 매체 특성과 기록으로서의 4대 속성이 반영된 8개 영역, 48개 요소를 설계하였다.

웹 기록물 아카이빙을 위한 메타데이터 요소 설계 연구는 오상훈, 최영선(2009), 안영희, 박옥화(2009)의 연구가 있다. 오상훈, 최영선(2009)은 국립중앙도서관의 웹 아카이빙 프로젝트인 OASIS를 대상으로 하였다. OASIS의 체계적인 수집, 관리와 더불어 이용자 서비스를 위한 보존단계에서의 단계별, 기능별 요구사항에 적합한 메타데이터 구조와 요소를 개발하고자 하였다. 해당 연구의 목적을 달성하기 위해 OASIS 프로젝트의 현황과 국외 메타데이터 표준인 Dublin Core, NLA, OCLC, PREMIS 등을 비교, 분석하였고 이를 ISO 14721 OAIS 참조모형 구조에 적용하여 아카이빙 메타데이터를 설계하였다. 그 결과 9개 범주인 설명, 구조, 권리, 관리, 기술, 출처, 참조, 문맥, 인증정보 범주의 상하위요소를 개발하였다.

안영희, 박옥화(2009)는 여러 대학에서 생산되는 이러닝 콘텐츠의 체계적인 관리 및 활용을 위해 이러닝 콘텐츠 아카이빙을 위한 메타데이터를 구축하고자 하였다. 이를 위해 13개 대학의 이러닝 센터 담당자들의 설문조사를 실시하였다. 그 결과 이러닝 콘텐츠 아카이빙에 대한 인식은 필요한 것으로 나타났으나, 보존관련 메타데이터 요소가 부재하여 아카이빙하기에 현실적으로 어렵다는 결과가 나왔다. 이를 통해 이러닝 콘텐츠 관련 메타데이터인 MIT OCW(Open CourseWare)와 KEM3.0을 LOM 1484.12.1.-2002에 기반하여 기존 KEM3.0 요소의 추가 및 확장한 KEM3.0+를 새로 정의하였다.

웹 기록물 아카이빙에 대한 연구는 이밖에도 유효림(2007), 김정심(2008), 최윤정, 차승준, 이규철(2009)의 연구가 있었다. 유효림(2007)과 김정심(2008)은 정부 및 행정부처의 웹사이트를 아카이빙하고자 국내외 사례분석을 통해 아카이빙의 전반적인 프로세스 및 수집정책을 제시하였다. 최윤정, 차승준, 이규철(2009)은 정적 기록물에 속하는 표면 웹 기록물 아닌 이용자 요구로 인해 갱신될 수 있는 심층 웹 기록물을 아카이빙하고자 더블린코어(Dublin Core), 국내 전자기록물 장기보존포맷 기술 규격을 기반으로 공공기관의 심층 웹 기록물을 나타내는 KoDeWeb의 메타데이터 항목을 설계하였다.

Ⅲ. 메타데이터 요소 설계

메타데이터 요소 설계는 요소 조사, 요소 매핑 및 요소 도출의 단계로 수행하였다. 첫 번째 메타데이터 요소 조사단계에서는 유튜브 동영상 아카이빙 메타데이터를 설계하기 위해 해당 분야의 특징이 반영된 국내외 메타데이터 표준들을 살펴보았다. 국내외에서 공식적으로 인정받은 다양한 분야의 메타데이터 표준을 살펴보고 이를 적용함으로써 상호호환성을 보장할 수 있다. 이를 위해 국내에서 살펴본 표준은 국가기록원의 기록관리 메타데이터(NAK 8:2021(v2.2))이다. 국제표준인 ISO 15489와 ISO 23081을 기반으로 작성되었기 때문에 기록의 진본성, 신뢰성, 무결성, 이용가능성을 보장할 수 있다. 또한, 본 연구에서 공공기관 유튜브 동영상 기록물은 공공기록물로서 관리 및 보존되어야 함을 지속적으로 언급하고 있다. 이에 국내 공공기록물을 관리하기 위한 메타데이터의 표준이 되는 NAK 8의 요소와 국외의 다양한 분야의 메타데이터 표준 요소들을 결합하고자 한다.

국외에서 살펴본 메타데이터 표준은 아카이빙을 위한 메타데이터라는 특성을 고려하여 보존기록과 관련된 표준과 동영상 기록물과 관련된 표준을 살펴보았다. 보존기록과 관련된 표준은 ISAD(G)와 PREMIS를 살펴보았다. 동영상 기록물과 관련된 메타데이터 표준은 방송, 영상분야의 메타데이터 표준인 PBCore를 조사하였다.

국내외 표준 이외에 유튜브의 소셜미디어 기록의 특수성을 고려하여 유튜브 내에서 영상을 발행하고 이용할 시, 생성되는 정보(유튜브 메타데이터)를 조사하였다.

두 번째는 메타데이터 요소 매핑 및 요소 도출이다. 국내외 표준과 유튜브 메타데이터를 아카이빙 메타데이터 구조를 토대로 설명, 구조, 관리, 보존 메타데이터로 구분하고 각 영역의 요소들을 비교하고 매핑하였다. 영역별로 요소들을 매핑하는 과정 중 유튜브 메타데이터의 소셜미디어 기록과 관련된 요소들이 매핑되지 않아 이용자 참여 메타데이터를 새롭게 추가하였다. 이에 따라 유튜브 동영상의 소셜미디어, 동영상, 아카이빙의 측면을 모두 만족하는 유튜브 동영상 아카이빙 메타데이터 요소를 도출하였다.

3.1 메타데이터 요소 조사

본 연구의 목적은 유튜브 동영상 아카이빙을 위한 메타데이터를 설계하고 핵심 요소를 선정하는 것이다. 윤선정(2008)은 새로운 분야의 메타데이터 개발을 위해서는 기존 메타데이터 표준의 참고와 해당 영역의 특성을 반영하는 요소의 비교, 분석을 해야한다고 하였다. 이후 기존 요소의 유지, 확장, 재정의 등의 과정을 거친 후 개발하고자 하는 영역의 특징을 적용하여 설계한다고 하였다. 본 연구는 안영희, 박옥화(2009), 오상훈, 최영선(2009)에서 진행한 아카이빙 메타데이터 구분을 토대로 국내 기록관리 메타데이터 표준 요소인 NAK 8:2021(v2.2), 국외 보존 기술 표준인 ISAD(G), 국외 보존 메타데이터 요소 사전인 PREMIS, 국외 영상 메타데이터 표준인 PBCore, 유튜브 내에서 영상을 발행하고 이용할 시, 생성되는 정보(유튜브 메타데이터)를 비교하고 새로운 요소를 추가하였다.

3.1.1 국내 기록관리 메타데이터 표준 요소

국내 기록관리 메타데이터 표준은 국가기록원의 NAK 8이다. NAK 8의 제정 목적은 장기간에 걸쳐 기록물의 진본성, 무결성, 신뢰성 및 이용 가능성을 보장하기 위해 공공기관이 생산 또는 접수하는 기록물에 대한 맥락과 내용, 구조 및 기록 생애주기 동안 기록물 관리기관이 관리할 사항을 기술하기 위함이다. 이 표준은 결재과정에서 생산된 문서류 위주의 기록물에 대한 메타데이터 요소를 제시하지만, 시청각기록물, 웹 기록물, 행정정보 데이터 세트 등의 기록물도 공통 항목에 대해서는 이 표준을 적용할 수 있다.

NAK 8은 장기 보존을 포함하여 기록물 생애주기 전체에 걸쳐 필요한 메타데이터가 획득될 수 있도록 적용 범위를 영구기록물관리기관으로 확장하고, 관련 법령 및 현장 실무에 적합하도록 2012년에 1차 개정하였다. 이 개정은

기록관리 표준인 KSXISO 15489-1(문헌 정보-기록관리-제1부: 일반사항)과, KS X ISO/TR(15489-2 문헌 정보-기록관리-제2부: 지침), KS X ISO 23081-1(문헌 정보-기록관리과정-기록 메타데이터-제1부: 원칙), KS X ISO 23081-2(문헌 정보-기록관리과정-기록 메타데이터-제2부: 개념과 실행 고려사항)가 그 대상이다.

2016년 2차 개정에서는 기록을 업무의 자산으로 인식하는 KS X ISO 30300(문헌 정보-기록경영시스템-제1부: 기본사항 및 용어)을 반영하여 기존 국제기록관리 표준에서 개정된 ISO 15489-1:2016(Information and documentation—Records management—Part 1: Concepts and principles)까지 적용할 수 있도록 개정되었다.

2021년 3차 개정에서는 목록 공개구분과 관련한 사항을 적용하기 위해 메타데이터 요소 및 선택값 등을 추가 반영하였다.

NAK 8의 요소는 23개 상위요소, 66개 하위요소, 58개 세부요소의 3계층으로 구성된다. 하위요소 중 세트로 작성되어야 할 요소들은 컨테이너로 묶여 하위요소-세부요소로 계층화된다. 상위요소 중 필수요소는 생산자, 기록계층, 기록 식별자, 기록물명, 전자기록물여부, 유형, 크기, 분류, 일시, 보존 기간, 보존장소, 권한으로 총 12개이다. 국가기록원의 기록관리 메타데이터 표준 요소는 다음 [표 4]와 같다.

[표 4] 국가기록원 기록관리 메타데이터 표준 요소

상위요소	하위요소	세부요소
1. 생산자	1.1 생산자유형	
	1.2 기관명	
	1.3 기관코드	
	1.4 부서명	
	1.5 부서코드	
	1.6 개인명	
	1.7 개인코드	
	1.8 직위(직급)명	
	1.9 직위(직급)코드	
2. 기록계층		
3. 기록식별자	3.1 기본식별자	
	3.2 시스템식별자	3.2.1 시스템식별자유형
		3.2.2 시스템식별자

	3.3 보조식별자	3.3.1 보조식별자유형 3.3.2 보조식별자
4. 기록물명	4.1 제목	
	4.2 기타제목	4.2.1 기타제목유형 4.2.2 기타제목명
5. 기술	5.1 기술유형	
	5.2 기술내용	
6. 주제	6.1 주제유형	
	6.2 주제명	
7. 전자기록물 여부		
8. 유형	8.1 기록유형	
	8.2 사본유형	
	8.3 기타문서유형	
	8.4 컴포넌트유형	
9. 포맷	9.1 포맷명	
	9.2 포맷버전	
	9.3 생성애플리케이션명	
	9.4 생성애플리케이션버전	
10. 매체		
11. 크기	11.1 용량	
	11.2 단위	
12. 분류	12.1 분류체계유형	
	12.2 분류값	12.2.1 분류ID 12.2.2 분류명 12.2.3 분류계층
13. 일시	13.1 생산일시	
	13.2 종료일시	
	13.3 등록일시	
	13.4 시행일시	
14. 생산이력	14.1 생산시스템	
	14.2 생산부서	14.2.1 기관명 14.2.2 기관코드 14.2.3 부서명 14.2.4 부서코드
	14.3 생산유형	
	14.4 비치기록물	14.4.1 비치사유 14.4.2 비치기간
	14.5 컴포넌트버전	
	14.6 생산경로	14.6.1 생산자 14.6.2 의견 14.6.3 생산컴포넌트유형 14.6.4 생산컴포넌트버전

		14.6.5 처리상태
		14.6.6 처리일시
		14.6.7 처리순번
15. 보존기간	15.1 보존기간	
	15.2 보존기간책정사유	
16. 보존장소		
17. 권한	17.1 비밀	17.1.1 비밀분류
		17.1.2 비밀분류 근거
		17.1.3 보호기간
	17.2 접근범위	
	17.3 공개	17.3.1 공개구분
		17.3.2 비공개사유
		17.3.3 공개제한부분
		17.3.4 공개예정일자
		17.3.5 공개관련근거
		17.4 공공저작물관리
		17.4.2 공공저작물 자유이용허락표시유형
18. 위치	18.1 소장처	
	18.2 소장위치	
19. 관리이력	19.1 관리유형	
	19.2 관리설명	
	19.3 관리일시	
	19.4 관리행위자	19.4.1 기관명
		19.4.2 기관코드
		19.4.3 부서명
		19.4.4 부서코드
		19.4.5 개인명
		19.4.6 개인코드
		19.4.7 직위(직급)명
		19.4.8 직위(직급)코드
		19.5 변경요소
		19.5.2 변경이전값
20. 이용이력	20.1 이용유형	
	20.2 이용내용설명	
	20.3 이용일시	
	20.4 이용자	20.4.1 기관명
		20.4.2 기관코드
		20.4.3 부서명
		20.4.4 부서코드
	20.4.5 개인명	
		20.4.6 개인코드

		20.4.7 직위(직급)명
		20.4.8 직위(직급)코드
21. 보존이력	21.1 보존처리유형	
	21.2 보존처리설명	
	21.3 보존처리일시	
	21.4 보존행위자	21.4.1 기관명
		21.4.2 기관코드
		21.4.3 부서명
		21.4.4 부서코드
		21.4.5 개인명
		21.4.6 개인코드
		21.4.7 직위(직급)명
	21.4.8 직위(직급)코드	
22. 관계	22.1 관계 유형	
	22.2 관계대상식별자	
	22.3 관계설명	
23. 무결성체크	23.1 무결성체크법	
	23.2 무결성체크값	

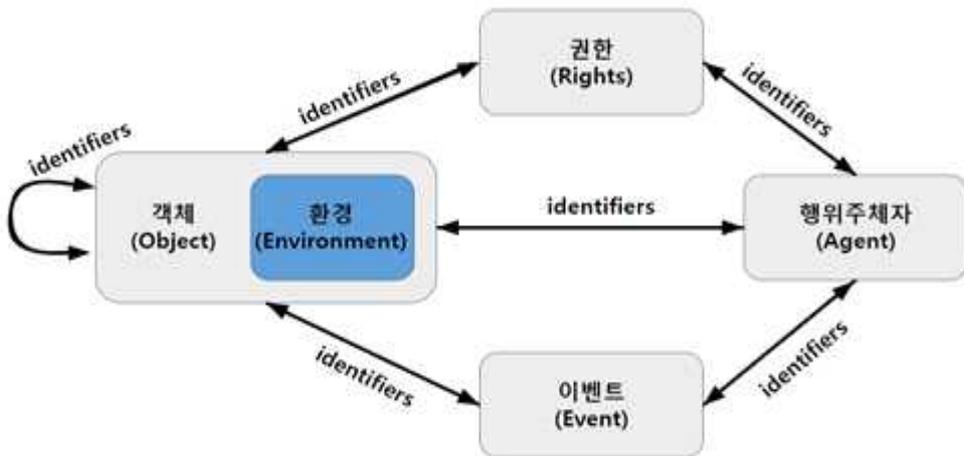
3.1.2 국외 메타데이터 요소

1) PREMIS

OCLC/RLG는 도서관, 기록 보관소, 정부 및 민간 분야의 각국의 전문가들이 속한 PREMIS 워킹 그룹을 구성하였다. 이를 통해 디지털 보존에 광범위하게 적용할 수 있고, 디지털 정보자원의 생성, 관리 및 사용에 대한 지침과 권장사항에 의해 지원되는 구현 가능한 보존 메타데이터의 핵심 요소를 개발하고자 하였다. 이러한 배경으로 2005년 ‘보존 메타데이터를 위한 데이터 사전: PREMIS 워킹그룹 최종 보고서((Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group)’로 발행된 데이터 사전이 바로 PREMIS(Preservation Metadata: Implementation Strategies)이다. 이후 PREMIS 편집위원회는 2008년에 PREMIS 2.0의 발표를 시작으로 두 차례에 걸쳐 PREMIS 2.2까지 출시하였다. 2015년에는 사용자들의 다양

한 피드백을 비롯해 PREMIS의 실행성과 상호운용성에 초점을 맞춘 PREMIS 3.0을 공개하였다.

PREMIS는 [그림 2]와 같이 객체(Object), 환경(Environment), 권한(Rights), 행위주체자(Agent), 이벤트(Event) 개체로 이루어져 있다. 객체 개체는 디지털 정보자원의 관리와 설명을 위해 지적 단위로 간주되는 요소 세트이며, 디지털 정보를 구현하기 위해 필요한 물리적 객체와 관련된 정보를 제공해 줄 수 있다. 환경 개체는 디지털 객체를 재현하거나 이용할 시 필요한 시스템 기술 정보를 지속적으로 확보할 수 있도록 해준다(박희진, 2020). 이벤트 개체는 하나 이상의 객체에서 작업한 활동에 대한 정보를 제공하는 개체이다. 이관, 입수 등과 같이 기록관리 행위가 이벤트 개체에 해당되며, 이벤트가 발생하는 시기만 그 활동에 관한 정보를 메타데이터로 획득할 수 있어야 한다(장보성, 2011). 행위주체자 개체는 데이터 객체의 생애주기에서 보존과 관련된 활동과 권한과 관련 있는 행위를 한 주체에 관한 정보이며, 개인이나 기관 이외의 시스템 특성에 관한 정보도 이에 포함될 수 있다(박옥남, 2012). 권한 개체는 저작권, 지적 재산권법, 협약에 따라 행위 주체에 적용되는 권한에 대한 정보이다. 즉, 아카이브에서 소장하고 있는 객체에 대해 아카이브가 어떤 권한과 허가를 받고 수행할 수 있는지에 대한 정보이다.



[그림 2] PREMIS 3.0 모델 (출처: Library Congress, 2016; 박희진, 2020)

PREMIS는 5개의 개체를 통해 보존활동에 필요하다고 판단한 다양한 정보를 정리하고 체계화시킨 메타데이터 요소 사전이다. 각 5개 개체에 대한 다양한 이름, 하위요소, 정의, 데이터 제약사항, 객체 범주, 적용가능성, 생성/유지 안내, 사용안내 정보를 정의하고 있다. 또한 PREMIS에서는 요소(element)라는 표현을 하지 않고 상위요소와 같은 의미인 의미 단위(semantic unit, 1단계)와 최소 2단계에서 최대 4단계까지의 하위요소를 기술할 수 있도록 하는 의미 구성 요소(semantic component)라는 표현을 사용한다([표 8 참조]). 의미 단위와 의미 구성 요소라는 표현을 사용함으로써 디지털 객체에 대한 파일뿐만 아니라 표현물과 비트스트림 수준까지 기술 할 수 있다(박희진, 2020). PREMIS 특징으로는 디지털 정보자원을 장기 보존하는데 있어 생존가능성, 실행가능성, 이해가능성, 진본성, 신뢰성을 지원할 수 있도록 관리, 구조, 기술, 출처와 맥락정보를 설명해 주는 요소가 존재한다는 것이다(강민정, 2021). 다음 [표 5]는 PREMIS 데이터 사전 3.0의 의미 단위와 의미 구성 요소를 정리한 것이다.

[표 5] PREMIS 의미 단위와 의미 구성 요소

개체	의미 단위	의미 구성 요소		
	1단계	2단계	3단계	4단계
Object	object identifier	object identifier type		
		object identifier value		
	object category			
	preservation level	preservation level type		
		preservation level value		
		preservation level role		
		preservation level rational		
		preservation level date assigned		
	significant	significant		

properties	properties type			
	significant properties value			
	significant properties extension			
object characteristic	composition level			
	fixity	message digest algorithm		
		message digest		
		message digest originator		
	size			
	format	format designation	format name	format version
		format registry	format registry name	
			format registry key	
			format registry role	
	format note			
	creating application	creating application name		
		creating application version		
		date created by application		
		creating application extension		
	inhibitors	inhibitor type		
		inhibitor target		
		inhibitor key		
object characteristics extension				
original name				
storage	content location	content location type		

			content location value	
		storage medium		
signature information	signature		signature encoding	
			signer	
			signature method	
			signature value	
			signature validation rules	
			signature properties	
	signature information extension			
environment function	environment function type			
	environment function level			
environment designation	environment name			
	environment version			
	environment origin			
	environment designation note			
	environment designation extension			
environment registry	environment registry name			
	environment registry key			
	environment registry role			
environment extension				
relationship	relationship type			
	relationship subtype			
	related object identifier	related object identifier type		

			related object identifier value		
			related object sequence		
		related event identifier	related event identifier type		
			related event identifier value		
			related event sequence		
		related environment purpose			
		related environment characteristic			
	linking event identifier	linking event identifier type			
		linking event identifier value			
	linking rights statement identifier	linking rights statement identifier type			
		linking rights statement identifier value			
	event	event identifier	event identifier type		
			event identifier value		
		event type			
event date time		event detail			
		event detail extension			
event outcome information		event outcome			
		event outcome detail	event outcome detail note		
			event outcome detail extension		
linking agent identifier		linking agent identifier type			
		linking agent identifier value			

		linking agent role		
	linking object identifier	linking object identifier type		
		linking object identifier value		
		linking object identifier role		
agent	agent identifier	agent identifier type		
		agent identifier value		
	agent name			
	agent type			
	agent version			
	agent note			
	agent extension			
	linking event identifier	linking event identifier type		
		linking event identifier value		
	linking rights statement identifier	linking rights statement identifier type		
		linking rights statement identifier value		
	linking environment identifier	linking environment identifier type		
		linking environment identifier value		
		linking environment role		
rights	rights statement	rights statement identifier	rights statement identifier type	
			rights statement identifier value	
	rights basis			
	copyright	copyright status		

	information	copyright jurisdiction		
		copyright status determination date		
		copyright note		
		copyright documentation identifier	copyright documentation identifier type	
			copyright documentation identifier value	
			copyright documentation identifier role	
		copyright applicable dates	start date	
			end date	
		license information	license documentation identifier	license documentation identifier type
				license documentation identifier value
	license documentation identifier role			
	license terms			
	license note			
	license applicable dates		start date	
			end date	
	statute information	statute jurisdiction		
		statute citation		
		statute information determination date		
		statute note		
		statute documentation identifier	statute documentation identifier type	
statute				

				documentation identifier value
				statute documentation identifier role
			statute applicable dates	start date
				end date
		other rights information	other rights information identifier	other rights information identifier type
				other rights information identifier value
				other rights information identifier role
			other rights basis	
			other rights applicable dates	start date
				end date
			other rights note	
		rights granted	act	
			restriction	
			term of grant	start date
				end date
			term of restriction	start date
			end date	
		rights granted note		
		linking object identifier	linking object identifier type	
			linking object identifier value	
			linking object role	
		liking agent identifier	liking agent identifier type	
			liking agent identifier value	
liking agent role				
rights extension				

2) ISAD(G)

국제기록관리협의회(International Council on Archives, ICA)는 기록물 기술에 관한 국제 표준 제정의 필요성을 인식하고 1990년 기술표준위원회(Ad Hoc Commission on Descriptive Standard, ICA/CDS)를 구성하였다. ICA/CDS는 미국의 APPM(Archives, Personal Papers, and Manuscripts), 영국의 MAD(Manual of Archival Description), 캐나다의 RAD(Rules for Archival Description)와 같은 국가별 표준들을 참고하여 1994년 ISAD(G)(General International Standard Archival Description) 초판을 출시하였다. 2000년에는 ISAD(G) 제2판을 확정 짓고 공포하였다. ICA(2000)에 따르면 ISAD(G)를 통해 보존기록의 기술을 위한 일반적인 가이드를 제공하고 기록의 형태나 매체에 관계없이 적용될 수 있다고 정의되어 있다.

ISAD(G)의 기록물 기술표준 작성의 목적은 다음과 같다. 첫째, 일관적이고 적절하며, 자명한 기술의 생산을 보장한다. 둘째, 기록 자료에 대한 정보 검색 및 교환을 촉진한다. 셋째, 전거데이터를 공유할 수 있게 한다. 넷째, 각 기관에서 작성한 기술을 수렴하여 통합정보시스템을 만들 수 있다. 이와 같은 목적으로 제정된 ISAD(G)는 국제 표준으로서 다양한 형태의 기록물이 가지고 있는 맥락을 보존하기 위해 다계층 기술(multilevel description) 정리 원칙을 채택하여 광범위하게 활용되고 있다. 다계층 기술 원칙은 가장 큰 기록물의 집단적 단위에서 기록물을 기술하는 것을 시작으로 점점 하위 계층으로 기술하는 것이다(이창민, 2015). ICA에서 밝히고 있는 다계층 기술의 4가지 원칙은 다음과 같다(ICA, 2000).

첫째, 일반적인 수준에서 구체적인 수준으로 기술해야 한다. 이는 기록물 군과 기록물 군의 맥락과 계층 구조를 나타내기 위한 것으로 기록물 군의 수준에서는 기록물 군 전체에 대한 정보를 기술해야 하며, 다음 수준에서는 기술되는 부분들의 정보를 제공하고, 포괄적인 것에서 좀 더 세분된 것을 기술해야 한다. 둘째, 해당 기술 수준과 관련된 데이터만을 기술해야 한다. 이는

기술단위의 맥락과 내용을 정확하게 표현하기 위한 것이다. 기술단위가 상위 계층일 경우 그에 대한 상세한 내용은 제공하지 않으며, 기술단위가 부서나 지점 같은 하위 계층일 경우 전체 부서에 대한 관리 이력에 관한 정보를 제공하지 않는다. 셋째, 기술 간의 연결이 이루어져야 한다. 이는 기록물의 계층 구조 속에서 기술단위의 명확한 위치를 나타내기 위한 것으로, 하위 계층의 기술을 상위 계층의 기술에 연결해야 한다. 넷째, 같은 정보를 반복해서 기술하지 않아야 한다. 이는 계층적으로 관련된 기록에 관한 기술의 중복을 피하기 위한 것으로 가장 적절한 수준에서 정보를 제공해야 한다.

이처럼 다계층 기술을 원칙으로 하는 ISAD(G)는 여러 기관에서 참고로 하는 기본이 되는 표준이다(유종연 2020). 기술요소는 7개의 영역과 26개 요소로 구성되어 있다. 7개의 영역은 식별, 배경, 내용 및 구조, 접근 및 이용조건, 연관자료, 주기, 기술통제 영역으로 이루어져있다. 식별 영역은 기술단위를 식별하는데 필수적인 정보로 ‘참조코드’, ‘제목’, ‘날짜’, ‘기술 계층’, ‘기술단위의 규모와 유형’의 요소가 존재한다. 배경 영역은 기술단위의 출처와 보관에 관한 정보로 ‘생산자명’, ‘행정연혁/개인이력’, ‘기록물 보관 이력’, ‘수집. 이관의 직접적 출처’ 요소가 존재한다. 내용 및 구조 영역은 기술단위의 주제와 배열에 관한 정보로 ‘범위와 내용’, ‘평가/폐기/처리일정 정보’, ‘추가’, ‘정리체계’ 요소가 존재한다. 접근 및 이용조건은 기술단위의 이용가능성에 관한 정보로 ‘접근환경’, ‘복제조건’, ‘언어/문자’, ‘물리적 특성과 기술적 요구사항’, ‘검색도구’ 요소가 존재한다. 연관자료 영역은 기술단위와 관련되는 중요한 자료에 관한 정보로 ‘원본의 존재와 위치’, ‘사본의 존재와 위치’, ‘관련 기술단위’, ‘출판 주기 요소’가 존재한다. 주기 영역은 모든 영역에 기술될 수 없는 정보와 특수 정보를 기술하는 영역이다. 기술통제 영역은 서지정보가 누가 언제 어떻게 작성된 것인가에 대한 정보로 ‘아키비스트 주기’, ‘규칙과 협약’, ‘기술일자 요소’가 존재한다. 이 중 ‘참조코드’, ‘제목’, ‘날짜’, ‘기술계층’, ‘기술단위의 규모와 유형’, ‘생산자명’이 필수요소이다(ICA, 2000; 양혜정, 2009). 이상의 내용을 표로 정리하면 [표 6]과 같다.

[표 6] ISAD(G) 영역과 기술요소

* : 필수요소

영역	요소
식별 (Identity statement)	참조코드*(Reference code)
	제목*(Title)
	날짜*(Date(s))
	기술 계층*(Level of description)
	기술단위의 규모와 유형(수량, 부피, 크기)* (Extent and medium of the unit of description(quantity, bulk, or size))
배경 (Context)	생산자명*(Name of creator(s))
	행정연혁/개인이력 (Administrative / Biographical history)
	기록물 보관 이력(Archival history)
	수집/이관의 직접적 출처 (Immediate source of acquisition of information)
내용 및 구조 (Content and Structure)	범위와 내용(Scope and content)
	평가, 폐기, 처리일정 정보 (Appraisal, destruction and scheduling information)
	추가(Accruals)
	정리체계(System of arrangement)
접근 및 이용조건 (Condition of Access)	접근환경(Condition governing access)
	복제조건(Conditions governing reproduction)
	언어/문자(Language/scripts of material)
	물리적 특성과 기술적 요구사항 (Physical characteristics and technical requirements)
	검색도구(Finding aids)
연관자료 (Allied materials)	원본의 존재와 위치 (Existence and location of originals)
	사본의 존재와 위치 (Existence and location of copies)
	관련 기술단위(Related units of description)
	출판주기(Publication note)
주기 (Notes)	주기(Note)
기술통제 (Description Control)	아키비스트 주기(Archivist's Note)
	규칙과 협약(Rules of Conventions)
	기술일자(Date(s) of descriptions)

3) PBCore

미국의 공영 방송 커뮤니티는 생산자와 지역 방송국이 미디어를 원활하게 공유하고 관리와 보존 할 수 있도록 PBCore(Public Broadcasting Metadata Dictionary Project)를 개발하였다. PBCore는 특정 방송 영상클립이나 전체 편집된 방송 영상 제작물을 포함하여 공공 방송과 관련된 미디어 자산의 기술과 교환을 지원하는 메타데이터 구조이다(봉지현, 2019). 또한 표준화된 메타데이터 요소 집합인 더블린 코어를 토대로 방송 영상에 적합하게 개발되었기 때문에 파일 크기나 화질 및 해상도 등과 같은 동영상과 관련된 요소를 충분히 반영하고 있다(박인우, 2018). 때문에 시청각 콘텐츠에 대한 정보 기록을 쉽게 공유할 수 있으며, 미디어 자산 및 컬렉션에 대한 정보를 조직과 미디어 시스템 간에 교환할 수 있다.

PBCore는 2005년에 v1.0 발표 후 2010년까지 3번의 업데이트를 거쳐 2011년에 v2.0으로 개정되었다. 2015년에는 v2.1로 개정되어 2022년인 현재 까지 이어지고 있다. v1.0과 v2.0의 가장 큰 변화는 모든 요소가 컨테이너에 속해있는 v1.0과 다르게 v2.0에는 ‘pbcore Asset Type’과 ‘pbcore Asset Date’, ‘pbcore Identifier’, ‘pbcore Title’, ‘pbcore Subject’ 등의 요소들이 상위 레벨로 구성되어 있다는 것이다. PBCore v2.1은 4개 영역, 19개의 상위요소, 36개의 하위요소, 21개의 세부요소로 구성된다. 4개의 영역은 ‘pbcore Description Documents’, ‘pbcore Annotation’, ‘pbcore Part’, ‘pbcore Extension’이다.

‘pbcore Description Documents’는 미디어 자료의 실제적인 지적 내용을 기술하는 포괄적 주제, 제목, 장르 등의 요소가 있다. 또한 미디어 자산의 제작, 이용과 관련하여 제작자, 기여자, 발행자, 권한 관련 요소들을 통하여 저작물의 사용에 대한 권리정보를 제공한다. 해당 영역의 ‘instantiation’ 요소에는 물리적 형식 또는 디지털화 형식의 미디어 자료의 특성을 식별하는 정보와 보존 및 관리하는 데 필요한 물리적 기술정보들로 구성되어 있다. 즉, 포맷, 미디어 유형, 파일 사이즈, 음성 및 영상 비율, 재생시간 등의 요소들을 통하여 미디어 객체에 대해 다양한 관점에서 설명할 수 있다. ‘pbcore

Annotation’과 ‘pbcore Part’ 영역을 통해서 는 명확하게 기술되지 않은 요소 들을 기술할 수 있다. ‘pbcore extension’ 영역은 다른 표준을 임베드(embed) 하거나 다른 표준의 요소의 값을 묶어서 표현할 수 있는 요소로 구성되어 있 다(박인우, 2018).

PBCore는 영상물의 지적 콘텐츠 및 지적 자산에 관한 요소를 통하여 영 상물과 관련한 서지정보, 내용, 저작권 등에 대해 정보를 제공한다. 또한 ‘instantiation’ 요소를 통하여 영상물의 물리적 형식과 디지털화 형식을 포함 하여 다양한 매체로 생산되는 자료를 관리할 수 있다. 즉, 물리적 기술 요소 를 통한 자료의 보존 및 관리적 측면을 강조하였다(봉지현, 2019). 다음 [표 7]은 PBcore의 메타데이터 요소 표이다.

[표 7] PBcore 메타데이터 표준 요소

*: 필수 요소

영역	요소	하위요소	
pbcore Description Documents	pbcore Asset Type		
	pbcore Asset date		
	pbcore Identifier*		
	pbcore Title*		
	pbcore Subject		
	pbcore Description*		
	pbcore Genre		
	pbcore Relation		pbcore relation Type
			pbcore relation Identifier
	pbcore Coverage		pbcore coverage
			pbcore coverage Type
	pbcore Audience Level		
	pbcore Audience Rating		
	pbcore Creator		creator
			creatorRole
	pbcore Contributor		contributor
			contributor Role
	pbcore Publisher		publisher
publisher Role			
pbcore RightsSummary		rights summary	
		rights Link	
		rights Embedded	

		Instantiation Identifier		
		Instantiation Date		
		Instantiation Dimensions		
		Instantiation Physical		
		Instantiation Digital		
		Instantiation Standard		
		Instantiation Location		
		Instantiation MediaType		
		Instantiation Generations		
		Instantiation FileSize		
		Instantiation TimeStart		
		Instantiation Duration		
		Instantiation DataRate		
		Instantiation Colors		
		Instantiation Tracks		
		Instantiation Channel Configuration		
		Instantiation Language		
		Instantiation Alternative Modes		
		pbcore Instantiation	Instantiation Essence Track	essence track type
				essence track identifier
				essence track standard
				essence track encoding
				essence track datarate
				essence track framerate
				essence track playbackspeed
				essence track samplingrate
				essence track bitdepth
				essence track framesize
essence track aspect ratio				
essence track timestart				
essence track duration				
essence track				

			language
			essence track annotation
			essence track extension
		Instantiation Relation	relation type
			relation identifier
		Instantiation Rights	rights summary
			rights Link
			rights Embedded
Instantiation annotation			
Instantiation part			
Instantiation extension			
pbcore Annotation			
pbcore Part			
pbcore Extensions	extension Wrap	extension Element	
		extension Value	
		extension Authority used	
	extension embedded		

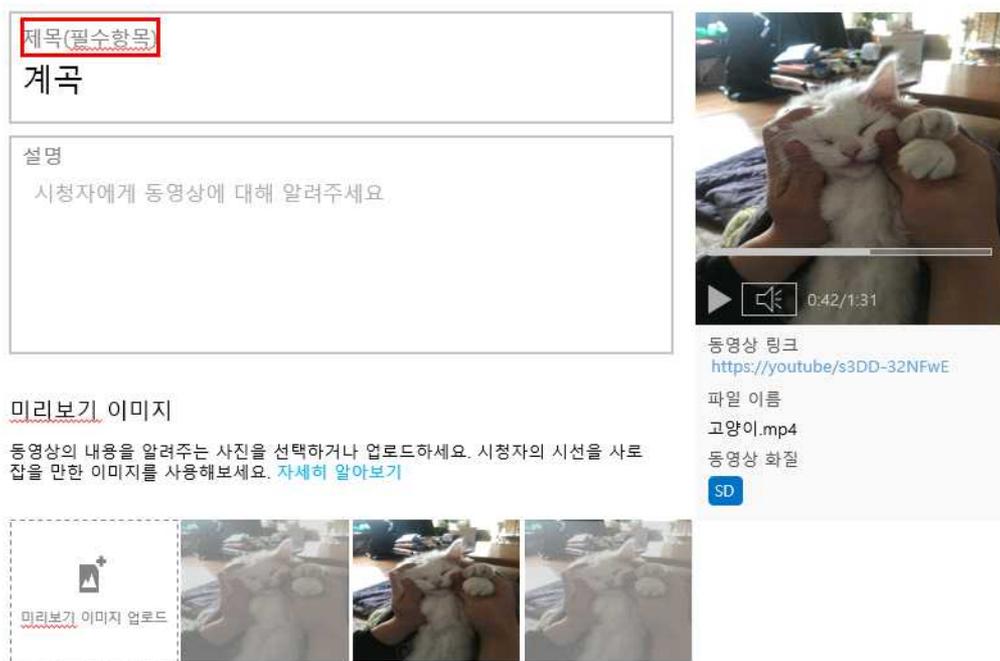
3.1.3 유튜브 메타데이터 요소

유튜브 동영상을 관리 및 보존하기 위한 메타데이터는 아직 존재하지 않는다. 본 연구는 유튜브 동영상 아카이빙을 위한 메타데이터 핵심 요소를 제안하는 것을 목적으로 한다. 이를 위해 우선 아카이빙할 대상의 특성을 파악해야 한다. 이에 따라 본 연구에서는 유튜브에 동영상을 등록하는 과정에서 발생하는 다양한 정보와 등록 후에 자동으로 추출되는 정보 및 이용자 참여 정보를 ‘유튜브 메타데이터(YouTube metadata)’로 지칭하였다.

유튜브에 동영상을 등록하는 과정은 1)동영상 세부정보, 2)동영상 요소, 3)공개 상태의 3단계로 구성되어 있다. 또한 등록 후에 생성되는 정보는 4)자동추출 정보 및 이용자 피드백 정보로 나타내어 본 연구에서 지칭하는 유튜브 메타데이터는 4가지 영역으로 구분하였다.

1) 동영상 세부정보

동영상 세부정보는 유튜브에 등록할 동영상 내용과 동영상 내용에 대한 설정 정보, 동영상을 등록한 후 활용할 수 있는 정보로 구성되어 있다. 이는 [그림 3]~[그림 6]에서 확인할 수 있다. [그림 3]에서와 같이 유튜브 내에서 필수적으로 입력해야 한다고 밝히고 있는 요소는 ‘제목’이다. ‘설명’ 요소는 동영상 내용을 텍스트로 입력하여 시청자들이 이해하기 쉽도록 짧고 간단하게 나타낼 수 있으며, ‘해시태그(#)’를 사용할 수 있다. 또한 소위 ‘썸네일(Thumbnail)’이라고 불리는 ‘미리보기 이미지’가 있으며, 이외에 ‘동영상 링크’, ‘파일 이름’, ‘파일 형식’ 요소가 존재한다.



[그림 3] 유튜브 메타데이터 - 동영상 세부정보 1

[그림 4]에서는 사용자가 기관 유튜브 페이지에서 영상을 더욱 빨리 발견할 수 있도록 주제별로 묶어주는 기능을 하는 ‘재생 목록’ 설정이 있다. ‘시청자 층’ 요소는 [그림 3]의 ‘제목’과 마찬가지로 필수적으로 입력해야 하는 정보로 동영상의 내용이 아동용 동영상인지의 여부를 선택하는 것이다. 이는 유튜브

에서 동영상을 등록하는 모두가 아동 온라인 개인정보 보호법(COPPA) 및 기타 법률을 준수해야 할 법적인 의무가 존재하기 때문에 반드시 지정해야 하는 요소이다. 등록할 동영상의 시청자 층이 아동용 동영상이 아닌 18세 이상 이상일 경우 ‘연령 제한(고급) 요소’를 통해 설정할 수 있다. 동영상에 대한 추가 설정은 ‘자세히 보기(유료 프로모션, 태그, 자막)’를 통해 가능하다.

재생목록

동영상을 1개 이상의 재생목록에 추가하세요. 시청자가 내 콘텐츠를 더 빨리 발견하는데 도움이 될 수 있습니다. [자세히 알아보기](#)

선택 ▼

시청자층

이 동영상이 아동용이 아니라고 설정됨 [크리에이터가 설정함](#)

모든 크리에이터는 위치에 상관없이 아동 온라인 개인정보 보호법(COPPA) 및 기타 법률을 준수해야 할 법적인 의무가 있습니다. 아동용 동영상인지 여부는 크리에이터가 지정해야 합니다. [아동용 콘텐츠란 무엇인가요?](#)

 아동용 동영상에서는 개인 맞춤 광고 및 알림 등의 기능을 사용할 수 없습니다. 크리에이터가 아동용으로 설정한 동영상은 다른 아동용 동영상과 함께 추천될 가능성이 높습니다. [자세히 알아보기](#)

- 예, 아동용입니다.
- 아니요, 아동용이 아닙니다.

▲ 연령제한 (고급)

동영상을 성인만 시청할 수 있도록 제한하고 싶으신가요?

연령 제한 동영상은 Youtube의 특정 영역에 표시되지 않습니다. 이러한 동영상은 광고 수익 창출이 제한되거나 배제될 수 있습니다. [자세히 알아보기](#)

- 예, 동영상 시청자를 만 18세 이상으로 제한하겠습니다.
- 아니요, 동영상 시청자를 만 18세 이상으로 제한하지 않겠습니다.

[자세히 보기](#)

유료 프로모션, 태그, 자막 등

[그림 4] 유튜브 메타데이터 - 동영상 세부정보 2

[그림 5]는 ‘자세히 보기(유료 프로모션, 태그, 자막)’에 대한 부분이다. ‘유료 프로모션’은 영상 제작의 목적이 직접/간접적인 광고의 경우 설정해야 하는 부분이다. 이외에 ‘자동 캡터’ 설정을 통해 영상 내용에 대한 목차를 영상을 보기 전 확인할 수 있고, ‘태그’ 기능을 통해 검색하면서 생기는 오타를 자동으로 수정해 검색시켜줄 수 있다.

유료 프로모션

제3자로부터 어떤 형태로든 동영상을 만드는데 대가를 받았다면 [Youtube](#)에 알려야 합니다. [Youtube](#)는 시청자에게 동영상에 유료 프로모션이 포함되어 있음을 알리는 메시지를 표시합니다.

동영상에 간접 광고, [스폰서십](#), [보증광고와](#) 같은 유료 프로모션이 포함되어 있음

이 체크박스를 선택하면 유료 프로모션이 [Youtube](#) 광고 정책 및 관련 법규와 규정을 준수한다고 확인하는 것입니다.

자동 캡터

캡터를 사용하면 동영상을 더 쉽게 시청할 수 있습니다. 언제든지 동영상 설명에서 자체적으로 캡터를 만들어 자동 캡터를 덮어쓸 수 있습니다. [자세히 알아보기](#)

자동 캡터 허용(가능한 경우)

태그

태그는 동영상의 콘텐츠에 일반적으로 맞춤법이 틀리는 단어가 있을 경우 유용합니다. 그 외에 시청자가 동영상을 찾는 데 있어 태그가 하는 역할은 제한적입니다. [자세히 알아보기](#)

태그 추가

각 태그의 뒤에 심표를 입력하세요.

[그림 5] 유튜브 메타데이터 - 동영상 세부정보 3

[그림 6]은 ‘언어 및 자막 면제 인증서’, ‘녹화 날짜 및 동영상 위치’, ‘라이선스’, ‘퍼가기 허용’, ‘Shorts 동영상 샘플링 허용 여부’, ‘카테고리’, ‘댓글 및 평가’가 있다. ‘언어 및 자막 면제 인증서’에는 영상 속 언어와 자막 면제 인증서의 여부를 설정할 수 있다. ‘녹화 날짜 및 동영상 위치’에서는 동영상을 촬영한 날짜와 촬영 장소에 대해 설정할 수 있다. ‘라이선스’에서는 표준 YouTube 라이선스와 크리에이티브 커먼즈 중 해당되는 저작권에 대해 설정할 수 있다. ‘퍼가기 허용’은 동영상 공유 여부를 설정할 수 있는 기능이다. ‘Shorts 동영상 샘플링’은 등록된 동영상의 일부를 Shorts 동영상으로 생산 가능한지 여부를 설정할 수 있다. ‘카테고리’는 유튜브에서 구분한 15개의 주제(게임, 비영리, 스포츠 등) 중 적합한 주제를 설정할 수 있는 기능이다. ‘댓글 및 평가’는 이용자들이 작성한 댓글에 대한 공개 여부 및 댓글 작성을 금지하는 설정 또는 이용자들이 누른 좋아요 수의 표시 여부를 설정할 수 기능이다.

언어 및 자막 면제 인증서

동영상 언어와 자막 면제 인증서(필요한 경우)를 선택하세요.

동영상 언어 선택	자막 면제 인증서 ? 없음
제목 및 설명 언어 선택	

다른 언어를 관리하려면 [자막](#)으로 이동하세요.

복화 날짜 및 위치

동영상을 촬영한 시기와 위치를 추가합니다. 시청자는 위치별로 동영상을 검색할 수 있습니다.

복화 날짜 2020. 7. 25.	동영상 위치 없음
-----------------------	--------------

라이선스

[라이선스 유형](#)에 대해 알아보세요.

라이선스 표준 YouTube 라이선스

- 퍼가기 허용 ?
- 구독 피드에 게시하고 구독자에게 알림 전송

Shorts 동영상 샘플링

다른 사용자가 이 동영상의 일부를 사용하여 shorts 동영상을 만들 수 있도록 허용합니다. 사용 중지할 경우 동영상의 콘텐츠가 샘플링된 모든 shorts 동영상이 영구적으로 삭제됩니다. 일부 동영상은 선택 해제할 수 없습니다. [자세히 알아보기](#)

- 다른 사용자가 이 콘텐츠를 샘플링하도록 허용

카테고리

시청자가 보다 쉽게 찾을 수 있도록 카테고리에 동영상을 추가하세요.

스포츠

댓글 및 평가

댓글 표시 여부 및 방법을 선택하세요.

- 모든 댓글 허용
- 부정적 댓글 수 있는 댓글은 검토를 위해 보류 ?
 - 검토 기준 높이기 실험
- 검토를 위해 모든 댓글 보류
- 댓글 사용 안함

정렬 기준 인기순

- 동영상에 좋아요를 표시한 시청자 수 표시

[그림 6] 유튜브 메타데이터 - 동영상 세부정보 4

2) 동영상 요소

동영상 요소 단계에서는 자막추가 설정을 비롯하여 해당 동영상과 연관된 다른 동영상 및 채널을 홍보할 수 있는 기능과 동영상의 공개 상태 설정으로 이루어져 있다. 아래 [그림 7]~[그림 9]는 동영상 요소의 ‘자막 추가’, ‘최종 화면 추가’, ‘카드 추가’에 해당한다.

[그림 7]과 같이 ‘자막’ 기능을 추가함으로써 청각이 불편한 이용자 또는 다른 국적의 이용자가 동영상의 내용을 이해하는데 수월해질 수 있다. 자막을 추가하기 위한 방법으로는 ‘파일 업로드’와 ‘자동 동기화 및 직접 입력’이 있는데, ‘파일 업로드’ 방법에서는 ‘자막 파일 유형’을 선택해야 한다. 단, 파일 형식은 반드시 ‘.txt’ 형식에만 적용할 수 있다. ‘자동 동기화 및 직접 입력’에서는 ‘스크립트’를 통하여 자막을 입력할 수 있으며, 원하는 타이밍에 자막을 노출시킬 수 있다.



[그림 7] 유튜브 메타데이터 - 동영상 요소 1

[그림 8]의 ‘최종화면 추가’는 동영상이 종료되었을 때 최근 업로드된 동영상, 자동 추천 동영상, 추천 동영상 직접 선택 중에 하나를 설정하여 해당 동영상의 썸네일을 노출시킬 수 있다. 또한 자신 또는 타채널을 구독을 할 수 있는 버튼 설정을 통해 유튜브 채널을 홍보할 수도 있다.



[그림 8] 유튜브 메타데이터 - 동영상 요소 2

[그림 9]는 ‘카드 추가’에 대한 설명이다. 이는 동영상 내용과 관련 있는 동영상을 홍보할 수 있는 요소들 존재한다. 시리즈물 영상으로 예시를 들면 1편 영상을 보지 않고 2편 영상을 처음 접한 시청자들에게 1편 영상을 직접 검색하지 않고, 사용자가 동영상 진행 중 오른쪽 상단에 위치한 동영상(재생 목록, 채널) 카드를 눌러 손쉽게 시청할 수 있다. ‘맞춤 메시지’와 ‘티저 텍스트’는 유튜브에 업로드한 제목의 영상이 아닌 직접적인 이해가 가능한 설명으로 연관 동영상 및 재생목록, 채널을 설명할 수 있다.



[그림 9] 유튜브 메타데이터 - 동영상 요소 3

3) 공개 상태

[그림 10]의 공개 상태는 ‘공개’, ‘비공개’, ‘일부공개’, ‘예약’으로 구성되어 있다. ‘공개’는 모든 사람이 영상을 시청할 수 있으며, ‘비공개’는 동영상을 등록한 사람이 선택한 사람만 시청할 수 있다. ‘일부 공개’는 동영상 내용 중 일부를 공개하는 것이 아닌 동영상의 링크를 알고 있는 사람만 영상을 시청할 수 있다. ‘예약’은 자신이 원하는 날짜 및 시간에 영상을 공개할 수 있는데, ‘(인스턴트)Premieres’ 기능은 영상을 공개하기 3분 전에 사람들이 입장하여 실시간 댓글 기능을 사용하면서 대기를 할 수 있는 기능이 있다.

The image shows two panels of YouTube video settings. The top panel is titled '저장 또는 게시' (Save or publish) and has four radio button options: '공개' (Public), '비공개' (Private), '일부 공개' (Unlisted), and '공개' (Public). Below these are checkboxes for '인스턴트 Premieres 동영상으로 설정' (Set as Instant Premiere video) and 'Premieres 동영상으로 설정' (Set as Premiere video). The bottom panel is titled '예약' (Schedule) and has a radio button selected for '예약' (Schedule). It includes a date selector (2021. 10. 29.), a time selector (오전 12:00), and a '시간대' (Time zone) dropdown. Below these are checkboxes for 'Premieres 동영상으로 설정' (Set as Premiere video) and a note: '동영상은 게시될 때까지 비공개 상태로 유지됩니다.' (The video will remain private until it is published).

[그림 10] 유튜브 메타데이터 - 공개 상태

4) 자동 추출 정보 및 이용자 피드백 정보

자동 추출 정보 및 이용자 피드백 정보 단계는 유튜브에 동영상을 등록한 후 자동으로 생성되는 다양한 정보가 포함되며 다음 [그림 11]과 같다. ‘Video ID’, ‘프레임크기’, ‘화면비율’, ‘색상’ 등이 표시되는 부분은 유튜브 영

상의 '전문 통계'에 해당하는 부분으로, 이를 통해 동영상 재생의 문제가 있을 경우, 화면에 나타나는 통계 수치를 보고 문제 원인을 발견하거나 문제를 해결할 수도 있다. 또한 '동영상길이'와 '조회수', 동영상을 등록한 유튜브 '채널명'에 대한 정보와 해당 채널의 '구독자수'가 존재하고, 이용자 피드백 정보에는 '좋아요수', '싫어요수', '댓글수', '댓글내용', '대댓글내용'이 존재한다.



[그림 11] 유튜브 메타데이터 - 자동 추출 정보 및 이용자 피드백 정보

유튜브 메타데이터 요소는 이상의 [그림 3]~[그림 11]에서 살펴본 것과 같이 1)동영상 세부정보, 2)동영상 요소, 3)공개 상태, 4)자동 추출 정보 및 이용자 피드백 정보로 구성되어 있다. 위의 내용에 기반하여 유튜브 메타데이터 요소를 정리하면 [표 8]과 같다.

[표 8] 유튜브 메타데이터 요소

NO	범주	요소명	설명
1	동영상 세부정보	제목	동영상의 제목
2		설명	동영상의 내용 및 키워드를 설명하는 부분
3		동영상링크	동영상의 URL 주소
4		파일이름	동영상 및 이미지, 자막 파일 이름
5		파일형식	동영상 및 이미지, 자막 파일 형식
6		미리보기 이미지	썸네일(Thumbnail) 이미지 사진
7		재생목록	동영상 목록을 분류해놓은 카테고리
8		아동용유무	동영상의 내용이 아동용인지에 대한 여부
9		연령제한	동영상을 시청할 수 있는 연령을 18세 이상으로 설정할지에 대한 여부
10		유료프로모션	유료 PPL, 스폰서십, 보증광고 또는 시청자에게 공개해야 하는 기타 콘텐츠를 포함하는지 유무
11		자동챕터허용	동영상에 구간이 나누어지고 각 구간의 미리보기가 생성되는 것. 챕터가 있으면 동영상 각 구간에 정보와 배경을 추가할 수 있어 동영상의 여러 다른 구간을 쉽게 다시 시청
12		태그	사용자가 콘텐츠를 검색하는 데 도움이 되도록 동영상에 추가하는 설명 키워드
13		동영상언어	동영상에 나오는 언어 종류
14		자막면제 인증서	미국 연방 통신 위원회 규정에 따라 미국 TV에서 2012년 9월 30일 이후에 방송된 특정 콘텐츠의 경우 YouTube에 자막을 제공해야함. 자막이 필요하지 않은 콘텐츠의 경우 인증을 선택하여 YouTube에 알림.
15		녹화날짜	동영상을 녹화한 날짜
16		동영상위치	동영상 속 장소의 위치
17		라이선스	동영상의 저작권 표시 (표준 YouTube라이선스 or 크리에이티브 커먼즈 중 선택)
18		배포	동영상을 모든 플랫폼에 제공할지 아니면 수

			의 창출 플랫폼에만 제공할지 선택
19		퍼가기허용	다른 사람의 사이트로 내 동영상을 퍼가도록 허용 여부
20		알림전송유무	동영상을 업로드할 시, 알림설정을 한 구독자에게 알림 전송에 대한 여부
21		Shorts 동영상 샘플링허용유무	다른 사용자가 동영상 일부를 shorts 동영상으로 만들 수 있도록 허용
22		카테고리	시청자가 찾을 수 있도록 유튜브에서 주제별로 분류한 것
23		댓글공개상태	모든 댓글 허용 or 부적절한 경우 보류 or 모든 댓글 보류 or 댓글 사용 안함 중 선택
24		댓글정렬기준	인기순 or 최신순 중 선택
25		시청자수표시유무	업로드된 동영상에 좋아요 및 싫어요를 누른 시청자의 수 표시 유무
26	동영상 요소	자막추가	동영상 자막 설정 유무 (파일 업로드 or 자동 동기화, 직접 입력 중 선택)
27		자막파일유형	타이밍 포함 or 제외 중 선택 * '제외' 선택 시 .txt 유형만 가능
28		시간대	자막을 입력했을 시, 자막이 나타나는 시간대 (예시: 0:16 ~ 0:53)
29		스크립트	자막을 입력 했을 시, 자막의 대본
30		최종화면추가	동영상 길이가 25초 이상일 때부터 사용할 가능한 기능으로 동영상 내용이 완전히 끝나고 홍보 및 제공할 동영상 및 구독 프로필 위치 유형 선택
31		최종화면 동영상요소	최근 업로드된 동영상 or 시청자 맞춤 or 특정 동영상 선택 중 선택
32		최종화면 재생목록요소	최종화면에서 사용할 재생목록 선택
33		최종화면 구독요소	최종화면에서 사용할 구독 홍보 선택
34		최종화면 채널요소	최종화면에서 사용할 특정 채널 선택
35			카드추가
36		맞춤메시지	'최종화면' 및 '카드'를 사용할 시 카드의 정보를 시청자들에게 보여줄 메시지

37		티저텍스트	'카드'를 사용할 시 카드의 제목을 시청자들에게 보여줄 텍스트
38	공개 상태	공개 상태	동영상의 공개 및 예약 여부. (비공개 or 일부공개 or 공개 or 예약 중 선택)
39		인스턴트 premieres 설정여부	동영상 예약이 완료되면 공유 가능한 보기 페이지 링크가 생성되어 Premiere가 시작될 때까지 그곳에서 팬들과 소통이 가능.
40		예약날짜 및시간	예약으로 설정했을 경우 공개할 날짜 및 시간
41	자동 추출 정보 및 이용자 피드백 정보	videoID /sCPN	비디오 ID와 특정 재생 인스턴스의 코드
42		Viewport/frames	비디오 플레이어의 해상도(비디오의 해상도가 아님)와 현재 있는 프레임.
43		Current/OptimalRes	비디오의 현재 해상도 대 원래 해상도
44		Volume/Normalized	보고 있는 비디오의 현재 볼륨과 오디오가 "표준화"된 정도. 괄호는 동영상이 YouTube 기준과 비교하여 얼마나 큰지를 나타냄.
45		Codecs	미디어의 비디오 및 오디오 압축 코덱
46		Color	현재 보고 있는 콘텐츠의 색상 범위.
47		Connection Speed	장치의 네트워크 연결 속도
48		Network Activity	YouTube 서버에서 기기로 데이터를 전송하는 활동. 비디오의 해상도가 높을수록 네트워크 활동(대역폭이라고도 함.)이 높아짐.
49		Buffer Health	연결 속도 저하로 인한 재생 끊김 및 끊김을 방지하기 위해 YouTube가 동영상을 버퍼링할 수 있는 정도
50		MysteryText	유튜브 특정 프로그래머 코드
51		동영상길이	동영상의 길이
52		채널명	동영상을 업로드한 채널의 이름
53		구독자수	채널을 구독한 사람의 수
54		조회수	동영상의 조회수
55		업로드날짜	동영상이 업로드 된 날짜
56		좋아요수	동영상의 '좋아요' 개수
57		싫어요수	동영상의 '싫어요' 개수
58		댓글수	동영상의 댓글 개수
59		댓글ID	동영상의 댓글 ID
60	댓글내용	동영상의 댓글 내용	
61	대댓글ID	동영상의 댓글에 달린 대댓글 ID	
62	대댓글내용	동영상의 댓글에 달린 대댓글 내용	

3.2 메타데이터 요소 매핑

메타데이터 요소 매핑에서는 3.1 메타데이터 요소 조사에서 살펴본 메타데이터들을 유튜브 동영상 아카이빙을 위한 메타데이터로 설계하기 위해 취합, 정리한다. 앞선 이론적 배경과 선행연구에서 검토한 것처럼 일반적으로 메타데이터는 설명, 구조, 관리 메타데이터로 세분화된다.

본 연구에서는 유튜브 동영상이 영상, 웹 기록물이라는 관점을 고려하여 안영희, 박옥화(2009), 오상훈, 최영선(2009) 연구에서와 같이 설명, 구조, 관리, 보존 메타데이터로 구분하였다. 다음 [표 9]는 2개의 선행연구에서 구분한 메타데이터의 내용과 예시를 종합하여 재구성한 것이다. [표 9]를 토대로 앞서 살펴본 국내외 표준과 유튜브 메타데이터를 매핑하여 유튜브 동영상 아카이빙 메타데이터를 정리하였다.

[표 9] 아카이빙 메타데이터 내용 및 예시

구분		내용	예시
설명 메타데이터	설명 영역	디지털 자원에 대한 설명으로 서지 정보 일부를 포함하는 자원을 유일하게 식별할 수 있는 기본 정보를 제공	서명, 주제, 기술, 자원, 언어, 범위
	구조 메타데이터	디지털 자원을 표현하기 위한 수단 및 방법에 따라 자원의 구조와 구성에 대한 정보를 제공	대상 유형, 대상 장르, 대상 식별자
관리 메타데이터	관리 영역	웹 아카이빙에서 입수된 자원을 관리하고 아카이빙을 운영하기 위해 요구되는 정보를 제공	아카이빙 등록번호, 날짜, 관리자, 사건 정보
	권리 영역	디지털 자원의 지적 재산권에 관련된 개인, 단체, 기관에 대한 정보를 제공	저작자, 출판사, 지적 재산권 정보
	기술 영역	디지털 자원을 사용하기 위해 요구되는 운영시스템, 소프트웨어, 하드웨어 정보 등 기술적인 환경에 대한 정보를 제공	operating system 정보, 파일 크기, 파일 포맷
보존 메타데이터	출처 영역	자원의 출처에 대한 정보를 제공	정보의 생산자 및 보관 처리, 마이그레이션 이력

참조 영역	디지털 자원의 내용 정보에 대한 이력 정보를 제공	자원의 이력 정보
문맥 정보	디지털 자원 내용 정보의 생산 맥락과 다른 내용 정보들과의 관련에 대한 정보를 제공	관계 정보
인증 영역	해당 디지털 자원의 진본성과 무결성에 대한 정보를 제공	진본성 확인, 무결성 확인

3.2.1 설명 메타데이터

설명 메타데이터는 일반적으로 책의 서지정보와 같은 것이다. 즉, 정보 자원이 수록하고 있는 내용과 형식을 자세하게 설명하여 정보자원을 기술하고 식별하는데 중점을 두는 메타데이터이다(남태우, 이승민, 2018). 이승호(2022)는 설명 메타데이터가 방대한 정보자원들 중에서 어떻게 찾아내고, 찾아낸 자원이 다른 자원과 어떻게 다른지, 이용자 요구에 적합한 자원을 어떻게 선별해 낼 것인지 등을 다룬다고 하였다. 안영희, 박옥화(2009), 오상훈 최영선(2009)의 연구를 종합하여 제시된 설명 메타데이터의 설명 영역은 ‘title’, ‘subject’, ‘description’, ‘source’, ‘language’, ‘coverage’이다. 다음 [표 10]을 통해 국내외 메타데이터 표준과 유튜브 메타데이터의 요소 중 설명 영역에 포함하는 요소를 정리하였다.

[표 10] 설명 메타데이터 요소 비교

설명 MD	요소	NAK 8	PREMIS	ISAD(G)	PBCore	유튜브
설명 영역	title	기록물명	original name	제목	pbcore Title	제목
	subject	주제			pbcore Subject	카테고리
	description	기술		범위와 내용	pbcore Description	설명
	source					
	language			언어	instantiation Language	동영상 언어
	coverage			범위와 내용	pbcore Coverage	

[표 10]을 통해 알 수 있듯이 모든 표준에서 존재하는 요소는 ‘title’ 뿐이었다. NAK 8의 ‘기록물명’, PREMIS의 ‘original name’, ISAD(G)의 ‘제목’, PBCore의 ‘pbcore title’, 유튜브 메타데이터의 ‘제목’이라는 요소가 있었다. PREMIS와 ISAD(G)에서 ‘subject’에 대한 요소가 부재한 것을 제외하고 나머지 표준에서는 모두 매핑되었고, PREMIS에서는 ‘description’ 요소가 부재한 것을 제외하고 나머지 표준에서는 모두 매핑되었다. 모든 표준에서 매핑되지 않은 요소는 ‘source’로 나타났다. ‘language’는 ISAD(G), PBCore, 유튜브에서 각각 ‘언어’, ‘instantiation Language’, ‘동영상 언어’로 매핑되었다. ‘coverage’는 ISAD(G)와 PBCore에서 각각 ‘범위와 내용’, ‘pbcore Coverage’로 매핑되었다.

3.2.2 구조 메타데이터

구조 메타데이터는 기술하고자 하는 대상을 내부적으로 유지하고 있는 구조적인 형태 정보를 관리하기 위한 메타데이터이다(노영희, 2014). 즉, 웹상의 정보자원이 지닌 전체적인 구조를 기술해주며, 정보자원의 특정 개체를 표현해주고, 이를 찾아갈 수 있도록 안내해주는 역할을 한다. 또한 다양한 개체들 간에 내부 관계를 표현하는 동시에 해당 정보자원과 관련된 다른 정보자원 사이의 외부적인 연결 관계를 표현해 줄 수 있다(Harriette Hemmasi, 2002). 남태우, 이승민(2018)은 기관의 목적이나 규모 또는 관리하고자하는 정보자원의 수준에 따라 구조 메타데이터의 복잡성과 요소의 세분성이 다양하게 나타날 수 있다고 하였다. 안영희, 박옥화(2009)의 연구에서 제시된 구조정보는 간단하게 제시되어 있는 반면, 오상훈, 최영선(2009)의 연구에서 제시된 아카이빙 메타데이터의 구조정보는 ‘object type’의 하위요소로 ‘text’, ‘image’, ‘video’, ‘audio’, ‘multimedia’의 요소까지 좀 더 세분하였다.

본 연구의 아카이빙 대상은 유튜브 동영상이기 때문에 동영상의 구조와 관련된 요소가 다양해야 한다고 판단하여 ‘object type’의 하위요소 중 ‘video’와

관련된 요소인 ‘frame dimensions’, ‘duration’, ‘frame rate’, ‘compression’, ‘video encoding structure’, ‘video sound’를 기준으로 조사한 표준들을 매핑하였으며 [표 11]과 같이 정리하였다.

[표 11] 구조 메타데이터 요소 비교

구조 MD	요소	NAK8	PREMIS	ISAD(G)	PBCore	유튜브
구조 영역	object type	유형	object category		instantiation media type	
	frame dimension				instantiation Dimension	
	duration				instantiation Duration	재생시간
	frame rate				Instantiation essence track framerate	viewport/ frames
	compression					codec
	video encoding structure				instantiation essence track encoding	
	video sound					Volume/Normalized
	object genre				pbcore genre	
	object identifier	기록 식별자	object identifier	참조코드	pbcore identifier	동영상 링크, video ID/SCPN

[표 11]을 살펴본 결과 ‘object type’은 NAK 8의 ‘기록 유형’, PREMIS의 ‘object category’, PBCore의 ‘Instantiation media type’과 매핑되었지만 ISAD(G)와 유튜브 메타데이터에는 존재하지 않았다.

다음 동영상의 구조를 표현해주는 요소들 매핑과 관련해서는 PBCore ‘instantiation’의 하위요소와 유튜브 메타데이터에서만 매핑되었다. ‘frame dimension’의 경우 PBCore의 ‘instantiation Dimension’이 매핑되었고, ‘duration’은 PBCore의 ‘instantiation Duration’과 유튜브 메타데이터의 ‘재생 시간’이 매핑되었다. ‘frame rate’는 PBCore의 ‘Instantiation essence track framerate’, 유튜브 메타데이터의 ‘viewport/frames’가 매핑되었다. ‘compression’은 유튜브 메타데이터의 ‘codec’에서만 매핑되었고, ‘video encoding structure’은 PBCore의 ‘instantiation essence track encoding’에서만 매핑되었으며, ‘video sound’는 유튜브 메타데이터의 ‘Volume/Normalized’에서만 매핑되었다. ‘object genre’의 경우 PBCore의 ‘pbcore genre’에서만 매핑되었다. ‘object identifier’ 요소는 모든 표준에서 존재하였다. NAK 8은 ‘기록 식별자’, PREMIS는 ‘object identifier’, ISAD(G)는 ‘참조코드’, PBCore는 ‘pbcore identifier’, YouTube에서는 ‘동영상 링크’와 ‘video ID/SCPN’ 요소가 매핑되었다.

3.2.3 관리 메타데이터

관리 메타데이터는 정보자원을 효율적인 관리를 위한 목적으로 생성되는 메타데이터이다(노영희, 2014). Hodge(2005)는 정보자원의 생성을 비롯하여 정보자원의 관리, 보존, 통제, 접근, 이용에 적용되는 제한사항 등과 같은 정보를 제공함으로써 효율적으로 정보자원의 관리를 할 수 있다고 하였다. 즉, 관리 메타데이터에는 정보자원을 누가 언제 어떻게 생성하였는지에 대한 정보가 포함된다. 뿐만 아니라 정보자원을 누가 관리하고, 이용할 수 있는 사람은 누구이며, 이용에 따른 조건은 어떤 것이 있는지에 관한 정보가 포함된다. 또한 향후 정보자원의 활용을 지원해야 하는 이유로 보존과 관련된 정보를 포함할 수 있고, 정보자원의 효과적인 운용과 관리를 지원하기 위한 목적으로 구축되었기 때문에 기술적인(technical) 정보도 포함된다(남태우, 이승민, 2018).

이에 따라 본 연구는 관리 메타데이터를 권리 영역, 관리 영역, 기술 영역으로 세분하였다. 관리 메타데이터의 각 하위 영역별 매핑 요소는 [표 12]와 같이 정리하였다.

[표 12] 관리 메타데이터 요소 비교

관리 MD	요소	NAK 8	PREMIS	ISAD(G)	PBCore	유튜브
권리 영역	creator	생산자, 생산자 유형, 기관명, 기관코드, 부서명, 부서코드, 개인명, 개인코드	agent, agent identifier, agent name, agent type, agent version, agent note, agent extension	생산자명	pbcore creator	채널명
	publisher				pbcore publisher	
	rights	권한, 비밀, 접근범위, 공개, 공개구분, 비공개사유, 공공저작물 관리, 공공저작물 여부, 공공저작물 자유이용허락표시 유형	rights statement, rights basis, copyright information, license information, other rights information, rights granted	접근 및 이용조건, 규칙과 협약	pbcore Rights Summary, pbcore Audience Level, pbcore Audience Rating	라이선스, 연령 제한, 공개 상태
관리 영역	date	일시, 생산일시, 종료일시, 등록일시, 시행일시	event datetime	날짜, 기술일자	pbcore Asset date	업로드 날짜, 녹화 날짜
	digital archiving					

	save number					
	service level			검색도구		
	management person	관리행위자	agent, agent identifier, agent name, agent type, agent version, agent note, agent extension			
	event	관리이력, 관리유형, 관리설명, 관리일시, 변경요소	event, event identifier, event type, event date time, event outcome information			
기술 영역	file	크기, 용량, 단위	size	기술단위의 규모와 유형	instantiation FileSize	파일 이름, 파일 형식
	object format	포맷, 포맷명, 포맷버전	format			
	operating system		environment function, environment designation, environment registry, environment extension			
	application	생성 어플리케이션명, 생성 어플리케이션버전	creating application			

[표 12]를 통해 관리 메타데이터의 권리 영역, 관리 영역, 기술 영역을 비교해본 결과는 다음과 같다. 우선 권리 영역의 ‘creator’와 ‘rights’에서 모든 표준이 매핑되는 요소가 있었다. ‘creator’는 NAK 8의 상위요소인 ‘생산자’와 그 하위요소인 ‘생산자 유형’, ‘기관명’, ‘기관코드’, ‘개인명’ 등이 매핑되었고, PREMIS의 경우 agent 영역과 그 의미 단위인 ‘agent identifier’, ‘agent name’, ‘agent type’, ‘agent version’, ‘agent note’, ‘agent extension’이 매핑되었다. ISAD(G)는 ‘생산자명’, PBCore는 ‘pbcore creator’, 유튜브 메타데이터는 ‘채널명’이 매핑되었다. 권리 영역의 ‘rights’와 매핑된 표준의 요소는 NAK 8의 ‘권한’과 그 하위요소인 ‘비밀’, ‘접근 범위’, ‘공개’, ‘공공저작물관리’와 ‘공개’의 세부요소인 ‘공개구분’, ‘비공개사유’, ‘공공저작물관리’의 세부요소인 ‘공공저작물 여부’, ‘공공저작물 자유이용허락표시 유형’이 매핑되었다, PREMIS는 ‘rights’ 영역의 의미 단위 ‘rights statement’에서 3번째 요소들인 ‘rights basis’, ‘copyright information’, ‘license information’, ‘other rights information’, ‘rights granted’와 매핑되었다. ISAD(G)는 ‘접근 및 이용조건’, ‘규칙과 협약’, PBCore는 ‘pbcore Rights Summary’, ‘pbcore Audience Level’, ‘pbcore Audience Rating’, 유튜브 메타데이터는 ‘라이선스’, ‘연령제한’, ‘공개 상태’가 매핑되었다.

관리 영역의 경우 ‘date’ 요소만 모든 표준에 매핑되었다. NAK 8의 상위요소 ‘일시’와 그 하위요소인 ‘생산일시’, ‘종료일시’, ‘등록일시’, ‘시행일시’, PREMIS의 ‘event date time’, ISAD(G)의 ‘날짜’, ‘기술일자’, PBCore의 ‘pbcore Asset date’, 유튜브 메타데이터의 ‘업로드 날짜’, ‘녹화날짜’가 매핑되었다. 그 외에 ‘management person’에서는 NAK 8의 ‘관리 이력’의 하위요소인 ‘관리행위자’, PREMIS의 agent 영역과 그 의미 단위인 ‘agent identifier’, ‘agent name’, ‘agent type’, ‘agent version’, ‘agent note’, ‘agent extension’을 제외하고 매핑되는 표준은 없었다. 이어서 ‘event’ 요소에서는 NAK 8의 ‘관리 이력’과 그 하위요소인 ‘관리 유형’, ‘관리 설명’, ‘관리 일시’, ‘변경 요소’, PREMIS의 event 영역과 그 의미 단위인 ‘event identifier’, ‘event type’, ‘event date time’, ‘event outcome information’만 매핑되었다.

기술 영역에서는 ‘file’과 ‘object format’ 순으로 표준들이 매핑되었다.

‘file’의 경우 NAK 8의 ‘크기’와 그 하위요소인 ‘용량’, ‘단위’, PREMIS의 의미 단위 ‘object characteristic’의 의미 구성 요소인 ‘size’, ISAD(G)의 ‘기술단위의 규모와 유형’, PBCore의 ‘instantiation’의 하위요소인 ‘instantiation FileSize’, 유튜브 메타데이터의 ‘파일이름’, ‘파일형식’이 매핑되었다. ‘object format’의 경우 NAK 8의 상위요소 ‘포맷’과 그 하위요소인 ‘포맷명’, ‘포맷버전’, PREMIS의 의미 단위 ‘object characteristic’의 의미 구성 요소인 ‘format’이 매핑되었다. ‘operating system’의 경우 PREMIS의 의미 단위인 ‘environment function’, ‘environment designation’, ‘environment registry’, ‘environment extension’만 매핑되었으며, ‘application’의 경우 NAK 8의 상위요소 ‘포맷’의 하위요소인 ‘생성 어플리케이션명’, ‘생성 어플리케이션버전’, PREMIS의 의미 단위 ‘object characteristic’의 ‘creating application’과 매핑되었다.

3.2.4 보존 메타데이터

박희진(2020)은 디지털 자원이 아카이브에 입수된 이후 자원의 무결성과 이용가능성을 보장하기 위해서 자원 상태변화와 관련된 이력 정보가 메타데이터에 기록되어야 한다고 하였다. 앞선 선행연구에 따라 보존 메타데이터는 출처 영역, 참조 영역, 문맥 영역, 인증 영역으로 구분하였다([표 12] 참조). 출처 영역은 자원의 출처에 대한 정보를 제공하는 정보이다. 참조 영역은 디지털 자원의 내용 정보에 대한 이력 정보를 제공한다. 문맥 영역은 디지털 자원 내용 정보의 생산 맥락과 다른 내용 정보들과의 관련에 대한 정보를 제공한다. 인증 영역은 해당 디지털 자원의 진본성과 무결성에 대한 정보를 제공한다. 각 영역의 요소들에 매핑되는 표준 요소는 다음 [표 13]과 같이 정리된다.

[표 13] 보존 메타데이터 요소 비교

보존 MD	요소	NAK 8	PREMIS	ISAD(G)	PBCore	유튜브
출처 영역	resource description			수집, 이관의 직접적 출처	instantiation generation	
참조 영역	origin	생산이력, 생산시스템, 생산부서, 생산유형, 생산경로, 의견, 처리상태, 처리일시				
	pre-ingest					
	ingest					
참조 영역	archival retention	보존기간, 보존기간책정 사유, 보존장소, 위치, 소장처, 소장위치, 보존이력, 보존처리유형, 보존처리설명, 보존처리일시, 보존행위자	preservation level, preservation level type, preservation level value, preservation level role, preservation level rational, preservation level date assigned, storage, content location, storage medium	원본의 존재와 위치, 사본의 존재와 위치, 기록물 보관 이력	instantiation location	
문맥 영역	relationship	관계, 관계유형, 관계대상 식별자, 관계설명	relationship, relationship type, related object identifier, related event identifier	관련 기술 단위	pbcore relation identifier, pbcore relation type	
인증 영역	object authentication	무결성체크, 무결성 체크법, 무결성 체크값	fixity, signature information, signature, signature information extension			

출처 영역에는 ISAD(G)의 ‘수집, 이관의 직접적 출처’와 PBCore의 ‘instantiation’의 하위요소인 ‘instantiation generation’이 매핑되었다.

참조 영역의 ‘origin’, ‘pre-ingest’, ‘ingest’ 요소는 NAK 8의 상위요소 ‘생산 이력’과 그 하위요소인 ‘생산시스템’, ‘생산부서’, ‘생산유형’, ‘생산경로’와 ‘생산경로’의 세부요소인 ‘의견’, ‘처리상태’, ‘처리일시’만 매핑되었다. ‘archival retention’의 경우는 유튜브 메타데이터를 제외하고 모두 매핑되었다. NAK 8에서는 상위요소 ‘보존기간’과 그 하위요소인 ‘보존기간’, ‘보존기간책정사유’, 상위요소인 ‘보존장소’, 상위요소 ‘위치’와 그 하위요소인 ‘소장처’, ‘소장위치’, 상위요소 ‘보존이력’과 그 하위요소인 ‘보존처리유형’, ‘보존처리설명’, ‘보존처리일시’, ‘보존행위자’가 매핑되었다. PREMIS에서는 의미 단위 ‘preservation level’과 그 의미 구성 요소인 ‘preservation level type’, ‘preservation level value’, ‘preservation level role’, ‘preservation level rational’, ‘preservation level date assigned’와 의미 단위 ‘storage’와 그 의미 구성 요소인 ‘content location’, ‘storage medium’이 매핑되었다. ISAD(G)에서는 ‘원본의 존재와 위치’, ‘사본의 존재와 위치’, ‘기록물 보관 이력’이 매핑되었고, PBCore에서는 ‘instantiation location’만 매핑되었다.

문맥 영역의 ‘relationship’ 역시 유튜브 메타데이터를 제외하고 모든 표준과 매핑되었다. NAK 8에서는 상위요소 ‘관계’와 그 하위요소인 ‘관계유형’, ‘관계대상식별자’, ‘관계설명’, PREMIS에서는 의미 단위 ‘relationship’과 그 의미 구성 요소인 ‘relationship type’, ‘related object identifier’, ‘related event identifier’가 매핑되었다. ISAD(G)의 경우 ‘관련 기술 단위’가 매핑되었으며, PBCore에서는 ‘pbcore relation’과 그 하위요소인 ‘pbcore relation identifier’, ‘pbcore relation type’이 매핑되었다.

인증 영역의 ‘object authentication’은 NAK 8의 상위요소 ‘무결성 체크’와 그 하위요소인 ‘무결성 체크법’과 ‘무결성 체크값’, PREMIS의 의미 단위 ‘object characteristic’의 ‘fixity’, ‘signature information’과 그 의미 구성 요소인 ‘signature’, ‘signature information extension’이 매핑되었다.

3.2.5 이용자 참여 메타데이터

이상에서 유튜브 동영상 아카이빙 메타데이터를 설계하기 위해 설명, 구조, 관리, 보존 메타데이터로 구분하여 국내외 표준들과 유튜브 메타데이터를 비교해보았다. 그중 구조 메타데이터의 video 관련 요소를 통해 동영상과 관련된 요소를 다양하게 기술 할 수 있음을 파악하였다. 하지만 소셜미디어에 속하는 유튜브 동영상의 이용자 참여와 관련된 요소는 설명, 구조, 관리, 보존 중 매핑할 수 있는 메타데이터가 없었다. 본 연구의 이론적 배경을 통해 알아본 유튜브 동영상의 소셜미디어 특징으로는 기관과 국민의 원활한 양방향 소통과 다양한 이용자 피드백 정보로 인한 사회, 문화적인 특징이 존재한다. 이에 따라 본 연구는 이용자 참여 메타데이터를 생성하였으며 유튜브 메타데이터에서 이에 해당되는 요소를 다음 [표 14]와 같이 추려보았다.

[표 14] 이용자 참여 메타데이터 요소

메타데이터 하위 영역	요소	요소 설명	출처
이용자 참여 영역	조회수	동영상의 조회수	유튜브
	좋아요수	동영상의 '좋아요' 개수	유튜브
	싫어요수	동영상의 '싫어요' 개수	유튜브
	댓글수	동영상의 댓글 개수	유튜브
	댓글ID	동영상의 댓글 ID	유튜브
	댓글내용	동영상의 댓글 내용	유튜브
	대댓글ID	동영상의 댓글에 달린 대댓글 ID	유튜브
	대댓글내용	동영상의 댓글에 달린 대댓글 내용	유튜브

3.3 메타데이터 요소 도출

국내외 표준과 유튜브 메타데이터의 매핑을 통해 설계한 유튜브 동영상 아카이빙을 위한 메타데이터는 [표 15]와 같이 설계하였다. 메타데이터의 요소명은 국내에서의 상호호환성을 고려하여 NAK 8의 매핑된 요소는 NAK 8 요소명을 사용하였다. NAK 8에 매핑되지 않은 요소는 매핑된 다른 표준의 요소명을 사용하였다. 또한 각 요소에 대한 설명은 NAK 8의 설명을 참고하여 작성하였다. NAK 8에 매핑되지 않은 요소는 매핑된 다른 표준의 요소 설명을 참고하여 작성하였다.

[표 15] 유튜브 동영상 아카이빙 메타데이터 요소 및 설명

설명 메타데이터 - 설명 영역		
상위 요소	하위 요소	요소 설명
1. 제목 (Title)		아카이브에 등록된 부여된 유튜브 동영상의 공식 제목을 의미한다. 이는 유튜브 동영상을 식별할 수 있도록 지원해주며, 이용자가 동영상을 검색할 수 있고, 활용할 수 있도록 접근 정보를 제공해준다.
2. 주제 (Subject)		유튜브 동영상에 포함된 핵심적인 내용을 태그를 통해서 간단하고 명료하게 기술 할 수 있다.
	2.1 태그(tag)	유튜브 동영상의 핵심 내용을 간결하게 표현해 줄 수 있는 단어
3. 설명 (Description)		유튜브 동영상의 내용 또는 동영상에 대한 자유로운 설명을 기술한다. 제목에서 설명하지 못한 유튜브 동영상 내용에 대한 검색을 지원할 수 있으며, 동영상을 이용할 때 이해도를 제고해준다.
4. 언어 (Language)		유튜브 동영상 속의 기본 언어를 식별하는 역할을 한다. 언어의 기술은 ISO 639-2 또는 ISO 639-3에 표준화된 3자리 코드를 사용하여 표시한다. 언어가 두 개 이상일 경우 세미콜론(예:kor;eng)으로 구분하여 하나로 표현할 수 있다. 한국어의 경우는 'kor'로 표시한다.
5. 범주 (Coverage)		유튜브 동영상의 시간적, 공간적 범주를 표현해주는 역할을 한다. 시간적 범주는 동영상에서 다루고 있는 내용의 시대에 대해 기술할 수 있으며, 공간적 범주는 동영상 속의 위치를 기술할 수 있다.

구조 메타데이터 - 구조 영역		
6. 비디오 (Video)		‘6.1.재생시간’ ~ ‘6.5.인코딩 정보’까지의 요소를 포함하는 상위요소이며, 본 연구는 유튜브 동영상에 아카이빙하는 것을 목적으로 하고 있기 때문에 비디오유형에 대한 구조만 작성하였다.
	6.1 재생시간 (duration)	유튜브 동영상의 전체 길이 즉, 재생 시간을 기술할 수 있으며, HH:MM:SS 형식으로 작성한다. (예시: 1:14:25)
	6.2 프레임비율 (frame rate)	유튜브 동영상의 트랙의 너비와 높이를 측정하는 값을 기술한다. 이는 표시된 동영상의 크기가 아닌 인코딩된 픽셀의 크기를 나타낸다. 수평으로 측정된 픽셀과 수직으로 쌓인 해상도 데이터의 픽셀수의 조합으로 표현된다. (예시 : 320x240)
	6.3 데이터 전송속도 (data rate)	매초 동안 인코딩되거나 전달 혹은 배포되는 유튜브 동영상의 디지털 미디어 파일의 데이터 양을 나타낸다. (예시 : 56 kbps)
	6.4 압축 (codec)	유튜브 동영상 또는 데이터 스트림에 대해 압축을 하거나 압축을 풀 수 있는 하드웨어 또는 소프트웨어 혹은 이것에 대한 알고리즘을 기술한다.
	6.5 인코딩정보 (encoding)	유튜브 동영상을 재생할 때 Coder를 사용하여 변환된 신호를 기술한다. 이와 같이 사용된 인코딩 정보를 식별함으로써 검색과 접근을 용이하게 하기 위한 디지털 자원의 데이터 구성 또는 다른 데이터 네트워크 전반에 걸쳐 정보를 효율적으로 배포할 수 있다.
7. 장르 (Genre)		동영상 속의 내용 정보에 대한 장르를 기술한다. 유튜브에서 동영상을 등록할 시 선택할 수 있는 장르는 게임, 과학기술, 교육, 노하우/스타일, 뉴스/정치, 비영리/사회운동, 스포츠, 애완동물/동물, 엔터테인먼트, 여행/이벤트, 영화/애니메이션, 음악, 인물/블로그, 자동차/교통, 코미디가 있다.
8. 식별자 (Identifier)		유튜브 동영상에 부여된 고유한 식별자, 혹은 검색이나 이용을 위해 추가적으로 생성된 식별자를 의미한다. 이는 유튜브 동영상을 고유하게 식별할 수 있으며, 이를 통해 유튜브 동영상을 검색할 시 접근점으로도 활용할 수 있다. 하위 요소로는 ‘비디오 ID’와 URL이 있다.
	8.1 비디오ID (VideoID)	특정 유튜브 동영상의 고유 정보 ID를 기술한다. (예시 : 8Udo8uQMSr8)

	8.2 URL	특정 유튜브 동영상의 웹상에서의 고유 위치 정보를 기술한다.
관리 메타데이터 - 권리 영역		
9. 생산자 (Creator)		유튜브 동영상을 생산하고 기관 유튜브 채널에 업로드 한 개인 또는 기관을 식별할 수 있는 정보로 '생산자 유형', '생산자 이름'의 하위요소를 가지고 있다.
	9.1 생산자유형 (creator type)	유튜브 동영상을 생산한 주체를 기술한다.
	9.2 생산자이름 (creator name)	유튜브 동영상을 생산한 주체가 소속된 기관의 이름 혹은 개인의 이름을 기술한다.
10. 권한 (Rights)		유튜브 동영상의 접근과 이용을 관리하고 통제하기 위한 권한 정보로 이루어진 요소들의 집합이다.
	10.1 권한 식별자 (rights identifier)	유튜브 동영상에 적용되는 권한 정보들을 고유하게 식별하는데 사용되는 요소이다.
	10.2 권한 근거 (rights basis)	'10.1 권한 식별자'에 의해 식별된 권한 또는 권한에 대한 근거를 지정하는 역할로 사용될 수 있다.
	10.3 공공저작물 여부 (Public works or not)	아카이빙된 유튜브 동영상이 공공기관에서 업무상 작성하여 공표한 저작물이나 계약에 따라 저작재산권의 전부를 보유한 공공저작물인지 아닌지에 대한 여부를 기술한다. 공공저작물은 기준은 저작권법 제 24조 2항에 따른다.
	10.4 공공저작물 자유이용허락표 시 유형 (type of Korea open government license)	공공저작물에 부여된 공공저작물 자유이용 허락 표시(공공누리)의 유형을 기술한다. 공공누리의 유형은 제1유형 ~ 제4유형까지 존재하는데 각 유형에 대한 설명은 다음과 같다. 제1유형은 출처 표시를 의무로 해야 하며, 상업적인 활용 여부와 무관하게 무료로 자유롭게 이용할 수 있고, 2차적 저작물을 생성하는 등 다양하게 변형하여 이용 가능한 유형이다. 제2유형은 출처 표시를 의무로 해야 하며, 무료로 자유롭게 이용할 수 있고, 2차적 저작물을 생성하는 등 다양하게 변형하여 이용가능하나 상업적 목적으로 이용하는 것은 불가능한 유형

		<p>이다.</p> <p>제3유형은 출처 표시를 의무로 해야 하며, 상업적인 활용 여부와 무관하게 무료로 자유롭게 이용 가능하지만 공공저작물의 내용을 변형하거나 변경할 수 없는 유형이다.</p> <p>제4유형은 출처 표시를 의무로 해야 하며, 상업적으로 활용하거나 자유롭게 이용할 수 없고, 2차적 저작물을 생성하는 등 변형해서 이용할 수 없다.</p>
	10.5 공개구분 (external access control)	유튜브 동영상을 공개 또는 비공개 하는지에 대한 여부를 기술한다.
	10.6 비공개사유 (external access reason)	동영상을 비공개로 설정한 것에 대한 사유를 작성해야 한다. 공공기관에서 업로드된 유튜브 동영상들은 기관의 홍보를 주된 목적으로 생산되기 때문에 일반적으로 모두 또는 전체에게 공개로 설정해서 업로드한다. 하지만 유튜브에 업로드한 영상에서의 문제점이 발견됐거나, 홍보영상에 출연한 유명인의 이미지가 추락한 경우, 해당 유튜브 영상을 비공개로 처리해야 하는 상황이 발생할 수도 있다. 이러한 경우, 영상이 비공개 처리된 이유를 적절히 작성한다면 추후에 재공개 및 폐기 등 해당 영상을 적절하게 활용할 수 있을 것이다.
	10.7 시청등급 (Audience level)	유튜브 동영상을 시청할 수 있는 연령에 대한 정보를 기술한다.
관리 메타데이터 - 관리 영역		
11. 일시 (Date)		유튜브 동영상의 생산과 업로드한 행위가 발생한 일시를 기술한다.
	11.1 생산일시 (time created)	공공기관에서 유튜브에 해당 동영상을 업로드한 일시를 기술한다.
	11.2 등록일시 (time registered)	아카이브에 특정 유튜브 동영상을 등록한 일시를 기술한다.
12. 관리이력		유튜브 동영상에 영향을 미치는 모든 관리행위

(Management history)		에 대한 이력정보를 기술한다. 유튜브 동영상이 아카이브로의 이동이나 동영상에 행해진 관리행위에 대한 내용을 누적함으로써 동영상에 대한 관리 및 통제행위에 대한 이력을 제공한다. 이를 통해 유튜브 동영상의 진본성, 무결성, 신뢰성, 이용가능성을 확보할 수 있다.
	12.1 관리유형 (event type)	유튜브 동영상에 영향을 미친 관리행위의 유형으로 ‘아카이브의 등록’, ‘공개 재분류’, ‘기술 재입력’, ‘이용자 피드백 정보 업데이트’ 등을 기술할 수 있다.
	12.2 관리설명 (event description)	‘12.1 관리 유형’에서 기술된 유형이 발생한 사유에 대해서 설명한다.
	12.3 관리일시 (event date time)	유튜브 동영상의 상태에 영향을 미치는 관리행위에 대한 발생 일시를 기술한다.
	12.4 관리행위자 (event agent)	유튜브 동영상의 상태에 영향을 미치는 관리 행위를 수행한 행위자에 대해 기술한다.
	12.5 변경요소 (changed element)	유튜브 동영상의 상태에 영향을 미치는 관리행위를 통해서 변경된 사항에 대해서 기술한다.
관리 메타데이터 - 기술 영역		
13. 포맷 (Format)		유튜브 동영상의 데이터 포맷 종류를 기술한다.
	13.1 포맷명 (format name)	유튜브 동영상의 데이터 파일 확장자를 기술한다.
	13.2 포맷버전 (format version)	유튜브 동영상의 파일 포맷의 버전의 정보를 기술한다.
14. 크기 (Size)		유튜브 동영상의 규모에 대한 정보를 제공한다.
	14.1 파일크기 (file size)	유튜브 동영상 파일에 대한 크기를 기술한다.
15. 운영 시스템 (Operating system)		유튜브 동영상을 이용하기 위해 요구되고 권장되는 운영체계에 대한 소프트웨어 또는 하드웨어를 기술한다.
16.		유튜브 동영상을 실행하기 위한 응용 프로그램

어플리케이션 (Application)		에 대한 정보를 제공한다.
	16.1 생성 어플리케이션 명 (creating application name)	유튜브 동영상 파일을 생성할 수 있는 어플리케이션 이름을 기술한다.
	16.2 생성 어플리케이션 버전 (creating application version)	유튜브 동영상 파일을 생성한 어플리케이션 버전 정보를 기술한다.
보존 메타데이터 - 출처 영역		
17. 수집, 이관의 직접적 출처 (Immediate source of acquisition of information)		아카이브에서 수집 또는 이관을 한 출처에 대한 상황을 기록하기 위한 요소로 수집된 날짜 및 방법과 기증자에 대한 정보를 기술한다.
보존 메타데이터 - 참조 영역		
18. 생산이력 (Creation history)		유튜브 동영상의 생산과정에서 발생한 이력정보에 대하여 기술한다.
	18.1 기관명 (corporate name)	유튜브 동영상을 생산한 기관에 대하여 기술한다.
	18.2 의견 (opinion)	유튜브 동영상의 생산 행위 과정 중 생산한 기관에서 유튜브 동영상에 대해 입력한 의견을 기술한다.
	18.3 처리상태 (status of creation action)	유튜브 동영상을 생산한 기관이 해당 동영상을 처리한 상태에 대해 기술한다.
	18.4 처리일시 (time of creation action)	유튜브 동영상을 생산한 기관이 해당 동영상을 처리한 일시를 기술한다.
19. 보존이력 (Preservation history)		유튜브 동영상이 아카이브로 등록된 이후 행해진 모든 보존처리행위에 대한 이력정보를 기술한다. 이와 같이 유튜브 동영상에 행해진 모든

		보존행위를 관리함으로써 유튜브 동영상을 유지 관리하기 위해 수행한 모든 업무에 대한 증거를 제공할 수 있으며, 특정 유튜브 동영상과 관련되어 보존할 필요가 있는 업무의 설명책임성을 증명하고, 진본성, 무결성, 신뢰성, 이용가능성을 보장한다.
	19.1 보존처리유형 (preservation action type)	유튜브 동영상의 원형을 그대로 유지한 채 수명을 연장시키기 위해 취한 행위로 마이그레이션(migration), 에뮬레이션(emulation) 등의 보존처리 유형을 기술한다.
	19.2 보존처리설명 (preservation action description)	유튜브 동영상의 상태에 대한 정보, 보존처리 사유 및 근거에 대한 상세한 설명을 기술한다.
	19.3 보존처리일시 (preservation action date)	유튜브 동영상에 보존처리 행위가 취해진 일시를 기술한다.
	19.4 보존처리행위자 (preservation agent)	유튜브 동영상에 보존 행위를 행한 행위자 정보에 대해 기술한다.
20. 보존장소 (Preservation place)		유튜브 동영상이 최종적으로 보존되는 장소에 대해 기술한다.
21. 위치 (Location)		아카이브에서 유튜브 동영상의 소장위치에 대한 정보로 소장 장소 및 상세 소장 위치 또는 시스템 저장위치에 대한 정보를 제공한다.
	21.1 소장처 (location)	유튜브 동영상의 원본 또는 저장 매체, 보존매체 등이 보존되어 있는 기관의 명칭을 기술한다.
	21.2 소장위치 (storage details)	유튜브 동영상이 저장된 기관의 서고 정보와 배열된 위치를 나타내는 스토리지 또는 매체번호, 매체가 배열된 위치에 대한 정보를 기술한다.
	21.3 원본의 존재와위치 (Existence and location of	유튜브 동영상에 대한 사본이 존재하는 경우, 소장처와 이용가능성 등을 제공하는 요소로 유튜브 동영상의 원본이 존재한다면 원본의 위치를 관리번호와 함께 기술한다. 원본이 없거나

	originals)	소재가 불분명하여도 그 자체를 기술한다.
	21.4 사본의 존재와위치 (Existence and location of copies)	유튜브 동영상의 복사본의 존재와 소장처, 이용가능성을 제공하는 요소로 유튜브 동영상의 사본을 기관 내/외부에서 이용할 수 있는 경우, 관리번호와 함께 소장처를 기술한다.
보존 메타데이터 - 문맥 영역		
22. 관계 (Relation)		유튜브 동영상과 다른 유튜브 동영상간의 관계, 사본, 참조 관계 등 다양한 관계에 대해 표현할 수 있다. 이는 업무행위의 증거로서 생산된 유튜브 동영상 간의 관계를 기술함으로써 유튜브 동영상의 생산과 관리에 대한 맥락정보를 제공할 수 있으며, 상호 관련 있는 유튜브 동영상과 정보자원들을 위한 접근점을 제공해준다.
	22.1 관계유형 (relation type)	유튜브 동영상과 다른 동영상의 관계, 사본, 참조관계 등 다양한 관계의 유형을 기술한다.
	22.2 관계대상 식별자 (relation item ID)	22.1 관계유형'에서 선택된 값에 따른 관계 대상의 식별자를 기술한다.
	22.3 관계설명 (relation description)	해당 관계에 대한 상세한 설명을 자유롭게 기술한다.
보존 메타데이터 - 인증 영역		
23. 진본성 확인 (Object authentication)		유튜브 동영상의 진본성 확인에 대한 정보를 기술한다.
	23.1 서명 (signature)	진본성 확인을 인증한 관리행위자의 서명을 기술한다.
24. 무결성체크 (Integrity check)		소셜미디어 기록인 유튜브 동영상의 무결성을 유지하기 위해 유튜브 동영상이 관리 및 보존되는 과정 속에서 인가받지 않은 접근이나 행위에 의해 위변조 혹은 훼손되지 않았음을 점검하는 방법이다.
	24.1 무결성체크법 (integrity check name)	유튜브 동영상의 무결성을 확인하기 위한 무결성 체크기법으로 기관에서 사용하는 무결성 체크기법을 기술한다. (예시 : SHA-256)
	24.2 무결성체크값	무결성 체크법에 의해 생성된 값을 기술한다.

	(integrity check value)	
이용자 참여 메타데이터 - 이용자 참여 영역		
25. 이용자 피드백정보 (User feedback information)		유튜브 동영상을 시청한 이용자들이 의해 나타나는 수치정보들을 기술함으로써 업로드한 동영상에 대한 긍정 또는 부정적인 평가를 한눈에 파악할 수 있을뿐더러 긍정적인 평가에 대한 부분은 발전시키고 부정적인 평가에 대한 부분은 지양할 수 있다.
	25.1 조회수 (views)	이용자들이 해당 유튜브 동영상을 시청한 횟수를 기술한다.
	25.2 좋아요수 (number of like)	이용자들이 해당 유튜브 동영상을 시청한 뒤 '좋아요'를 표시한 수치정보를 기술한다.
	25.3 싫어요수 (number of dislike)	이용자들이 해당 유튜브 동영상을 시청한 뒤 '싫어요'를 표시한 수치정보를 기술한다.
	25.4 댓글 수 (number of reply)	이용자들이 해당 유튜브 동영상을 시청한 뒤 작성한 '댓글'에 대한 수치정보를 기술한다.
26. 댓글 (Reply)		이용자가 유튜브 동영상을 시청한 후 이용자의 의견이 들어가 있는 정보로 이를 통해 정책에 대한 솔직한 이용자들의 의견을 알 수 있으며, 당시 사회의 트렌드 및 이용자들 간의 공론의 장으로써 새로운 문화가 형성될 수 있다.
	26.1 댓글ID (reply ID)	유튜브 동영상에 댓글을 작성한 이용자 ID를 기술한다.
	26.2 댓글내용 (reply content)	유튜브 동영상에 작성된 댓글의 내용을 기술한다.
	26.3 대댓글ID (re-reply ID)	특정 댓글에 대한 대댓글을 작성한 이용자 ID를 기술한다.
	26.4 대댓글내용 (re-reply content)	특정 댓글에 대한 대댓글의 내용을 기술한다.

IV. 메타데이터 요소 평가

메타데이터 요소 평가와 핵심 요소 선정방법에 관한 자세한 설명은 다음 [표 16]과 같다.

[표 16] 메타데이터 요소 평가 상세 절차

평가 절차	수행 방법
설문 조사	5점 리커트 평가
데이터 정제	통일되지 않은 답변 정제
기술 통계 분석	빈도 분석 및 평균 분석
타당도 검증	요인분석
신뢰도 검증	크론바흐 알파값(chronbach's alpha)

첫째, 연구자가 설계한 메타데이터 요소에 대해 적합성을 평가받기 위해 메타데이터 전문가를 대상으로 온라인 설문조사를 실시하였다. 전문가 선정방법은 선행연구를 참조하였다. 이를 통해 디지털 아카이빙 및 메타데이터의 실무 경험과 연구(학문) 경험을 토대로 15명을 선정하였다. 설문지는 1. 일반현황, 2. 메타데이터 요소 평가, 3. 추가 요소 및 기타 의견으로 구성하였으며, 2. 메타데이터 요소 평가는 5점 리커트 방법으로 진행하였다.

둘째, 설문결과에 대해 정제작업을 실시하였다. 수집된 데이터 중 통일되지 않은 다양한 응답에 대해 통계처리가 가능하도록 수정하고, 파일을 변환하였다.

셋째, 설문결과에 대해 기술 통계 분석을 실시하였다. 수집된 데이터의 인구통계학적 분석을 위해 1. 일반현황(소속기관 유형, 학력, 경력)응답을 빈도 분석하였다. 또한 2. 메타데이터 요소 평가에 대한 응답은 평균분석을 실시하여 최솟값, 중위값, 최댓값, 평균값, 표준편차를 분석하였다.

넷째, 타당도 검증을 실시하였다. 본 연구는 전문가 평가한 메타데이터 요소 중 통계학적으로 유의한 결과가 나타난 요소들을 핵심 요소로 추출하였다. 타당도 검증은 메타데이터 하위 영역별로 요인분석을 통해 이루어졌다.

우선 KMO 및 Bartlett의 구형성 검정을 통해 요인분석을 수행해도 되는지 알아보았다. KMO는 변수(요소)들 간의 편상관을 확인하는 작업으로, 변

수의 숫자와 케이스의 숫자의 적절성을 표현하는 표본의 적합도를 의미한다. KMO의 값이 높을수록 좋지만, 일반적으로 0.5보다 크면 요인분석을 실시하는 것이 적절하다고 판단할 수 있다(노경섭, 2019). Bartlett의 구형성 검정은 요인분석을 실시할 때, 상관계수의 행렬이 대각행렬일 경우 요인분석을 실시하는 것이 부적절하다고 판단한다. Bartlett 값이 $p < 0.05$ 인 경우는 대각행렬이 아님을 의미하므로 요인분석을 실시하는데 적절하다고 할 수 있다(노경섭, 2019).

요인을 추출하는 방법은 공통요인분석(Common Factor Analysis, CFA) 방법 중 가장 많이 쓰이는 방법인 최대우도(노경섭, 2019) 방법을 사용하였다. CFA는 ‘자료의 축소’라는 차원을 포함하여 자료 내재적으로 존재하는 속성까지 찾아내는 방법이다. 즉, 변수 간의 공통요인을 추출한 후 이를 통해 각 변수 간의 상관관계를 발견하고, 각 변수의 성질을 축소하여 설명하는 방법이다. 요인 회전은 요인 간 연관 관계를 유지해서 회전하는 방법 즉, 상관관계가 전혀 없다고 가정하지 않는((상관계수) $\neq 0$) 사각회전 방법을 사용하였으며 그 중 직접 오블리민(oblimin) 방법을 적용하였다.

이상의 방법을 사용하여 추출되는 요소 값 중 타당도 검증에서 유의미하다고 판단하는 기준값은 다음과 같다. 최대우도법을 통해 추출된 공통성 값이 0.4 이상이면서 초기 고유값(eigen value)이 1이상인 요소이다. 공통성은 추출된 요인들에 의해 설명되는 변수의 분산(노경섭, 2019)을 나타낸다. 즉, 해당 요소가 추출된 공통성 값만큼의 설명력을 지닌다고 볼 수 있다. 또한 고유값은 요인의 설명력으로서 요인 적재값의 제곱합으로 구할 수 있는데 해당 수치가 높으면 요인이 측정요소들의 분산을 잘 설명한다는 의미이며, 일반적으로 1이상인 것을 기준으로 한다(송지준, 2008; 강병서, 김계수, 2009). 마지막으로 직접 오블리민 회전방법으로 회전된 요인의 성분(적재량)의 기준값은 다음과 같다. 일반적으로 0.3 이상이면 추출된 요소가 통계적으로 유의미하는 것으로 판단하고, 0.5 이상인 경우 매우 유의한 것으로 판단한다(노경섭, 2019).

다섯째, 요인분석 결과에서 유의미하다고 나타난 요소들을 추출하여 메타데이터별, 하위 영역별로 신뢰도 검증을 실시하였다. 요인분석의 결과를 제시

할 때 해당 결과가 신뢰성도 확인해야 한다(노경섭, 2019). 신뢰도 검증의 기준은 크론바흐 알파(chronbach's alpha) 값을 활용하였다. 크론바흐 알파값이 0.7 이상이면 신뢰도가 아주 높다고 판단하는데 사회과학분야에서는 일반적으로 0.6 이상이면 높은 값으로 본다. 통계분석을 위한 최소한의 기준값은 0.5이다(박태섭, 2017).

이상에서 설명한 요인분석 및 신뢰도 검증의 방법과 추출 기준값을 [표 17]과 같이 정리하였다.

[표 17] 요인분석 및 신뢰도 검증 방법 및 추출 기준값

구분	방법	기준값
요인분석	KMO 및 Bartlett의 구형성 검증	KMO > 0.5
		Bartlett의 $p < 0.05$
	최대우도법	추출 공통성 값 ≥ 0.4 초기 고유값 ≥ 1
	직접 오블리민 회전법	성분(적재량) 값 ≥ 0.3
신뢰도 검증	크론바흐 알파값 측정	크론바흐 알파값 ≥ 0.5

본 연구의 마지막 단계인 최종 정리 단계에서는 [표 17]의 기준값을 모두 충족하는 요소들을 메타데이터 핵심 요소로 선정하였다.

메타데이터 요소 평가 단계에서 수행한 통계 분석 도구는 다음 [표 18]과 같이 정리하였다. 데이터 정제작업은 excel 2016을 통해 수행하였으며, 기술 통계 분석, 타당도 검증, 신뢰도 검증은 spss 28을 통해 수행하였다.

[표 18] 통계 분석 도구

분석 방법	분석 도구
데이터 정제	excel 2016
기술 통계 분석	spss 28
타당도 검증	
신뢰도 검증	

4.1 설문조사 개요

설문조사는 Ⅲ. 메타데이터 요소 설계에서 구조화한 메타데이터의 적정성을 검토하는 단계이다. 이를 통해 타당도와 신뢰도를 검증하였다. 설문은 온라인으로 진행하였으며 설문지는 [표 19]와 같이 구성하였다. 설문지 전체 내용은 [부록 1]에 첨부하였다.

[표 19] 설문지 구성

구분	문항			응답 방법
1. 일반 현황	1.1 소속 기관 유형			주관식
	1.2 최종 학력			주관식
	1.3 경력 사항			객관식
2. 메타데이터 요소 평가	2.1 설명 메타데이터	설명 영역	5 문항	5점 리커트 1점 : 매우 부적절 2점 : 부적절 3점 : 보통 4점 : 적절 5점 : 매우 적절
	2.2 구조 메타데이터	구조 영역	8 문항	
	2.3 관리 메타데이터	권리 영역	9 문항	
		관리 영역	7 문항	
		기술 영역	6 문항	
	2.4 보존 메타데이터	출처 영역	1 문항	
		참조 영역	13 문항	
		문맥 영역	3 문항	
		인증 영역	3 문항	
	2.5 이용자 참여 메타데이터	이용자 참여 영역	8 문항	
3. 추가 요소 및 기타 의견				주관식

4.1.1 전문가 선정

전문가를 판별하는 명확한 기준을 찾기는 어렵다. 따라서 본 연구에서는 본 연구주제와 유사한 기존 연구에서 활용한 전문가 선정방법을 조사하여 기준을 설정하였다. 기존 연구의 주제와 전문가 선정방법은 [표 20]처럼 정리된다.

[표 20] 연구자별 메타데이터 전문가 선정방법

연구자	연구 주제	전문가 선정방법
한국 영상자료원 (2003)	국내 영상 자료 메타데이터 표준안 마련	<ol style="list-style-type: none"> 영상자료 데이터 목록 작성자 <ul style="list-style-type: none"> - 현재 비도서 자료 목록 담당자 - 비도서 자료 목록 경험자(2년 이상) - 현재 영상 자료 취급 담당자 - 영상자료 취급 경험자(2년 이상) 영상자료 데이터 기술자/시스템 관리자 <ul style="list-style-type: none"> - 현재 영상자료 데이터/시스템 담당자 - 영상자료 데이터/시스템 경험자(2년 이상)
윤선정 (2008)	교육용 게임에 대한 메타데이터 개발	<ol style="list-style-type: none"> 게임 기획 & 개발 전문가 <ul style="list-style-type: none"> - 2년 이상의 현업 실무자 게임 전공 연구원 <ul style="list-style-type: none"> - 게임 전공 4학년 이상 학부생과 대학원생 교수 학습 설계자 <ul style="list-style-type: none"> - 교육 경력 5년 이상인 교수
장보성 (2011)	디지털 연구성과물의 장기보존을 위한 메타데이터 연구	<ol style="list-style-type: none"> 장기보존 전문가 <ul style="list-style-type: none"> - 도서관 및 기록관에서 실무경험 및 학술저서 경험이 있는 자 학계 전문가 <ul style="list-style-type: none"> - 대학이나 대학원에서 장기보존, 디지털 아카이빙을 가르치거나, 관련 연구를 수행한 자
곽성호 (2012)	국내 방송 콘텐츠 메타데이터 표준 개발	<ol style="list-style-type: none"> 방송국 실무자 <ul style="list-style-type: none"> - 방송국에서 직접 메타데이터를 관리 및 운영 인터넷 영상 유통업체의 실무자 <ul style="list-style-type: none"> - 인터넷 영상 유통업체에서 직접 메타데이터를 이용해 영상을 유통 및 관리 소프트웨어 솔루션 업체의 실무자 <ul style="list-style-type: none"> - 소프트웨어 솔루션 업체에서 메타데이터를 관리하는 시스템에 참여하는 실무자
이혜은 (2013)	고지도 정보시스템 구축을 위한 메타데이터 요소	<ol style="list-style-type: none"> 대학이나 대학원에서 고지도에 대한 강의 경험 및 관련 연구를 수행한 경험이 있는 연구자 도서관 및 박물관에서 고지도를 관리하고 고

	설계	지도의 수집, 고지도 관련 시스템 운영 등의 경험이 있는 실무자
황혜림 (2018)	공개교육자원 관리와 활용을 위한 Open Course Ware(OCW) 메타데이터 스키마 개발	1. KERIS의 OCW 담당자 및 OCW 서비스를 제공하고 있는 대학에서 메타데이터를 담당하는 사서 전문가 2. 메타데이터에 대해 전반적인 지식을 가지고 실제로 메타데이터를 설계해 본 경험이 있는 문헌정보학 전공 대학원생 이용자
봉지현 (2019)	일본군 '위안부' 구술기록 관리를 위한 메타데이터 요소 선정 연구	1. 실제 일본군 '위안부' 관련 기록을 소장하고 있는 일본군 '위안부' 관련 기관 담당자 2. 대학 도서관 내 메타데이터를 담당하고 있는 담당자

[표 20]을 통해서 알 수 있듯이 전문가 선정방법의 중요한 기준은 실무경험과 연구(학문) 경험이다. 본 연구에서도 설문대상인 전문가의 기준은 기관에서 디지털 아카이빙 또는 메타데이터 업무 관련 경험과 연구경험으로 두었다. 해당 기준을 토대로 본 연구는 공공기관을 비롯한 디지털 아카이브 및 기타 분야에서 아카이빙 또는 메타데이터 업무 담당자 15명에게 온라인으로 응답하도록 하였으며, 2022년 4월 19일부터 4월 29일까지 10일간 수행하였다.

4.2 설문조사 데이터 취합 및 정제

설문지는 15부 배포하여 총 14부(93%)가 회수되었다. 설문지는 네이버 폼으로 제작 하였으며, 1. 일반 현황과 3. 추가 요소 및 기타 의견은 선택사항으로 설정하고, 2. 메타데이터 요소 평가는 필수사항으로 설정했다. 14명의 응답결과는 1.3 경력 사항에 대해 한 명이 응답하지 않았으며, 2. 메타데이터 요소 평가 문항에 대해서는 모두 응답하였다. 3. 추가 요소 및 기타의견에서는 7명이 응답하였다.

정제작업을 수행한 문항은 '1. 일반 현황' 중 '1.1 소속기관 유형'이다. 14개의 응답 현황을 보면 공공기관 8건, 국립도서관 1건, IT회사 1건, 대학 1건, 금융기관 1건, 법인회사 1건, 실제 기관명 응답 1건이다. 본 연구는 응답

자 기관 유형에 따른 차이를 분석하지는 않지만, 인구통계학적 분석을 위해 국립도서관은 공공기관으로, IT회사, 법인회사, 금융기관, 실제 기관명은 기타 기관으로 수정하였다. 정제 내역은 [표 21]과 같다.

[표 21] 메타데이터 요소 평가 설문지 결과 데이터 취합 정제 내용

문항	응답자	정제 전	정제 후
1.1 소속 기관 유형	R1	• 공공기관	• 공공기관
	R2	• 국립도서관	• 공공기관
	R3	• 공공기관	• 공공기관
	R4	• 공공기관	• 공공기관
	R5	• 공공기관	• 공공기관
	R6	• 공공기관	• 공공기관
	R7	• 공공기관	• 공공기관
	R8	• 공공기관	• 공공기관
	R9	• 공공기관	• 공공기관
	R10	• 실제 기관명으로 응답	• 기타기관
	R11	• 법인회사	• 기타기관
	R12	• 금융기관	• 기타기관
	R13	• IT회사	• 기타기관
	R14	• 대학	• 대학

4.3 설문조사 결과

4.3.1 응답자 분석

응답자의 인구통계학적 분석은 설문지에서와 같이 소속기관 유형, 최종 학력, 경력 사항 순서로 정리하였다. 총 14명의 응답자 중 소속기관 유형에 대한 응답은 공공기관이 9명(64%), 기타기관이 4명(29%), 대학이 1명(7%)으로 나타났으며 [표 22]와 같다.

[표 22] 소속기관 유형 응답 결과

구분	공공기관	기타기관	대학	계
응답 수	9	4	1	14
비율(%)	64%	29%	7%	100%

둘째, 응답자의 학력 구분은 박사 4명(29%), 박사(수료) 2명(14%), 석사 4명(29%), 학사 4명(29%)으로 나타났으며 [표 23]과 같다.

[표 23] 최종 학력 응답 결과

구분	박사		석사	학사	계
	박사 졸업	박사(수료)			
응답 수	4	2	4	4	14
비율(%)	28.57%	14.29%	28.57%	28.57%	100%

셋째, 응답자 경력 사항 구분은 1명이 응답하지 않아 13명으로 분석을 실시하였다. 그 결과, 0~5년 3명(23%), 6~10년 4명(31%), 11~15년 2명(15%), 그 이상 4명(31%)으로 나타났다. 경력 사항 응답결과는 [표 24]와 같이 정리된다.

[표 24] 경력 사항 응답 결과

구분	0~5년	6~10년	11~15년	그 이상	계
응답 수	3	4	2	4	13
비율(%)	23%	31%	15%	31%	100%

4.3.2 기술 통계 분석

4.3.2.1 설명 메타데이터 기술 통계

설명 메타데이터의 설명 영역 요소들의 기술 통계는 다음 [표 25]와 같다. ‘제목’은 14명 모두에게 5점(매우 적절)을 응답받았고, ‘언어(4.57)’, ‘설명(4.50)’, ‘범주, 태그(4.21)’ 순으로 평균값이 높았다.

[표 25] 설명 메타데이터 - 설명 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
1. 제목(Title)		5	5	5	5.00	0.000
2. 주제(Subject)	2.1 태그(tag)	2	4	5	4.21	0.893
3. 설명 (Description)		3	5	5	4.50	0.760
4. 언어 (Language)		3	5	5	4.57	0.756
5. 범주(Coverage)		2	4.5	5	4.21	0.975

4.3.2.2 구조 메타데이터 기술 통계

구조 메타데이터의 구조 영역 요소들의 기술 통계는 다음 [표 26]과 같이 정리된다. 평균값은 ‘재생시간(4.93)’, ‘URL(4.5)’, ‘압축(4.07)’, ‘프레임비율(4)’, ‘비디오 ID(4)’ 순으로 나타났다. 구조 영역에서 가장 평균값이 낮은 요소는 ‘데이터 전송 속도(3.86)’로 나타났다.

[표 26] 구조 메타데이터 - 구조 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준편차
6. 비디오 (Video)	6.1 재생시간 (duration)	4	5	5	4.93	0.267
	6.2 프레임비율 (frame rate)	1	4	5	4.00	1.177
	6.3 데이터전송속도 (data rate)	1	4	5	3.86	1.231
	6.4 압축 (codec)	1	4.5	5	4.07	1.269
	6.5 인코딩정보 (encoding)	1	4.5	5	3.93	1.385
7. 장르 (Genre)		2	4	5	3.79	1.122
8. 식별자 (Identifier)	8.1 비디오ID (video ID)	2	4	5	4.00	1.109
	8.2 URL	2	5	5	4.50	0.941

4.3.2.3 관리 메타데이터 기술 통계

1) 권리 영역

관리 메타데이터의 하위 영역은 권리 영역, 관리 영역, 기술 영역으로 구성된다. 하위 영역별로 평균값 및 표준편차 등을 살펴본 결과는 다음과 같다. 먼저, 권리 영역의 기술 통계는 [표 27]과 같이 정리하였다. 평균값이 가장 높은 요소는 해당 콘텐츠의 공개 여부인 ‘공개구분(4.57)’으로 나타났으며, 다음으로 유튜브 동영상의 저작권을 설명하는 ‘공공저작물여부’와 ‘공공표시유형’이 4.29로 나타났다. 권리 영역에서 평균값이 가장 낮은 요소는 ‘생산자유형(3.93)’이었다.

[표 27] 관리 메타데이터 - 권리 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준편차
9. 생산자 (Creator)	9.1 생산자유형 (creator type)	1	4	5	3.93	1.207
	9.2 생산자이름 (creator name)	1	4.5	5	4.07	1.207
10. 권한 (Rights)	10.1 권한식별자 (rights identifier)	2	5	5	4.21	1.051
	10.2 권한근거 (rights basis)	2	5	5	4.07	1.207
	10.3 공공저작물여부 (Public works or not)	2	5	5	4.29	0.994
	10.4 공공저작물 자유이용허락표시유형 (type of Korea open government license)	2	5	5	4.29	0.994
	10.5 공개구분 (external access control)	3	5	5	4.57	0.756
	10.6 비공개사유 (external access reason)	2	4	5	4.14	0.949
	10.7 시청등급 (Audience level)	3	4	5	4.14	0.864

2) 관리 영역

관리 영역의 기술 통계는 다음 [표 28]과 같이 정리하였다. ‘생산일시(4.50)’의 평균값이 가장 높았으며, ‘등록일시(4.43)’, ‘관리유형(3.43)’, ‘변경요소(3.43)’, ‘관리설명(3.29)’, ‘관리일시(3.14)’ 순으로 나타났다. 평균값이 가장 낮은 요소는 ‘관리행위자(3.07)’로 나타났다. 해당 요소는 5개 메타데이터에 속한 요소들 중 평균값이 제일 낮은 요소로 나타났다.

[표 28] 관리 메타데이터 - 관리 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
11. 일시 (Date)	11.1 생산일시 (time created)	3	5	5	4.50	0.855
	11.2 등록일시 (time registered)	3	5	5	4.43	0.756
12. 관리 이력 (Management History)	12.1 관리유형 (event type)	2	3	5	3.43	1.016
	12.2 관리설명 (event description)	1	3	5	3.29	1.069
	12.3 관리일시 (event date time)	1	3	5	3.14	1.099
	12.4 관리행위자 (event agent)	1	3	5	3.07	1.141
	12.5 변경요소 (changed element)	1	4	5	3.43	1.284

3) 기술 영역

기술 영역의 기술 통계는 다음 [표 29]와 같다. 평균값이 4점 이상으로 나타난 요소들은 ‘파일크기(4.43)’과 ‘포맷명(4.36)’ 뿐이었다. 나머지 요소들은 평균값이 3점대였으며, 그중 평균값이 가장 낮은 요소는 ‘생성어플리케이션버전(3.36)’으로 나타났다.

[표 29] 관리 메타데이터 - 기술 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
13. 포맷 (Format)	13.1 포맷명 (format name)	3	4	5	4.36	0.633
	13.2 포맷버전 (format version)	2	4	5	3.71	1.267

14. 크기 (Size)	14.1 파일크기 (file size)	2	5	5	4.43	1.089
15. 운영 시스템 (Operating system)		1	4	5	3.43	1.505
16. 어플리케이션 (Application)	16.1 생성 어플리케이션명 (creating application name)	1	3	5	3.43	1.284
	16.2 생성 어플리케이션버전 (creating application version)	1	3	5	3.36	1.216

4.3.2.4 보존 메타데이터 기술 통계

1) 출처 영역

보존 메타데이터는 출처 영역, 참조 영역, 문맥 영역, 인증 영역으로 구성된다. 우선 출처 영역의 기술 통계는 [표 30]과 정리하였다. 출처 영역의 요소는 ‘수집,이관의직접적출처’뿐이다. 평균값은 4.36으로 나타났다.

[표 30] 보존 메타데이터 - 출처 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
17. 수집,이관의 직접적출처 (Immediate source of acquisition of information)		2	5	5	4.36	1.082

2) 참조 영역

참조 영역의 기술 통계는 [표 31]과 같이 정리하였다. ‘소장처(4.5)’의 평균값이 제일 높았으며, ‘기관명(4.43)’, ‘원본의존재와위치(4.43)’, ‘소장위치(4.21)’, ‘보존장소(4)’ 등의 순으로 나타났다. 평균값이 가장 낮은 요소는 ‘처리상태(3.29)’로 나타났다.

[표 31] 보존 메타데이터 - 참조 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
18. 생산이력 (Creation history)	18.1 기관명 (corporate name)	3	5	5	4.43	0.756
	18.2 의견 (opinion)	2	3.5	5	3.57	1.016
	18.3 처리상태 (status of creation action)	2	3	5	3.29	0.994
	18.4 처리일시 (time of creation action)	2	3	5	3.43	0.938
19. 보존이력 (Preservation history)	19.1 보존처리유형 (preservation action type)	1	4	5	3.64	1.216
	19.2 보존처리설명 (preservation action description)	1	4	5	3.57	1.158
	19.3 보존처리일시 (preservation action date)	1	4	5	3.71	1.139
	19.4 보존처리 행위자 (preservation agent)	1	3	5	3.36	1.336
20. 보존장소 (Preservation place)		1	4.5	5	4.00	1.301

21. 위치 (Location)	21.1 소장처 (location)	1	5	5	4.50	1.092
	21.2 소장위치 (storage details)	1	4.5	5	4.21	1.122
	21.3 원본의존재와위치 (Existence and location of originals)	1	5	5	4.43	1.158
	21.4 사본의존재와위치 (Existence and location of copies)	1	4	5	3.71	1.204

3) 문맥 영역

문맥 영역의 기술 통계표는 [표 32]와 같다. ‘관계유형(3.93)’의 평균값이 가장 높았으며, 다음으로 ‘관계대상식별자(3.71)’, ‘관계설명(3.5)’순으로 나타났다.

[표 32] 보존 메타데이터 - 문맥 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
22. 관계 (Relation)	22.1 관계유형 (relation type)	2	4	5	3.93	0.997
	22.2 관계대상식별자 (relation item ID)	2	3.5	5	3.71	0.994
	22.3 관계설명 (relation description)	1	3.5	5	3.50	1.160

4) 인증 영역

인증 영역의 기술 통계표는 [표 33]과 같다. 평균값이 가장 높은 요소는 ‘서명(3.64)’으로 확인되었다. ‘무결성체크법’, ‘무결성체크값’의 평균값은 각각 3.5, 3.43으로 나타났다.

[표 33] 보존 메타데이터 - 인증 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
23. 진본성확인 (Object authentication)	23.1 서명 (signature)	2	4	5	3.64	1.151
24. 무결성체크 (Integrity Check)	24.1 무결성체크법 (integrity check name)	1	4	5	3.50	1.345
	24.2 무결성체크값 (integrity check value)	1	3.5	5	3.43	1.342

4.3.2.5 이용자 참여 메타데이터 기술 통계

이용자 참여 메타데이터의 기술 통계표는 다음 [표 34]와 같이 정리하였다. ‘조회수’와 ‘좋아요수’ 요소의 평균값은 각각 4.29, 4.21로 나타났다. 나머지 요소의 평균값은 3점대로 나타났다. 그중 평균값이 가장 낮은 요소는 ‘대댓글 ID(3.21)’로 나타났다.

[표 34] 이용자 참여 메타데이터 - 이용자 참여 영역 응답 기술 통계표

상위 요소	하위 요소	최솟값	중위값	최댓값	평균값	표준 편차
25. 이용자 피드백 정보 (User feedback information)	25.1 조회수 (views)	2	5	5	4.29	0.994
	25.2 좋아요수 (number of like)	2	4.5	5	4.21	0.975
	25.3 싫어요수 (number of dislike)	1	4	5	3.71	1.267

	25.4 댓글 수 (number of reply)	1	4	5	3.79	1.311
26. 댓글 (Reply)	26.1 댓글ID (reply ID)	1	4	5	3.29	1.490
	26.2 댓글내용 (reply content)	1	3.5	5	3.29	1.590
	26.3 대댓글ID (re-reply ID)	1	3	5	3.21	1.626
	26.4 대댓글내용 (re-reply content)	1	3	5	3.29	1.684

4.3.2.6 추가 요소 및 기타 의견

14명의 전문가들 중 7명이 추가 요소 및 기타 의견에 응답하였다. 추가 요소에 대한 응답으로 나타난 요소는 채널 이름, 구독자 수, 콘텐츠 활용 연령, 삭제가 있었다. 기타 의견에 대한 응답은 다음과 같다. “디지털 영상의 보존 프로세스에 적합한 요소 추가 필요”, “상, 하위 요소의 불명확성과 요소의 필수/선택 조사 여부”가 있었다. 또한 “아카이빙 할 유튜브 영상에 대한 명확한 기준 제시”라는 의견이 있었다.

4.3.3 타당도 검증

설문결과 응답에 대한 타당도 검증은 spss v28을 통해 요인분석 방법으로 진행하였다. 14명의 전문가가 응답한 설문결과를 토대로 각 메타데이터의 하위 영역별 타당도를 확인하였다. 이를 위해 요소(변수)간의 공통요인을 추출하여 상관관계를 발견하고 각 요소의 성질을 축소하는 방법인 최대우도분석을 사용하였다. 또한 요인회전은 요인 간의 상관관계가 있다고 가정하고 회전시키는 방법인 사각 회전 방법 중 직접 오블리민 회전법을 적용하였다. 이를 통해 공통성 값이 0.4 이상이면서 성분(적재량) 값이 0.3 이상인 요소들만 추출하였다.

4.3.3.1 설명 메타데이터 타당도 검증

각 메타데이터의 타당도 검증은 기술 통계 방법과 마찬가지로 메타데이터의 하위 영역별로 확인하였다. 요인분석의 적합성은 KMO 및 Bartlett의 검정값을 기준으로 한다. 이는 각 문항에 대해 요인분석을 실시하는 것이 적절한지 확인하는 기준이 되는데 $KMO > 0.5$, Bartlett의 $p < 0.05$ 가 만족해야 바람직하다고 할 수 있다(노경섭, 2019).

설명 메타데이터의 설명 영역 요소의 타당도 검증 결과는 다음 [표 35]와 같다. ‘제목’ 요소는 기술 통계에서 확인된 것과 같이 모든 응답자들에게 5점(매우 적절)을 받아 핵심요소로써 유의미한 요소로 확인되었다. 이에 따라 ‘제목’ 요소를 제외하고 요인분석을 진행하였다. 그 결과 설명 메타데이터의 KMO는 .666, Bartlett은 .001로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 ‘태그(0.053)’를 제외하고 ‘설명’, ‘언어’, ‘범주’가 0.742, 0.874, 0.645로 공통성 기준값인 0.4 이상이였다. 추출된 하나의 요인에 대한 성분 수치는 ‘언어’, ‘설명’, ‘범주’ 순으로 0.935, 0.861, 0.803으로 나타나 기준값인 0.3을 넘는 것으로 나타났다. 반면에 ‘태그’ 요소의 성분 수치는 0.231로 나타났다. 따라서 설명 메타데이터에서 타당도 검증에서 유의한 결과로 추출된 요소는 ‘언어’, ‘설명’, ‘범주’로 나타났다.

[표 35] 설명 메타데이터 - 설명 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)
			1
4	언어	0.645	0.935
3	설명	0.742	0.861
5	범주	0.645	0.803
2.1	태그	0.053	0.231
총계			2.314
분산의 %			57.857

누적률(%)	57.857	
KMO 및 Bartlett의 검정		
표본 적절성의 Kaiser-Meyer-Olkin 측도	.666	
Bartlett의 구형성 검정	근사 카이제곱	23.971
	자유도	6
	유의수준	.001

4.3.3.2 구조 메타데이터 타당도 검증

구조 메타데이터의 구조 영역 요소들의 타당도 검증 결과는 [표 36]과 같다. KMO는 .725, Bartlett은 .000으로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 ‘URL(0.351)’, ‘장르(0.133)’, ‘비디오ID(0.207)’를 제외하고 ‘프레임비율’, ‘데이터전송속도’, ‘재생시간’, ‘압축’, ‘인코딩정보’가 0.999, 0.956, 0.578, 0.861, 0.998로 0.4 이상이였다. 공통성 기준값을 충족하는 요소들의 성분 수치를 확인결과 모두 0.3 이상으로 나타났다.

[표 36] 구조 메타데이터 - 구조 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)	
			1	2
6.2	프레임비율	0.999	1.077	-0.291
6.3	데이터전송속도	0.956	0.919	0.131
6.1	재생시간	0.578	0.762	-0.004
6.4	압축	0.861	0.761	0.312
8.2	URL	0.351	0.570	0.052
6.5	인코딩정보	0.998	0.550	0.647
7	장르	0.133	-0.051	0.381
8.1	비디오ID	0.207	0.199	0.338
총계			4.394	0.689
분산의(%)			54.924	8.607
누적률(%)			54.924	63.531
KMO 및 Bartlett의 검정				
표본 적절성의 Kaiser-Meyer-Olkin 측도			.725	

Bartlett의 구형성 검정	근사 카이제곱	87.136
	자유도	28
	유의수준	.000

4.3.3.3 관리 메타데이터 타당도 검증

1) 권리 영역

관리 메타데이터의 하위 영역 중 권리 영역의 타당도 검증 결과는 다음 [표 37]과 같다. KMO는 .398, Bartlett은 <.001로 나타났다. Bartlett의 기준값(< .005)은 충족하였지만 KMO의 기준값(>0.5)을 충족하지 못했다. 이에 따라 권리 영역은 요인분석으로 활용하기 적합하지 않은 것으로 나타났다.

[표 37] 관리 메타데이터 - 권리 영역 타당도 검증 결과

KMO 및 Bartlett의 검정		
표본 적절성의 Kaiser-Meyer-Olkin 측도		.398
Bartlett의 구형성 검정	근사 카이제곱	77.427
	자유도	36
	유의수준	<.001

2) 관리 영역

관리 영역의 타당도 검증 결과는 [표 38]과 같다. KMO는 .741, Bartlett은 .000로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 ‘등록일시(0.388)’를 제외하고 ‘관리일시’, ‘관리행위자’, ‘변경요소’, ‘생산일시’, ‘관리설명’, ‘관리유형’이 0.999, 0.872, 0.770, 0.532, 0.979, 0.849로 0.4 이상이였다. 공통성 값을 충족하는 요소들에 한해서 성분 수치를 확인결과 모두 0.3 이상으로 나타났다.

[표 38] 관리 메타데이터 - 관리 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)	
			1	2
12.3	관리일시	0.999	1.061	-0.158
12.4	관리행위자	0.872	0.830	0.194
12.5	변경요소	0.770	0.797	0.153
11.1	생산일시	0.532	0.758	-0.068
12.2	관리설명	0.979	0.581	0.579
12.1	관리유형	0.849	0.049	0.898
11.2	등록일시	0.388	-0.042	0.641
총계			3.818	1.572
분산의(%)			54.537	22.454
누적률(%)			54.537	76.991
KMO 및 Bartlett의 검정				
표본 적절성의 Kaiser-Meyer-Olkin 측도			.741	
Bartlett의 구형성 검정			근사 카이제곱	
			78.469	
			자유도	
		21		
		유의수준		
		.000		

3) 기술 영역

기술 영역의 타당도 검증 결과는 [표 39]와 같다. KMO는 .646, Bartlett은 .000으로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 ‘포맷명(0.364)’, ‘파일크기(0.375)’를 제외하고 ‘생성어플리케이션명’, ‘생성어플리케이션버전’, ‘운영시스템’, ‘포맷버전’이 0.999, 0.870, 0.920, 0.999로 0.4 이상이었다. 공통성 값을 충족하는 요소들에 한해서 성분 수치를 확인한 결과 모두 0.3 이상으로 나타났다.

[표 39] 관리 메타데이터 - 기술 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)	
			1	2
16.1	생성어플리케이션명	0.999	1.034	-0.065
16.2	생성어플리케이션버전	0.870	0.914	0.032
15	운영시스템	0.920	0.452	0.631

13.1	포맷명	0.364	-0.051	0.630
14.1	파일크기	0.375	-0.020	0.623
13.1	포맷버전	0.999	0.532	0.600
총계			3.851	0.675
분산의(%)			64.180	11.249
누적률(%)			64.180	75.429
KMO 및 Bartlett의 검정				
표본 적절성의 Kaiser-Meyer-Olkin 측도				.646
Bartlett의 구형성 검정	근사 카이제곱			81.405
	자유도			15
	유의수준			.000

4.3.3.4 보존 메타데이터 타당도 검증

1) 출처 영역

출처 영역에는 ‘수집, 이관의 직접적 출처’요소만 존재한다. 이에 따라 요인분석을 할 수 없었다.

2) 참조 영역

참조 영역 요소들은 요인분석 결과 추출되는 요소가 존재하지 않았다. 따라서 핵심 요소로써 유의한 요소는 존재하지 않은 것으로 확인되었다.

3) 문맥 영역

문맥 영역 요소들에 대한 타당도 검증 결과는 다음 [표 40]과 같다. KMO는 .696, Bartlett은 .000으로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 ‘관계대상식별자’, ‘관계설명’, ‘관계유형’이 1.000, 0.934, 0.909로 공통성 기준값인 0.4 이상이였다. 요인 회전은 하

나의 요인만 추출되어 실시할 수 없었다. 추출된 하나의 요인에 대한 성분 수치는 모두 0.3 이상으로 나타났다.

[표 40] 보존 메타데이터 - 문맥 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)	
			1	
22.2	관계대상식별자	0.999	1.000	
22.3	관계설명	0.872	0.934	
22.1	관계유형	0.826	0.909	
총계			2.697	
분산의(%)			89.898	
누적률(%)			89.898	
KMO 및 Bartlett의 검정				
표본 적절성의 Kaiser-Meyer-Olkin 측도			.696	
Bartlett의 구형성 검정	근사 카이제곱		45.520	
	자유도		3	
	유의수준		.000	

4) 인증 영역

인증 영역 요소들에 대한 타당도 검증 결과는 [표 41]과 같다. KMO는 .691, Bartlett은 .000으로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 ‘무결성체크법’, ‘서명’, ‘무결성체크값’이 1.000, 0.920, 0.895로 공통성 기준값인 0.4 이상이었다. 요인 회전은 하나의 요인만 추출되어 실시할 수 없었다. 추출된 하나의 요인에 대한 성분 수치는 모두 0.3 이상으로 나타났다.

[표 41] 보존 메타데이터 - 인증 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)	
			1	
24.1	무결성체크법	0.999	1.000	
23.1	서명	0.847	0.920	
24.2	무결성체크값	0.802	0.895	
총계			2.647	

분산의(%)	88.243	
누적률(%)	88.243	
KMO 및 Bartlett의 검정		
표본 적절성의 Kaiser-Meyer-Olkin 측도	.691	
Bartlett의 구형성 검정	근사 카이제곱	39.058
	자유도	6
	유의수준	.000

4.3.3.5 이용자 참여 메타데이터 타당도 검증

이용자 참여 메타데이터의 이용자 참여 영역 요소들의 타당도 검증 결과는 [표 42]와 같다. KMO는 .602, Bartlett은 .000으로 나타나 요인분석에 활용하기 적합한 것으로 나타났다. 각 요소의 공통성 값은 0.999, 0.798, 0.71, 0.996, 0.980, 0.849, 0.908, 0.511로 공통성 기준값인 0.4 이상이였다. 공통성 값을 충족하는 요소들에 한해서 성분 수치를 확인결과 모두 0.3 이상으로 나타났다.

[표 42] 이용자 참여 메타데이터 - 이용자 참여 영역 타당도 검증 결과

요소 번호	측정 요소명	추출 (공통성)	성분(적재량)	
			1	2
25.2	좋아요수	0.999	1.048	-0.087
25.1	조회수	0.798	0.841	0.085
25.3	싫어요수	0.871	0.531	0.518
26.4	대댓글내용	0.996	-0.067	1.036
26.3	대댓글ID	0.980	-0.023	1.003
26.1	댓글ID	0.849	0.002	0.920
25.4	댓글수	0.908	0.318	0.732
26.2	댓글내용	0.511	-0.028	0.731
총계			4.51	2.402
분산의(%)			56.377	30.029
누적률(%)			56.377	86.406
KMO 및 Bartlett의 검정				
표본 적절성의 Kaiser-Meyer-Olkin 측도			.602	

Bartlett의 구형성 검정	근사 카이제곱	146.496
	자유도	28
	유의수준	.000

4.3.4 신뢰도 검증

타당도 검증을 통해 추출된 요소들은 총 32개로 나타났으며, 이들의 내적 일관성을 알아보기 위해 메타데이별, 하위 영역별 신뢰도 검증을 실시하였다. 측정방법은 크론바흐 알파(Cronbach's α) 측정방법을 사용하였다. 크론바흐 알파값이 0.7 이상이면 신뢰도가 아주 높다고 판단하는데 사회과학분야에서는 일반적으로 0.6 이상이면 높은 값으로 본다. 통계 분석을 위한 최소한의 기준 값은 0.5이다(박태섭, 2017). 전체 결과는 다음 [표 43]과 같다. 신뢰도 검증 결과 5개 메타데이터의 7개 하위영역 모두 0.7 이상으로 아주 바람직한 요소로 확인되었다.

[표 43] 신뢰도 검증 결과

구분		신뢰도(chronbach's alpha)	
설명 메타데이터	설명 영역	0.773	0.773
구조 메타데이터	구조 영역	0.912	0.912
관리 메타데이터	관리 영역	0.795	0.919
	기술 영역	0.953	
보존 메타데이터	문맥 영역	0.958	0.933
	인증 영역	0.952	
이용자참여 메타데이터	이용자 참여 영역	0.947	0.947

4.4 최종 메타데이터 요소 도출

최종적으로 메타데이터 핵심 요소를 정리하면 다음 [표 44]와 같다. 설문 응답 결과에서 14명의 전문가에게 5점을 받은 ‘제목’ 요소를 포함하고, 타당도 검증과 신뢰도 검증을 통해 핵심 요소로써 유의한 결과가 나타난 요소들을 종합한 결과, 총 5개 메타데이터의 7개 영역(총 33개 요소)으로 나타났다.

[표 44] 유튜브 동영상 아카이빙을 위한 메타데이터 핵심 요소

메타데이터 구분	영역 구분	상위 요소	하위 요소
설명 메타데이터	설명 영역	제목	
		설명	
		언어	
		범주	
구조 메타데이터	구조 영역	비디오	재생시간
			프레임비율
			데이터전송속도
			압축
			인코딩정보
관리 메타데이터	관리 영역	일시	생산일시
		관리이력	관리유형
			관리설명
			관리일시
			관리행위자
	변경요소		
	기술 영역	포맷	포맷버전
운영시스템			
어플리케이션		생성어플리케이션명 생성어플리케이션버전	
보존 메타데이터	문맥 영역	관계	관계유형
			관계대상식별자
			관계설명
	인증 영역	진본성확인	서명
		무결성체크	무결성체크법 무결성체크값
이용자 참여 메타데이터	이용자 참여 영역	이용자피드백정보	조회수
			좋아요수
			싫어요수
			댓글 수

		댓글	댓글ID
			댓글내용
			대댓글ID
			대댓글내용

다음 [표 45]는 연구자가 설계했던 메타데이터 측정 요소와 핵심 요소로써 유의한 결과로 나타난 요소들을 비교하여 정리한 것이다. 메타데이터별로 측정 요소 현황을 살펴보면 첫째, 설명 메타데이터의 경우 기존 5개의 측정 요소 중 기술 통계에서 14명의 전문가에게 모두 5점을 받은 ‘제목’ 요소와 타당도 및 신뢰도 검증에서 3개 요소가 추출되어 총 4개 측정 요소가 핵심 요소로 나타났다. 둘째, 구조 메타데이터의 경우 기존 8개의 측정 요소 중 5개 측정 요소가 핵심 요소로 나타났다. 셋째, 관리 메타데이터의 경우 기존 3개 영역(권리, 관리, 기술)의 22개 측정 요소 중 2개 영역(관리, 기술)의 10개 측정 요소가 핵심 요소로 나타났다. 보존 메타데이터의 경우 기존 4개 영역(출처, 참조, 문맥, 인증)의 17개 측정 요소 중 2개 영역(문맥, 인증)의 6개의 측정 요소가 핵심 요소로 나타났다. 다섯째, 이용자 참여 메타데이터의 경우 기존 8개 측정 요소 모두 핵심 요소로 나타났다.

[표 45] 기존 메타데이터 요소와 핵심 요소 결과 비교

메타데이터 구분	기존 메타데이터 영역별 구분 (측정 요소 수)	핵심 메타데이터 영역별 구분 (측정 요소 수)	채택률	
			채택률	채택률
설명 메타데이터	설명 영역(5)	설명 영역(4)	80%	80%
구조 메타데이터	구조 영역(8)	구조 영역(5)	63%	63%
관리 메타데이터	권리 영역(9)	-	0%	45%
	관리 영역(7)	관리 영역(6)	86%	
	기술 영역(6)	기술 영역(4)	67%	
보존 메타데이터	출처 영역(1)	-	0%	33%
	참조 영역(13)	-	0%	
	문맥 영역(3)	문맥 영역(3)	100%	
	인증 영역(3)	인증 영역(3)	100%	

이용자 참여 메타데이터		이용자 참여 영역(8)	이용자 참여 영역(8)	100%	100%
계	메타데이터 수	5	5	100%	
	정보 수	10	7	70%	
	측정요소 수	63	33	53%	

이를 종합하면 연구자가 설계한 5개 메타데이터의 10개 영역의 63개 측정 요소 중 핵심 요소로 나타난 측정 요소는 5개 메타데이터 7개 영역의 33개 측정 요소로 확인되었다. 메타데이터, 영역, 측정요소별 채택률은 각각 100%, 70%, 53%로 나타났다. 기존 요소와 핵심 요소의 구분 표는 [부록 2]를 통해 확인할 수 있다.

V. 결 론

본 연구는 공공기관에서 운영하는 유튜브 동영상에 대한 관리 및 보존의 필요성에 주목하여 아카이빙 메타데이터를 설계하고 핵심 요소를 제안하는 것을 목적으로 하였다. 이를 위해 연구절차는 메타데이터 요소 설계와 메타데이터 요소 평가로 나누어 진행하였다. 메타데이터 요소 설계에서는 국내외 기록, 보존, 영상 분야의 메타데이터 표준과 유튜브에 영상을 등록하고 이용 시 생성되는 정보(유튜브 메타데이터)를 조사하였다. 다음으로 아카이빙 메타데이터의 구조를 토대로 조사한 표준들을 매핑하여 유튜브 동영상 아카이빙을 위한 메타데이터를 설계하였다. 메타데이터 요소 평가에서는 메타데이터 전문가를 대상으로 온라인 설문조사를 실시하였다. 이를 통해 설계한 메타데이터의 타당도와 신뢰도 검증에서 유의한 결과로 나타난 요소를 핵심 요소로 선정하였다.

본 연구에서 전문가를 선정한 방법은 디지털 아카이빙 및 메타데이터 관련 실무 또는 연구 경험의 유무로 설정하였다. 설문지는 1. 일반현황(개방형, 폐쇄형 질문), 2. 메타데이터 요소 평가(5점 리커트), 3. 추가 요소 및 기타 의견(개방형 질문)으로 구성하였다. 본 연구는 설문 결과에 대해 기술 통계 분석과 타당도와 신뢰도 검증을 실시하였다. 기술 통계 분석은 빈도분석과 평균분석을 통해 분석하였고, 타당도는 KMO 및 Bartlett의 구형성 검증과 요인 분석을 통해 검증하였으며, 신뢰도 검증은 크론바흐 알파 측정값을 통해 검증하였다.

본 연구의 주요 결과를 절차별로 설명하면 다음과 같다.

첫째, 국내외 표준을 매핑하여 본 연구에서 설계한 메타데이터 요소는 5개 메타데이터 10개 하위 영역의 63개 요소로 구성되어있다. 5개 메타데이터 영역은 설명 메타데이터, 구조 메타데이터, 관리 메타데이터, 보존 메타데이터, 이용자 참여 메타데이터로 구분하였다. 또한 설명 메타데이터는 설명 영역, 구조 메타데이터는 구조 영역, 관리 메타데이터는 권리, 관리, 기술 영역, 보존 메타데이터는 출처, 참조, 문맥, 인증 영역, 이용자 참여 메타데이터는

이용자 참여 영역으로 세분화 하였다.

둘째, 메타데이터 요소 평가를 위한 설문지는 15부를 배포하여 14부(92%)를 회수하였다. 또한 설문지 문항 중 ‘1.1 소속기관 유형’에 대해 정제 작업을 실시하였다.

셋째, 각 응답에 대한 기술 통계 분석 결과는 다음과 같다. 먼저 일반 현황에 대한 결과는 빈도분석을 실시하였다. 그 결과 소속기관 유형에 대한 응답은 공공기관 9명(64%), 기타기관 4명(29%), 대학 1명(7%)이 응답하였다. 최종학력에 대한 응답은 박사(수료 포함) 6명(42.86%), 석사 4명(28.57%), 학사 4명(28.57%) 이었다. 경력 사항의 경우 응답하지 않은 1명을 제외하고 13명으로 분석하였다. 그 결과 0~5년이 3명(23%), 6~10년이 4명(31%), 11~15년이 2명(15%), 그 이상이 4명(31%)로 나타났다.

다음은 메타데이터 하위 영역별 평균 분석을 실시한 결과이다. 설명 메타데이터의 설명 영역 요소의 ‘제목’ 요소는 5개 메타데이터의 전체 요소 중 모든 응답자들에게 5점(매우 적절)을 받은 유일한 요소로 나타났다. 구조 메타데이터의 구조 영역 요소 중 ‘재생시간’의 평균값이 가장 높았으며, 다음으로 ‘URL’, ‘프레임 비율’, ‘비디오 ID’ 순으로 나타났다. 평균값이 가장 낮은 요소는 ‘장르’로 나타났다. 관리 메타데이터의 권리 영역, 관리 영역, 기술 영역의 응답을 종합한 결과는 다음과 같다. ‘공개 구분’ 요소의 평균값이 관리 메타데이터의 전체 하위 영역 요소들 중 가장 높았으며 다음으로 ‘생산일시’, ‘등록일시’, ‘파일크기’, ‘포맷명’ 순으로 나타났다. ‘관리 행위자’ 요소는 5개 메타데이터 전체 요소 중 평균값이 가장 낮은 요소로 나타났다. 보존 메타데이터의 출처 영역, 참조 영역, 문맥 영역, 인증 영역의 응답을 종합한 결과는 다음과 같다. ‘소장처’ 요소의 평균값이 보존 메타데이터의 전체 하위 영역 요소들 중 가장 높았으며 다음으로 ‘원본의 존재와 위치’, ‘기관명’, ‘소장 위치’ 순으로 나타났다. 마지막 이용자 참여 메타데이터의 평균값은 ‘조회수’, ‘좋아요수’ 순으로 높게 나타났으며, ‘대댓글 ID’ 요소의 평균값은 해당 영역 내에서 가장 낮게 나타났다.

추가 요소 및 기타 의견에 대한 응답으로 7명의 전문가가 응답하였다. 추가 요소 응답에서는 “채널 이름”, “구독자 수”, “콘텐츠 활용 연령”, “삭제

요소라고 응답하였으며, 기타 의견 응답에서는 “보존 메타데이터의 추가 요소 필요”, “상, 하위 요소의 불명확성과 필수/선택 조사 여부”, “아카이빙 기준 제시”라고 응답하였다.

넷째, 5개의 메타데이터의 10개 하위 영역별로 타당성 검증을 진행한 결과는 다음과 같다. 우선 하위 영역별로 KMO 및 Bartlett의 구형성 검증을 실시하여 요인분석을 수행하는 것이 적합한지 확인하였다. 그 결과, 7개 하위 영역(설명, 구조, 관리, 기술, 문맥, 인증, 이용자 참여)이 기준값인 $KMO > 0.5$, Bartlett의 $p < 0.05$ 를 충족하였다. 반면에 권리 영역은 기준값을 충족하지 못해 요인분석을 수행하는데 적합하지 않는 것으로 나타났다. 출처 영역과 참조 영역은 요인분석이 실시되지 않았다.

다음으로 7개 하위 영역에 대해 요인분석을 실시한 결과는 다음과 같다. 요인분석을 통해 유의한 요소로 판단할 수 있는 기준은 추출 공통성 값이 0.4 이상이며, 초기 고유값이 1요인 이상이면서 성분 수치가 0.3 이상인 경우이다. 설명 메타데이터의 설명 영역의 경우 ‘제목’ 요소를 제외하고 요인분석을 실시한 결과 ‘태그’를 제외한 3개 요소가 기준값을 충족하는 것으로 나타났다. 구조 메타데이터의 구조 영역의 경우 5개 요소가 요인분석 기준값을 충족하였다. 관리 메타데이터의 관리, 기술 영역의 경우 각각 6개, 4개 요소의 요소가 기준값을 충족하였다. 보존 메타데이터의 문맥, 인증 영역과 이용자 참여 메타데이터의 이용자 참여 영역은 측정된 모든 요소가 추출 기준값을 충족하였으며 각각 3개, 3개, 8개 요소로 나타났다.

다섯째, 신뢰도 검증은 메타데이터별, 하위 영역별로 실시하였다. 그 결과 모두 0.7 이상으로 나타났다. 본 연구에서 설계한 메타데이터는 5개 메타데이터의 10개 하위 영역, 63개 측정 요소로 구성되었다. 그중 타당성 검증과 신뢰도 검증을 통해 핵심 요소로써 유의한 결과가 나타난 요소는 요인분석 과정에서 제외시킨 ‘제목’ 요소를 포함하여 총 5개 메타데이터의 7개 하위 영역, 33개 측정 요소로 나타났다. 각 메타데이터별, 하위 영역별, 요소별 채택률은 각각 100%, 70%, 53%이다.

본 연구에서는 메타데이터 핵심 요소 도출을 위해서 요인분석을 수행하였다. 하지만 요인분석 결과가 핵심 요소의 여부를 가르기에는 한계가 존재한

다. 또한 본 연구는 소셜미디어 메타데이터 요소도출을 지향하지만 대상은 대표적인 서비스인 유튜브로 한정하였다. 유튜브의 규모가 크기는 하지만 전체 동영상 소셜미디어를 대표한다고 보기는 어려울 수도 있다. 본 연구는 소셜미디어 동영상 아카이빙 메타데이터의 초기 연구로서 의의를 가지고 있으며, 선행연구가 부족한 상황 속에서 소셜미디어 기록과 동영상 기록의 특징이 결합된 유튜브 기록에 주목했다는 점에서도 다른 메타데이터 연구와 차별화를 가진다고 볼 수 있다. 향후 다양한 후속 연구들이 이루어지고 소셜미디어 메타데이터 대한 요소들이 다양화되면 소규모 전문가 집단을 활용한 요인분석보다는 대규모 집단을 대상으로 한 확인적 요인분석 등 방법이 가능하리라 본다.

모든 기록관리의 궁극적인 목적은 기록을 활용하는 것으로 귀결된다. 유튜브를 포함한 소셜미디어 기록 역시 아카이빙을 통해 이용자에게 서비스를 제공하고, 수집된 데이터를 통해 유의미한 현상을 파악함으로써 해당 기록의 가치를 높일 수 있다. 본 연구는 유튜브 동영상 기록물의 아카이빙에서 메타데이터에 주목했으며, 후속 연구를 통해 다양한 연구가 진행되기를 바란다.

참 고 문 헌

1. 국내 문헌

- 강민정. (2021). O AIS 참조모형 기반 민증가요 기록관리 모형에 관한 연구. 석사학위논문. 전남대학교 기록관리협동과정.
- 강병서, 김계수. (2009). 「(SPSS 17.0)사회과학 통계분석」. [서울]: 한나래.
- 곽성호. (2012). 국내 방송 메타데이터 표준화 모델 연구. 석사학위논문. 서강대학교 대학원.
- 국가기록원. (2005). 전자기록물 영구보존 기반기술 연구 프로젝트 보고서
- 김아랑. (2020). 공공기관 유튜브 PR사례 연구. 석사학위논문. 고려대학교 언론대학원.
- 김정심. (2008). 행정부처 웹사이트의 아카이빙 방안에 관한 연구. 석사학위논문. 중앙대학교 대학원.
- 김태영, 양동민, 최상기, 오효정. (2019). 소셜미디어 아카이빙을 위한 수집 및 보존방안. 『정보관리학회지』. 36.2 : 79-104.
- 김혜영. (2010). 웹 기록 보존 메타데이터 요소 설계에 관한 연구. 석사학위논문. 명지대학교 기록정보과학전문대학원.
- 나스미디어. (2021). NPR 2021 인터넷 이용자 조사.
- 남태우, 이승민. (2010). 메타데이터의 의미론적 확장에 관한 연구. 『한국문헌정보학회지』. 44.4 : 373-393.
- 남태우, 이승민. (2018). 「정보자원의 기술과 메타데이터」. 서울: 한국도서관협회.
- 노경섭. (2019). 「제대로 알고 쓰는 논문 통계분석 SPSS & AMOS」. 서울 : 한빛아카데미
- 노영희. (2014). 「메타데이터 이론과 실제」. 서울 : 조은글터.
- 류지연. (2019). 유튜브 교육용 동영상 검색 개선을 위한 메타데이터 요소 설계. 석사학위논문. 명지대학교 대학원.

- 박성미. (2002). 기록보존서로서의 영상실록 아카이브. 『역사민속학』, 0(14), 21-41.
- 박신애. (2010). TV 뉴스 아카이빙을 위한 표준 메타데이터 요소 설계. 석사학위논문. 한국외국어대학교 대학원.
- 박옥남. (2012). PREMIS 기반 보존 메타데이터 요소 개발에 관한 연구 - 국립중앙도서관 디지털 자료를 중심으로 -. 『한국문헌정보학회지』. 46.2 : 83-113.
- 박유경. (2014). 공공기관 소셜미디어의 메타데이터 표준 요소 분석. 석사학위논문. 경북대학교 대학원.
- 박인우. (2018). 스포츠 동영상 기록물 메타데이터에 관한 연구. 석사학위논문. 중앙대학교 대학원.
- 박태섭. (2017). IPA를 활용한 지방자치단체 기록연구사의 직무분석에 관한 연구. 석사학위논문. 한성대학교 대학원.
- 박희진. (2020). 디지털 보존을 위한 PREMIS 3.0 기반 메타데이터에 관한 연구. 『한국기록관리학회지』. 20.4 : 185-203.
- 배소연. 애니메이션 IP를 활용한 공공기관의 캠페인 애니메이션 사례 연구. 석사학위논문. 중앙대학교 첨단영상대학원.
- 봉지현. (2019). 일본군 '위안부' 구술 기록의 관리를 위한 메타데이터 요소 선정에 관한 연구. 석사학위논문. 중앙대학교 대학원.
- 서은경. (2005). 디지털자원의 보존 메타데이터 요소세트 개발에 관한 연구. 『정보관리학회지』. 22.3 : 233-260.
- 설문원. (2005). 기록의 품질 기준 분석 - 진본성, 신뢰성, 무결성, 가용성을 중심으로 -. 『기록학연구』. 0(11): 41-89.
- 송주형. (2012). 기록관리 대상으로서 SNS 연구: 페이스북, 트위터, 블로그, 유튜브를 중심으로. 석사학위논문. 한국외국어대학교 대학원.
- 송지준. (2008). 「논문작성에 필요한 SPSS/AMOS 통계분석방법」. 파주: 21세기사.
- 심보미. (2018). 문화 분야 공공영상 아카이브의 현황과 시사점. 『도서관』. 392.45 : 117-139.

- 안영희, 박옥화. (2009). 이러닝 콘텐츠 아카이빙 구축을 위한 메타데이터 요소에 관한 연구. 『한국문헌정보학회지』. 43.3 : 147-162.
- 양혜정. (2009). 국제표준기록물기술규칙(ISAD(G))의 1판과 2판에 관한 비교 연구. 석사학위논문. 중앙대학교 대학원.
- 오상훈, 최영선. (2009). ISO 14721 OAIS 참조모형을 활용한 웹 아카이빙의 메타데이터 구조 및 요소 정의. 『정보처리학회논문지』. 16.5 : 651-660.
- 유종연. (2020). 독도 고지도의 기술요소 개선방안에 관한 연구. 석사학위논문. 이화여자대학교 정책과학대학원.
- 유효립. (2007). 정부부처 웹 아카이빙 방안 연구. 석사학위논문. 명지대학교 기록과학대학원.
- 윤선정. (2008). 게임 기반 LCMS를 위한 교육용 게임 콘텐츠 메타데이터 개발. 박사학위논문. 동서대학교 디자인&IT전문대학원.
- 윤승욱, 장준갑, 김건. (2019). SNS 게시물에 대한 공공기록화 수용에 관한 연구. 『디지털융복합연구』, 17.9 : 1-12.
- 이귀영. (2015). 중앙행정기관의 소셜미디어 아카이브 도입 방안에 관한 연구. 석사학위논문. 이화여자대학교 정책과학대학원.
- 이승호. (2022). 개인수집기록물 메타데이터 요소 설계 방안. 석사학위논문. 경북대학교 대학원.
- 이인혁. (2017). 복합전자기록물 아카이빙을 위한 메타데이터에 관한 연구. 석사학위논문. 한성대학교 대학원.
- 이창민. (2015). 다중 출처 기반의 공연예술자료 기술에 관한 연구. 석사학위논문. 한성대학교 대학원.
- 이혜은. (2013). 고지도 정보시스템 구축을 위한 메타데이터 요소 설계. 박사학위논문 숙명여자대학교. 대학원
- 임다은. (2017). SNS 기록의 수집방안 연구: '강남역 여성혐오 살해사건' 관련 트위터 기록을 중심으로. 석사학위논문. 한국외국어대학교 대학원.
- 장보성. (2011). 디지털 연구성과물 장기보존을 위한 메타데이터에 관한 연구. 박사학위논문. 중앙대학교 대학원.

- 장운석. (2019). 유튜브 정치 동영상 채널의 기능, 그리고 유튜브 정치 동영상의 유권자의 정치사회화에 미치는 영향에 관한 정치인의 인식 연구. 석사학위논문. 한국외국어대학교 정치행정언론대학원.
- 정성숙. (2012). 오마이뉴스 아카이빙 연구. 석사학위논문. 한남대학교 대학원.
- 정혜정. (2018). SNS와 민원에 기반한 기록정보서비스 활성화에 관한 연구. 석사학위논문. 명지대학교 기록정보과학 전문대학원.
- 천현순. (2011). 문화산업으로서의 영화에 대한 비판적 고찰. 『카프카연구』. 0.26 : 181-202.
- 최동희. (2021). 지방자치단체의 유튜브 콘텐츠 특성이 지역 이미지에 미치는 영향에 관한 연구. 『관광경영연구』. 103.0 : 567-583.
- 최두원, 이수진, 윤은하, 오현정. (2018). 대통령 SNS기록물 관리 방안에 관한 연구. 『한국기록관리학회지』. 18.2 : 29-59.
- 최민재, 양승찬, 김강석. (2009). 「인터넷 소셜미디어와 저널리즘」. 서울: 한국언론재단.
- 최윤정, 차승준, 이규철. (2009). 공공기관의 심층 웹 기록물 아카이빙을 위한 메타데이터 개발. 『한국정보과학회 학술발표논문집』. 36.2C : 87-91.
- 최지윤, 정운재. (2017). 뷰티 인플루언서 마케팅 활용 전략 : 매스미디어와 소셜미디어의 비교를 중심으로. 『광고학연구』. 28.4 : 47-72.
- 최지혜. (2021). 유튜브 채널 운영을 통한 장애인의 사회참여에 관한 사례 연구. 석사학위논문. 대구대학교 대학원.
- 최진호. (2017). 기록관의 동영상 기록물 기술에 관한 연구. 석사학위논문. 한남대학교대학원.
- 케빈 알로카(Kevin Alloca), 엄성수 옮김. (2018). 「유튜브컬처」. 서울: 스타리치 북스.
- 태보라. (2020). 팽수콘텐츠의 담론과 사회문화적 가치: 유튜브 채널 <자이언트팽TV> 를 중심으로-. 『영상문화콘텐츠연구』. 20.0 : 39-66.
- 한국기록학회. (2008). 「기록학 용어 사전」. 경기: 역사비평사.

- 한국영상자료원. (2003). 영상정보 메타데이터(Metadata) 표준(안) 연구.
- 한규린. (2021). 유튜브 기록의 수집방안에 관한 연구. 석사학위논문. 한국외국어대학교 대학원.
- 한희정, 오효정, 양동민. (2020). 전자기록물의 장기보존을 위한 보존포맷 선정 방안에 관한 연구. 『한국기록관리학회지』 20(1): 69-87.
- 황운영, 장인호, 이규철. (2012). 정부부처 페이스북 페이지 기록물 장기보존을 위한 장기보존 메타데이터 및 장기보존포맷. 『한국정보과학회 학술발표논문집』. 39.2C : 61-63.
- 황혜림. (2018). 공개교육자원 관리와 활용을 위한 Open Course Ware(OCW) 메타데이터 스키마 개발 연구. 석사학위논문 성균관대학교 대학원.

2. 국외 문헌

- Boyd Danah M, Ellison Nicole B. (2007). SocialNetworkSites :Definition, History, and Scholarship. JournalofComputer- Mediated Communication. 13(1), 210-230.
- Gail Hodge. (2005). Metadata for Electronic Information Resources: From Variety to Interoperability. Information Services & Use, 25(1), 35-45.
- Harriette Hemmasi. (2002). Why not MARC?. In Proceedings of the Third International Conference on Music Information Retrieval, October, 13-17, 2002, Paris, France, P.243.
- Hong SH. The criteria for selecting appropriate fit indices in structuralequation modeling and their rationales. Korean J Clin Psychol 2000;19:161-177.
- Horton, D. & Wohl, R.R. (1956). Mass communication and parasocial interaction: Observations on intimacy at a distance. Psychiatry, 19(3), 215-229.
- Kääriäinen, M., Kanste, O., Elo, S., Pölkki, T., Miettunen, J., &Kyngäs,

- H. (2011). Testing and verifying nursing theory by confirmatory factor analysis. *Journal of Advanced Nursing*, 67, 1163–1172
- Kara Van Malssen. (2016). Digital Preservation Metadata Practice for Audio–Visual Materials. In Kara Van Malssen. (Eds.), *Digital Preservation Metadata for Practitioners – Implementing PREMIS*. 45–58.
- Jenn Riley. (2017). *Understanding Metadata: What is Metadata, and What is it For?: A Primer*. NISO Press.
<https://groups.niso.org/higherlogic/ws/public/download/17446/Understanding%20Metadata.pdf>.
- OCLC/RLG Working Group. (2001). *Preservation Metadata for Digital Objects: A Review of the State of the Art*.
https://www.oclc.org/content/dam/research/activities/pmwg/presmeta_wp.pdf.
- OCLC/RLG Working group. (2002). *Preservation Metadata and the OAIS Information Model : A Metadata Framework to Support the Preservation of Digital Objects*.
https://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf.

3. 법령 및 표준

- International Council on Archives. (2000). *ISAD(G)2 : General. International Standard Archival Description 2nd ed.*
<https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf>.
- PBCore. (n.d.). <https://pbcore.org/elements/element-hierarchy>.
- PREMIS Editorial Committee. (2015). *Data Dictionary for Preservation Metadata version 3.0*,
<https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.

- 공공기록물 관리에 관한 법률. (2020. 6. 4.).
- 공공기관의 운영에 관한 법률 시행령. (2022. 3. 29.).
- 국가기록원. (2021). 공공기관 기록물관리 지침.
- 국가기록원. (2021). 기록관리 메타데이터 표준 : 일반(NAK8:2021(v2.2)).
- 국가기록원. (2008). 전자기록물 문서보존포맷 기술규격(NAK 30:2008(v1.0)).
- 국가기록원. (2020). 전자기록물 장기보존패키지 기술규격-제1부 XML로 포맷화된 방식(NEO2)(NAK 31-1:2020(v2.2)).
- 국가기록원. (2020). 전자기록물 장기보존패키지 기술규격 - 제2부 디렉토리로 구조화된 방식(NEO3)(NAK 31-2:2020(v1.0)).
- 국가기록원. (2016). 문헌정보 — 기록관리 — 제1부: 개념과 원칙(KS X ISO15489-1).

4. 웹 페이지

- YouTube for Press. (2022). <https://blog.YouTube/press/>.
- 공공기관의 현황. (2022). <https://www.alio.go.kr/guide/publicAgencyList.do>
- 국내 10명 중 8명 유튜브 쓴다… 월평균 30시간. (2020).
<http://www.mediatoday.co.kr/news/articleView.html?idxno=210934>.
- 유튜브·넷플릭스 추천 알고리즘 원리. (2020).
<https://www.etnews.com/20201005000062>.

부 록 1

공공기관 유튜브 동영상 아카이빙을 위한 메타데이터 요소 설계를 위한 설문

안녕하십니까?

저는 한성대학교 대학원 문헌정보학과 기록관리전공 석사과정에 재학 중인 김락근입니다.

본 설문은 공공기관의 유튜브 동영상 기록물을 아카이빙 하기 위한 핵심 메타데이터 요소 설계하는 논문을 위한 것입니다. 본 연구는 유튜브 동영상 아카이빙 메타데이터를 설계하기 위해 NAK 8, PREMIS, ISAD(G), PBCore, 유튜브를 대상으로 설명, 구조, 관리, 보존, 이용자 참여별로 메타데이터 요소를 도출하였습니다. 본 설문은 이렇게 도출한 메타데이터에 대한 타당성을 평가받고자 하는 것입니다.

설문지는 제안된 메타데이터 요소의 타당도를 5점 척도로 평가합니다. 본 설문지는 연구목적 이외의 다른 용도로 사용되지 않으며, 통계법 33조에 의거하여 개인정보와 설문내용의 비밀이 보장됩니다. 본 조사에 관한 문의 및 기타 궁금한 점이 있으실 경우, 아래의 연락처로 문의해 주시기 바랍니다. 귀중한 시간을 내어주신 선생님께 깊이 감사드립니다.

한성대학원 문헌정보학과 기록관리전공 김락근 올림
(dnrm123@naver.com, 010-4662-9246)

지도교수: 한성대학교 도서관정보문화트랙 박진호 교수

1. 응답자 일반현황

응답자 현황 정보를 파악하기 위한 기본 정보만 조사합니다.

1.1 응답자 소속기관 유형: _____(예, 대학, 공공기관, 도서관 등)

1.2 응답자 최종 학위: _____(예, 학사, 석사, 박사 등)

1.3 응답자 경력 사항: 1) 0~5년, 2) 6~10년, 3) 11~15년, 4) 그 이상

2. 메타데이터 요소 타당성 평가

<유튜브 동영상 아카이빙을 위한 메타데이터 요소 설계>의 요소들에 대한 타당도 조사를 위해 적합성 여부를 5점 척도로 평가해 주십시오.

* 1점: 매우 부적절, 2점: 부적절, 3점: 보통, 4점: 적절, 5점: 매우 적절

2.1 설명 메타데이터

설명 영역							
요소	요소 설명	평가 척도(5점)					출처
		1	2	3	4	5	
1. 제목	유튜브 동영상의 공식 제목						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
2. 주제	유튜브 동영상에 포함된 핵심적인 내용을 태그를 통해서 간단하고 명료하게 기술						NAK 8, PBCore
2.1 태그	유튜브 동영상의 핵심 내용을 간결하게 표현해 줄 수 있는 단어						유튜브
3. 설명	유튜브 동영상의 내용 또는 동영상에 대한 자유로운 설명						NAK 8, ISAD(G), PBCore, 유튜브
4. 언어	유튜브 동영상 속의 기본 언어 식별						ISAD(G), PBCore, 유튜브
5. 범주	유튜브 동영상의 시간적, 공간적 범주						ISAD(G), PBCore

2.2. 구조 메타데이터

구조 영역							
요소	요소 설명	평가 척도(5점)					출처
		1	2	3	4	5	
6. 비디오	유튜브 동영상 구조에 대한 하위요소 집합						PBCore
6.1 재생 시간	유튜브 동영상의 전체 길이 즉, 재생 시간						PBCore, 유튜브
6.2 프레임 비율	유튜브 동영상의 트랙의 너비와 높이를 측정한 값 (예시 : 320x240)						PBCore, 유튜브
6.3 데이터 전송속도	유튜브 동영상의 디지털 미디어 파일의 데이터 양 (예시 : 56 kbps)						PBCore, 유튜브
6.4 압축	유튜브 동영상 또는 데이터 스트림에 대해 압축을 하거나 압축을 풀 수 있는 하드웨어 또는 소프트웨어 혹은 이것에 대한 알고리즘						PBCore, 유튜브
6.5 인코딩 정보	유튜브 동영상을 재생할 때 Coder를 사용하여 변환된 신호						PBCore, 유튜브
7. 장르	동영상 속의 내용 정보에 대한 유형						PBCore, 유튜브
8. 식별자	유튜브 동영상에 부여된 고유한 식별자, 혹은 검색이나 이용을 위해 추가적으로 생성된 식별자						NAK 8, PREMIS, ISAD(G), PBCore
8.1 비디오 ID	특정 유튜브 동영상의 고유 정보 ID						유튜브
8.2 URL	특정 유튜브 동영상의 웹상에서의 고유 위치 정보						NAK 8, PREMIS, ISAD(G), PBCore

2.3. 관리 메타데이터

관리 영역							
요소	요소 설명	평가 척도(5점)					출처
		1	2	3	4	5	
9. 생산자	유튜브 동영상을 생산하고 기관 유튜브 채널에 업로드 한 개인 또는 기관을 식별할 수 있는 정보						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
9.1 생산자 유형	유튜브 동영상을 생산한 주체						NAK 8, PREMIS, PBCore
9.2 생산자 이름	유튜브 동영상 생산한 주체가 소속된 기관의 이름 혹은 개인의 이름						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
10. 권한	유튜브 동영상의 접근과 이용을 관리하고 통제하기 위한 권한 정보로 이루어진 요소들의 집합						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
10.1 권한 식별자	유튜브 동영상에 적용되는 권한 정보들을 고유하게 식별할 수 있는 정보						PREMIS
10.2 권한 근거	식별된 권한 또는 권한에 대한 근거를 지정하는 역할						PREMIS, PBCore
10.3 공공 저작물 여부	저작권법 제 24조 2항에 따른 유튜브 동영상이 공공저작물인지 아닌지에 대한 여부						NAK 8
10.4 공공 저작물 자유이용 허락표시 유형	공공저작물에 부여된 공공저작물 자유 이용 허락표시(공공누리)의 유형						NAK 8
10.5 공개	유튜브 동영상을 공개 또는 비공개 하는지에 대한 여부						NAK 8, 유튜브

	구분							
	10.6 비공개 사유	동영상을 비공개로 설정한 것에 대한 사유						NAK 8
	10.7 시청 등급	유튜브 동영상을 시청할 수 있는 연령에 대한 정보						ISAD(G), PBCore, 유튜브
관리 영역								
	11. 일시	유튜브 동영상의 생산과 업로드한 행위가 발생한 일시						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
	11.1 생산 일시	공공기관에서 유튜브에 해당 동 영상을 업로드 한 일시						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
	11.2 등록 일시	아카이브에 특정 유튜브 동영상 을 등록한 일시						NAK 8, PREMIS, ISAD(G), PBCore, 유튜브
	12. 관리 이력	유튜브 동영상의 영향을 미치는 모든 관리행위에 대한 이력정보						NAK 8, PREMIS
	12.1 관리 유형	유튜브 동영상에 영향을 미친 관리행위의 유형						NAK 8, PREMIS
	12.2 관리설명	‘12.1 관리 유형’에서 기술된 유 형이 발생한 사유						NAK 8, PREMIS
	12.3 관리일시	유튜브 동영상의 상태에 영향을 미치는 관리행위에 대한 발생 일시						NAK 8, PREMIS
	12.4 관리 행위자	유튜브 동영상의 상태에 영향을 미치는 관리 행위를 수행한 행 위자						NAK 8, PREMIS
	12.5 변경 요소	유튜브 동영상의 상태에 영향을 미치는 관리행위를 통해서 변경 된 사항						NAK 8, PREMIS
기술 영역								
	13. 포맷	유튜브 동영상의 데이터 포맷 종류						NAK 8, PREMIS
	13.1 포맷명	유튜브 동영상의 데이터 파일 확장자						NAK 8, PREMIS

13.2 포맷버전	유튜브 동영상의 파일 포맷의 버전의 정보							NAK 8, PREMIS
14. 크기	유튜브 동영상의 규모에 대한 정보							NAK 8, PREMIS, ISAD(G), PBCore
14.1 파일크기	유튜브 동영상 파일에 대한 크기							NAK 8, PREMIS, ISAD(G), PBCore
15. 운영 시스템	유튜브 동영상을 이용하기 위해 요구되고 권장되는 운영체계에 대한 소프트웨어 또는 하드웨어							PREMIS
16. 어플리케이션	유튜브 동영상을 실행하기 위한 응용 프로그램에 대한 정보							NAK 8, PREMIS
16.1 생성 어플리케이션 명	유튜브 동영상 파일을 생성할 수 있는 어플리케이션 이름							NAK 8, PREMIS
16.2 생성 어플리케이션 버전	유튜브 동영상 파일을 생성한 어플리케이션 버전 정보							NAK 8, PREMIS

2.4. 보존 메타데이터

출처 영역							
요소	요소 설명	평가 척도(5점)					출처
		1	2	3	4	5	
17. 수집, 이관의 직접적 출처	아카이브에서 수집 또는 이관을 한 출처에 대한 상황을 기록						ISAD(G)
참조 영역							
18. 생산이력	유튜브 동영상의 생산과정에서 발생한 이력 정보						NAK 8
18.1 기관명	유튜브 동영상을 생산한 기관						NAK 8
18.2	유튜브 동영상의 생산 행위 과						NAK 8

	의견	정 중 생산한 기관에서 유튜브 동영상에 대해 입력한 의견						
	18.3 처리상태	유튜브 동영상을 생산한 기관이 해당 동영상을 처리한 상태						NAK 8
	18.4 처리일시	유튜브 동영상을 생산한 기관이 해당 동영상을 처리한 일시						NAK 8
	19. 보존이력	유튜브 동영상이 아카이브로 등록된 이후 행해진 모든 보존처리행위에 대한 이력정보						NAK 8, PREMIS, ISAD(G)
	19.1 보존처리 유형	유튜브 동영상의 원형을 그대로 유치한 채 수명을 연장시키기 위해 취한 행위						NAK 8, PREMIS, ISAD(G)
	19.2 보존처리 설명	유튜브 동영상의 상태에 대한 정보, 보존처리 사유 및 근거에 대한 상세한 설명						NAK 8, PREMIS, ISAD(G)
	19.3 보존처리 일시	유튜브 동영상에 보존처리 행위가 취해진 일시						NAK 8, PREMIS, ISAD(G)
	19.4 보존처리 행위자	해당 유튜브 동영상에 보존 행위를 행한 행위자						NAK 8, PREMIS, ISAD(G)
	20. 보존장소	유튜브 동영상이 최종적으로 보존되는 장소						NAK 8, PREMIS
	21. 위치	아카이브에서 유튜브 동영상의 소장위치에 대한 정보						NAK 8, PREMIS, PBCore, ISAD(G)
	21.1 소장처	유튜브 동영상의 원본 또는 저장 매체, 보존매체 등이 보존되어 있는 기관의 명칭						NAK 8, PREMIS, PBCore
	21.2 소장 위치	유튜브 동영상이 저장된 기관의 서고 정보와 배열된 위치를 나타내는 스토리지 또는 매체번호, 매체가 배열된 위치						NAK 8, PREMIS, PBCore
	21.3 원본의 존재와 위치	유튜브 동영상에 대한 사본이 존재하는 경우, 소장처와 이용가능성 등을 제공						ISAD(G)
	21.4 사본의 존재와 위치	유튜브 동영상의 복사본의 존재와 소장처, 이용가능성을 제공하는 요소						ISAD(G)
문맥 영역								
	22. 관계	유튜브 동영상과 다른 유튜브 동영상간의 관계, 사본, 참						NAK 8,

		조 관계 등 다양한 관계							PREMIS, PBCore, ISAD(G)
	22.1 관계 유형	유튜브 동영상과 다른 동영상의 관계, 사본, 참조관계 등 다양한 관계의 유형							NAK 8, PREMIS, PBCore
	22.2 관계대상 식별자	'관계유형'에서 선택된 값에 따 른 관계 대상의 식별자							NAK 8, PREMIS, PBCore
	22.3 관계 설명	해당 관계에 대한 상세한 설명							NAK 8, PREMIS, PBCore
인증 영역									
	23. 진본성 확인	유튜브 동영상의 진본성 확인에 대한 정보							NAK 8, PREMIS
	23.1 서명	본성 확인을 인증한 관리행위자 의 서명							PREMIS
	24. 무결성체크	유튜브 동영상의 무결성을 유지하기 위해 유튜브 동영상 이 관리 및 보존되는 과정 속에서 인가받지 않은 접근이 나 행위에 의해 위변조 혹은 훼손되지 않았음을 점검하 는 방법							NAK 8
	24.1 무결성 체크법	유튜브 동영상의 무결성을 확인 하기 위한 무결성 체크기법 (예 시 : SHA-256)							NAK 8
	24.2 무결성 체크값	무결성 체크법에 의해 생성된 값							NAK 8

2.5. 이용자 참여 메타데이터

이용자 참여 영역								
요소	요소 설명	평가 척도(5점)					출처	
		1	2	3	4	5		
25. 이용자 평가 정보	유튜브 동영상을 시청한 이용자들이 의해 나타나는 수치정 보						유튜브	
25.1 조회수	이용자들이 해당 유튜브 동 영상을 시청한 횟수						유튜브	
25.2 좋아요 수	이용자들이 해당 유튜브 동 영상을 시청한 뒤 '좋아요' 를 표시한 수치정보						유튜브	

25.3 싫어요 수	이용자들이 해당 유튜브 동영상을 시청한 뒤 '싫어요'를 표시한 수치정보							유튜브
25.4 댓글 수	이용자들이 해당 유튜브 동영상을 시청한 뒤 작성한 '댓글'에 대한 수치정보							유튜브
26. 댓글	이용자가 유튜브 동영상을 시청한 후 이용자의 의견이 들어 있는 정보							유튜브
26.1 댓글ID	유튜브 동영상에 댓글을 작성한 이용자 ID							유튜브
26.2 댓글 내용	유튜브 동영상에 작성된 댓글의 내용							유튜브
26.3 대댓글 ID	특정 댓글에 대한 대댓글을 작성한 이용자 ID							유튜브
26.4 대댓글 내용	특정 댓글에 대한 대댓글의 내용							유튜브

3. 본 연구에서 제안한 메타데이터 요소 이외에 필요하다고 생각하신 메타데이터 요소(들) 또는 기타 의견이 있으면 기재해 주시기 바랍니다.

* 질문에 응해주셔서 대단히 감사합니다.

부 록 2

공공기관 유튜브 동영상 아카이빙을 위한 메타데이터 요소

: 핵심 요소

설명 메타데이터 - 설명 영역	
상위 요소	하위 요소
1. 제목(Title)	
2. 주제(Subject)	2.1 태그(tag)
3. 설명(Description)	
4. 언어(Language)	
5. 범주(Coverage)	
구조 메타데이터 - 구조 영역	
6. 비디오(Video)	6.1 재생시간(duration)
	6.2 프레임비율(frame rate)
	6.3 데이터전송속도(data rate)
	6.4 압축(codec)
	6.5 인코딩정보(encoding)
7. 장르(Genre)	
8. 식별자(Identifier)	8.1 비디오ID(Video ID)
	8.2 URL
관리 메타데이터 - 권리 영역	
9. 생산자 (Creator)	9.1 생산자유형(creator type)
	9.2 생산자이름(creator name)
10. 권한(Rights)	10.1 권한식별자(rights identifier)
	10.2 권한근거(rights basis)
	10.3 공공저작물여부(Public works or not)
	10.4 공공저작물자유이용허락표시유형 (type of Korea open government license)
	10.5 공개구분(external access control)
	10.6 비공개사유(external access reason)

	10.7 시청등급(Audience level)
관리 메타데이터 - 관리 영역	
11. 일시(Date)	11.1 생산일시(time created)
	11.2 등록 일시(time registered)
12. 관리이력 (Management History)	12.1 관리유형(event type)
	12.2 관리설명(event description)
	12.3 관리일시(event date time)
	12.4 관리행위자(event agent)
	12.5 변경요소(changed element)
관리 메타데이터 - 기술 영역	
13. 포맷(Format)	13.1 포맷명(format name)
	13.2 포맷버전(format version)
14. 크기(Size)	14.1 파일크기(file size)
15. 운영시스템 (Operating system)	
16. 어플리케이션 (Application)	16.1 생성 어플리케이션 명 (creating application name)
	16.2 생성 어플리케이션 버전 (creating application version)
보존 메타데이터 - 출처 영역	
17. 수집,이관의직접적출처 (Immediate source of acquisition of information)	
보존 메타데이터 - 참조 영역	
18. 생산이력 (Creation history)	18.1 기관명(corporate name)
	18.2 의견(opinion)
	18.3 처리상태(status of creation action)
	18.4 처리일시(time of creation action)
19. 보존이력 (Preservation history)	19.1 보존처리 유형(preservation action type)
	19.2 보존처리설명 (preservation action description)

	19.3 보존처리일시(preservation action date)
	19.4 보존처리행위자(preservation agent)
20. 보존장소 (Preservation place)	
21. 위치(Location)	21.1 소장처(location)
	21.2 소장 위치(storage details)
	21.3 원본의존재와위치 (Existence and location of originals)
	21.4 사본의존재와위치 (Existence and location of copies)
보존 메타데이터 - 문맥 영역	
22. 관계(Relation)	22.1 관계유형(relation type)
	22.2 관계대상식별자(relation item ID)
	22.3 관계설명 (relation description)
보존 메타데이터 - 인증 정보	
23. 진본성 확인 (Object authentication)	23.1 서명(signature)
24. 무결성체크 (Integrity Check)	24.1 무결성체크법(integrity check name)
	24.2 무결성체크값(integrity check value)
이용자 참여 메타데이터 - 이용자 참여 영역	
25. 이용자피드백정보 (User feedback information)	25.1 조회수(views)
	25.2 좋아요수(number of like)
	25.3 싫어요수(number of dislike)
	25.4 댓글 수(number of reply)
26. 댓글(Reply)	26.1 댓글ID(reply ID)
	26.2 댓글내용(reply content)
	26.3 대댓글ID(re-reply ID)
	26.4 대댓글내용(re-reply content)

ABSTRACT

A Study on the Core Elements of Metadata for YouTube Video Archiving in Public Institutions

Kim, Rack-Keun

Major in Archives & Records Management

Dept. of Library and Information Science

The Graduate School

Hansung University

YouTube is used not only as a content distribution platform for sharing video data, but also as a path to obtain various information. In Korea, YouTube usage time is much higher than other social media types (Twitter, Facebook, etc.). In addition, YouTube is at the top of the list in information retrieval. Most of public institutions use YouTube as a tool for promoting themselves. Therefore, managing a YouTube channel in public institutions can be applied to as a public service, and videos uploaded on YouTube can be treated as the result of performing public tasks.

YouTube videos enable the public institutions to communicate with the public. Public institutions can interactively communicate with the

public in two directions by uploading YouTube videos. Moreover, social media is highly valuable as a record in that the results of communication remain as a record. In addition to uploading videos to YouTube by public institutions, the trend of society at that time also could be figured through subtitles or buzzwords in the uploaded videos. Furthermore, social and cultural characteristics could be also found in comments written by users who watched the video. As a video record, YouTube videos can increase users comprehension more than the text recordings of the same content. Considering this, it is necessary to manage and preserve YouTube videos of public institutions as a digital public record. In other words, YouTube videos should be archived to preserve and manage valuable videos. In this case, the video itself and metadata describing the video must be archived together. However, it is currently difficult to find metadata for YouTube video for archiving, and the related researches are also insufficient. Therefore, the purpose of this study is to design metadata for archiving YouTube videos of public institutions and select the core elements.

For this purpose above mentioned, the NAK 8:2021(v2.2), a domestic record management metadata standard, was investigated. In addition, the elements of ISAD(G), an international record metadata standard, and the semantic units and semantic components of PREMIS were investigated. Also, the elements of PBCore, the standard for broadcasting and video metadata, were investigated. Lastly, data generated after the videos were uploaded on YouTube were also monitored. And those were named YouTube metadata. After that, the investigated metadata elements were mapped based on the archive metadata structure synthesized through previous studies. As a result, metadata for YouTube video archiving was designed, consisting of 5 metadata categories and 10 sub-categories. The five metadata categories are divided into descriptive metadata, structural

metadata, management metadata, preservation metadata, and user participation metadata. The sub-region was subdivided into a description category for descriptive metadata and a structural category for structural metadata. In the case of management metadata, it was subdivided into rights, management, and technology categories, and in the case of preservation metadata, it was subdivided into source, reference, context, and authentication areas. Finally, in the case of user participation metadata, it was subdivided into user participation category.

In order to be evaluated for the designed metadata, a survey was conducted through 15 experts with experience related to digital archiving or metadata work or researches. The questionnaire consisted of a 5-point Likert evaluation method, additional elements, and other opinions for the general status and 63 metadata elements. Fifteen copies of the questionnaire were distributed and 14 copies (93%) were returned. In addition, the data purification work was conducted to refine the questionnaire item "Type of institution belonging to 1.1" by public institution, other institution, and university.

This study was performed by descriptive statistical analysis. During the descriptive statistical analysis, the type of institution, final education, and career of the respondent were analyzed through frequency analysis. In addition, through average analysis, the minimum, medium, maximum, average, and standard deviation of the data for the 5-point Likert evaluation results were analyzed. The results for the descriptive statistical analysis are as follows. First, the results of frequency analysis are as follows. The institutions belonging to the respondents were 9 public institutions (64%), 4 other institutions (29%), and 1 university (7%). The responses to the final academic background were six PhDs (42.86%), four masters (28.57%), and four bachelor's (28.57%). As for the response to career details, one person did not respond, and as a result of analyzing

13 people, 3 people (23%) in 0–5 years, 4 people (31%) in 6–10 years, 2 people (15%) in 11–15 years, and 4 people (31%) or more.

Secondly, the results of descriptive statistical analysis centered on the average value are as follows. Among the descriptive category, the 'title' was the only element in the entire metadata element that received 5 points (very appropriate) from all respondents. Accordingly, the 'title' element showed significant results as a core element without further analysis. In the response results of the structural category of the structural metadata, the average value of the 'duration' element was the highest, followed by 'URL', 'frame rate', and 'video ID'. The element with the lowest average value was found to be a 'genre'. The results of combining responses from the rights, management, and technology categories of the management metadata showed that the average value of the 'external access control' element was the highest among the three categories, followed by 'time created', 'time registered', 'file size', 'format name', 'Public works or not', and 'type of Korea open government license'. The 'event agent' element was found to have the lowest average value in the entire metadata element. The results of combining responses from the source, reference, context, and authentication category of the preservation metadata showed that the average value of the 'location' element was the highest among the four elements, followed by 'Existence and location of originals', 'corporate name', and 'storage details'. In the case of user participation metadata, the factors with the high average value were in the order of 'views' and 'number of like', and the average value of the 'reply ID' element was the lowest.

Third, 7 people responded to 'additional elements and other opinions'. In additional element responses, the "channel name", "number of subscribers", "age to use content" and "delete" elements were answered, and in other opinion responses, "need additional elements of retention

metadata”, “unclearance of top and sub-elements and whether to investigate required/select” and “provide archiving criteria”.

Next, validity verification and reliability verification were conducted. Validity verification was performed by a factor analysis method, and reliability verification was performed using the Cronbach's α value. The core elements were selected as a factor that satisfied both the validity test and the reliability test. The results of validity verification and reliability verification are as follows.

First, validity verification and reliability verification were conducted for each of the 10 sub-categories of the five metadata for the questionnaire response. As a result of conducting the sphericity test of KMO and Bartlett to find out whether factor analysis is suitable for each sub-category, the Right category of management metadata was found in categories that did not satisfy $KMO > 0.5$ and Bartlett's $p < 0.05$. Factor analysis was not possible on the preservation metadata because the source category of the preservation metadata had only one element ('Immediate source of acquisition of information') and the reference category had no value extracted as a result of factor analysis.

Second, the results of factor analysis are as follows. In the explanation category of the explanation metadata, only the remaining elements except for the 'title' element were analyzed, and as a result, 'language', 'description', and 'coverage' met the standard values of extraction commonality and component values. In the structural category of structural metadata, only 'frame rate', 'data rate', 'duration', 'codec', and 'encoding' met the reference values of commonality and component values. Elements that met the extraction reference value in the management category of the management metadata were found to be 'event date time', 'event agent', 'changed element', 'time created', 'event description', and 'event type'. In the technical category, it appeared as

'creating application name', 'creating application version', 'operating system', and 'format version'. The areas where all elements met the extraction criteria were the context category of the preserved metadata ('relation item ID', 'relation type', 'relation description'), the authentication category ('integrity check name', 'signature'), and the user participation category ('views', 'number of like', 'number of dislike', 'number of reply', 'reply ID', 'reply content', 're-reply ID', 're-reply content').

Third, reliability verification was conducted for each metadata and area only for factors that met the validity verification reference value. As a result, all of them were confirmed to be 0.7 or more, indicating that it was very desirable, and it can be said that the factors were significant as core elements.

The metadata designed in this study consisted of 63 measurement elements included in 10 areas of 5 metadata. Among them, 33 elements included in 7 categories of 5 metadata were found to be significant factors through validity verification and reliability analysis. Accordingly, the adoption rates by metadata, area, and measurement element were 100%, 70%, and 53%, respectively.

It may be difficult to say that the YouTube videos selected as the subject of this study derive metadata elements represents the entire social media videos. In addition, there is a limitation in that it is difficult to determine whether or not the core element is based on the results of factor analysis conducted to derive the core element. However, this study is meaningful as an initial study of metadata for archiving social media videos among various metadata studies, and it is distinguishable compared to the existing metadata studies in that the study was conducted on the YouTube videos. If the elements of social media metadata are diversified through various follow-up studies in the future, methods such as

confirmatory factor analysis for large groups will be possible.

It can be said that the ultimate purpose of record management is in the use of records. YouTube records can be also increase their value as the records by provision of services to users through archiving and identifying meaningful phenomena through collected data. This study focused on metadata in the archiving of YouTube video records, and it is hoped that various studies will be conducted through follow-up studies.

【Keywords】 : Public institutions, social media records, YouTube records, archiving, metadata, core elements