

# 소셜미디어 아카이빙을 위한 수집 및 보존방안\*

## Acquisition and Preservation Methods for Social Media Archiving

김태영 (Tae-Young Kim)\*\* , 양동민 (Dongmin Yang)\*\*\*  
최상기 (Sang-Ki Choi)\*\*\*\* , 오효정 (Hyo-Jung Oh)\*\*\*\*\*

### 초 록

최근 공공기관에서 국민과 소통하는 방법의 하나로 소셜미디어를 활발하게 활용함에 따라 기록관리 영역에서도 이를 수집하고 보존하기 위한 다양한 방안들이 논의되고 있다. 미국에서는 정부 차원에서 소셜미디어를 수집·보존하고자 하는 움직임이 일고 있고, 영국 국립기록관에서는 이미 공공기관의 트위터와 유튜브를 대상으로 아카이브를 구축하여 이용자에게 서비스를 제공하고 있다. 이에 본 연구에서는 소셜미디어 아카이빙의 수집 및 보존 측면에서의 사례조사 결과를 토대로 소셜미디어의 유형별 특징과 그에 따른 수집 방안과 보존 모형을 제안하고자 한다. 또한 제안방안의 실효성을 검증하기 위해 경상남도 도청의 소셜미디어를 대상으로 시범 적용하였다. 본 연구는 실제 수집한 결과를 통해 수집 및 보존방안을 제시하였다는 점에서 그 의의가 있으며, 차후 소셜미디어 아카이빙을 위한 모델 수립에 도움이 될 수 있을 것으로 기대한다.

### ABSTRACT

Recently, various methods of collecting and preserving social media have been discussed in the field of archives and records management as social media is actively used as a way of communicating with the public in the government. In the United States, there is a move to acquire and preserve social media at the government level, and the National Archives (UK) already provides social media archives to users through Twitter and YouTube. In this study, we proposed the features, acquisition methods of social media by type and the preservation model based on case study results in terms of acquisition and preservation of social media archiving. In order to verify the effectiveness of the proposed methods, this study was applied to the social media of Gyeongsangnamdo provincial government. The results of this study is meaningful in that it suggested acquisition and preservation methods through actual collected results and it is expected that it will be useful for establishing the models for future social media archiving.

키워드: 소셜미디어, 소셜미디어 아카이빙, 자동 수집, 보존 포맷  
social media, social media archiving, automatic acquisition, preservation format

\* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임  
(NRF-2016S1A5B8913575).

이 논문은 2018년 『경상남도기록원 영구기록관리시스템 구축』 연구 결과의 일부를 확장·보완한 것임.

\*\* 전북대학교 일반대학원 기록관리학과 박사수료(fnty127@hanmail.net) (제1저자)

\*\*\* 전북대학교 기록관리학과 조교수, 문화융복합아카이빙연구소 연구원(dmyang@jbnu.ac.kr) (공동저자)

\*\*\*\* 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 연구원(choisk@jbnu.ac.kr) (공동저자)

\*\*\*\*\* 전북대학교 문헌정보학과 부교수, 문화융복합아카이빙연구소 연구원(ohj@jbnu.ac.kr) (교신저자)

■ 논문접수일자: 2019년 5월 17일 ■ 최초심사일자: 2019년 6월 13일 ■ 게재확정일자: 2019년 6월 18일

■ 정보관리학회지, 36(2), 79-104, 2019. [http://dx.doi.org/10.3743/KOSIM.2019.36.2.079]

## 1. 서론

### 1.1 연구필요성 및 목적

미국 트럼프 대통령은 2017년 1월 취임 이후 약 1년 동안 대략 3천개의 트윗을 게시하면서 (황준범, 2018), 트위터를 대중과의 소통, 그리고 정책 제안의 도구로 활용하고 있다. 트럼프 대통령 외에도 여러 국가원수 및 정치인들이 여러 소셜미디어를 활용해 외교, 복지 등 다양한 분야의 정책에 대한 결정사안이나 견해를 알리고 있다. 특히, 소셜미디어는 빠른 확산력과 넓은 노출 범위와 같은 특징으로 인해 공적인 커뮤니케이션 장으로서의 역할이 증대되고 있으며(이귀영, 김지현, 2015), 이러한 환경 변화에 맞춰 최근 국내 대부분의 공공기관에서도 국민과 소통하는 방법의 주요한 매체로 소셜미디어를 활용하고 있다. 더불어 이러한 기조에 발맞춰 기록관리 영역에서는 공공영역의 소셜미디어에 대한 수집과 보존에 대한 논의가 중요 쟁점으로 각광받게 되었다.

해외에서는 이 같은 흐름에 발맞춰 많은 움직임이 활발히 일고 있는데, 이미 미국에서는 정부 차원에서 소셜미디어를 수집·보존하고자 하는 시도가 있었고, 영국 국립기록관(TNA)에서는 공공기관의 트위터와 유튜브를 대상으로 아카이브를 구축하여 이용자에게 서비스를 제공하고 있다. 이와 달리 국내는 아직 국가 차원에서 소셜미디어를 수집하고 보존하기 위한 논의를 시작하는 단계에 머물러 있어, 해외 소셜미디어 아카이빙 사례를 기반으로 도출된 시사점들을 종합적으로 검토하여 국내 소셜미디어 수집 및 보존을 위한 실행방안을 마련할 필요성

이 대두된다.

따라서 본 연구에서는 수집 및 보존 측면에서 해외 소셜미디어 아카이빙 프로젝트 사례를 조사·분석하였고, 이를 토대로 소셜미디어의 유형별 특징과 그에 따른 수집 프로세스와 전략, 그리고 보존 모형을 제안하고자 한다. 이후 제안방안의 실효성을 검증하기 위해 본 연구에서는 경상남도 도청(이하 경남도청)에서 운영 중인 소셜미디어를 대상으로 시범 적용해봤다. 경남도청을 연구 대상으로 선정한 이유는 최근 경상남도기록원에서 경상남도 내 공공기관에서 운영 중인 소셜미디어 기록을 정부차원에서 적극적으로 수집 및 보존할 의지를 표명하였고(경상남도기록원, 2018), 해당 공공기관 중 가장 대표적인 기관이 경남도청이기 때문에 본 연구의 시범 적용 대상으로 선정하였다. 본 연구 결과로 제시된 소셜미디어 수집 및 보존 방안은 실제 수집한 결과를 통해 제시되었다는 점에서 그 의의가 있다.

### 1.2 연구방법 및 범위

본 연구의 목적은 소셜미디어 아카이빙을 위한 소셜미디어 기록의 수집 및 보존 방안을 제안하는 것으로, 이를 위해 다음과 같이 연구가 진행되었다.

첫째, 문헌연구를 통해 소셜미디어에 대한 개념과 특징, 유형 등을 정의하고, 소셜미디어 아카이빙 도구 및 전략을 조사하여 국내 소셜미디어 아카이빙에의 적용 가능성을 살펴보았다. 또한 문서보존포맷 및 장기보존포맷에 대한 해외 현황에 대해 조사함으로써 소셜미디어 보존에 적합한 포맷 형태를 제안하였다.

둘째, 소셜미디어 아카이빙 프로젝트 사례를 선행연구를 통해 조사·분석하여 국내 소셜미디어 수집 및 관리 시스템 도입에 있어서의 시사점을 도출하였다. 본 연구에서는 소셜미디어 아카이빙 프로젝트 사례로 미국 의회도서관의 트위터 아카이브, 호주 국립도서관의 PANDORA 아카이브, 영국 국립기록관의 트위터 및 동영상 아카이브를 선정하여 조사·분석하였다. 또한 경상남도기록원 관리 대상 기관의 소셜미디어 현황을 분석하기도 하였다.

셋째, 문헌연구 및 사례조사한 결과를 바탕으로 소셜미디어의 유형에 따라 구체적인 수집 방안을 제시하였다. 본 연구에서는 페이스북(Facebook), 트위터(Twitter), 인스타그램(Instagram), 유튜브(Youtube) 등 총 4가지 유형별로 각각의 페이지 구조와 수집 전략에 대해 제안하였으며, 소셜미디어 특성에 맞는 보존을 위한 보존 포맷도 함께 제안하였다.

넷째, 제안방안의 실효성 검증을 위해 경남도청에서 운영 중인 트위터 계정을 시범 적용 대상으로 선정하여 수집도구를 개발, 관련 트윗들을 자동 수집하였으며, 수집 결과에 대한 활용 방안으로 시계열 분석을 수행하였다. 이러한 분석 결과를 기반으로 수집 전략 및 주기에 대한 몇 가지 시사점을 결론으로 도출하였다.

## 2. 이론적 배경 및 선행연구

### 2.1 소셜미디어 개념 및 유형

#### 2.1.1 소셜미디어 개념

소셜미디어는 사람들의 의견, 생각, 경험, 관

점 등을 서로 공유하기 위해 사용하는 온라인 톨과 플랫폼으로, 가입자 기반 커뮤니케이션 서비스로 시작된 소셜네트워킹 서비스가 가입자 기반이 확대되면서 단순한 커뮤니케이션에 머무르지 않고, 가입한 이용자들이 서로 정보와 의견을 양방향으로 공유하고 확산시킬 수 있는 미디어로 진화된 현상을 대변하는 신조어(FKII 조사연구팀, 2006; 김윤화, 2013)이다.

『공공기록물 관리에 관한 법률』에서 정의하는 ‘기록물’ 및 ‘전자기록물’의 범주에는 소셜미디어가 직접적으로 포함되어 있지 않지만(〈표 1〉 참조), 현재 국가적 차원에서는 소셜미디어를 ‘웹기록물’의 범주로 포함하여 다루고자 하고 있다(네이버시스템, 2017). 특히, 우리나라 대부분의 공공기관에서는 국민과의 소통 수단으로서 페이스북, 트위터 등의 소셜미디어를 활발히 사용하고 있기 때문에 소셜미디어를 통해 제공되는 전문적인 정책내용과 이미지, 동영상 등의 다양한 콘텐츠, 그리고 시민들의 제안이 담긴 댓글들은 중요기록물로서 보존되어야 할 높은 가치를 지니고 있다(임진희, 2013; 장인호, 황윤영, 이규철, 2014).

현재까지 계속적으로 소셜미디어 수집 및 아카이브에 대한 사례들이 발표되고 있는 상황을 고려하면 소셜미디어를 통해 생산·유통되는 정보가 기록 관리의 적극적인 대상으로 간주되고 있음을 알 수 있으며(이귀영, 김지현, 2015), 일부 공공기관의 웹사이트에서는 기관의 소셜미디어 게시글을 웹사이트와 연결하여 웹기록물 형태로 이용자에게 제공하는 경우도 있다. 따라서 앞으로는 이메일, 소셜미디어 등 새롭게 등장할 기록들을 포괄할 수 있도록 기록물 개념에 대한 재정립이 필요한 상황이다. 따라서 본

〈표 1〉 법률상에서의 기록물 정의

용어	정의
기록물	• 공공기관이 업무와 관련하여 생산하거나 접수한 문서·도서·대장·카드·도면·시청각물·전자문서 등 모든 형태의 기록정보 자료와 행정박물(行政博物) (공공기록물 관리에 관한 법률 제3조)
전자기록물	• 정보처리능력을 가진 장치에 의하여 전자적인 형태로 작성하여 송신·수신 또는 저장되는 전자문서, 웹기록물 및 행정정보 데이터세트 등의 기록정보자료 (공공기록물 관리에 관한 법률 시행령 제2조)
웹기록물	• 공공기관의 웹사이트에 포함된 모든 형태의 기록정보자료와 웹사이트 운영 및 구축과 관련된 관리정보 (공공기록물 관리에 관한 법률 시행령 제2조)

연구에서는 소셜미디어 기록을 웹기록물 범주 중 하나의 유형으로 보고 웹기록물이 갖는 특성을 상속한다고 간주하였다.

2.1.2 소셜미디어 유형

소셜미디어는 어떤 기준으로 구분하는가에

따라 다양한 유형으로 분류될 수 있으며, 현재는 기본적으로 블로그, 소셜네트워크 서비스, 콘텐츠 커뮤니티, 위키, 마이크로 블로그 등으로 소셜미디어를 분류할 수 있다(〈표 2〉 참조). 이 외에도 호주의 Public Record Office Victoria에서는 웹 기반의 외부 호스팅(web-based externally

〈표 2〉 소셜미디어의 유형

구분	내용
블로그 (Blog)	• Web(웹)과 Log(기록)의 합성어로 특정 주제에 대해 시간 순으로 자유롭게 등록 가능한 사이트 • 개인 미디어인 블로그를 중심으로 소셜 네트워크 기능이 결합된 서비스 • 운영자의 취향에 따라 글의 주제, 업데이트 주기, 관리 운영에 대한 제약이 없는 1인 미디어라고 할 수 있음 • 블로그는 다른 웹 사이트로 링크가 가능한 이미지와 비디오, 음악 등이 많이 포함된 콘텐츠들을 통합하는 온라인 공간임
소셜 네트워크 서비스 (SNS)	• 주로 인적 네트워크 형성 목적의 서비스로 페이스북, 마이스페이스, 싸이월드, 카카오톡 등이 대표적 • 개인의 관심주제와 같은 관심사항 및 활동을 하고 있는 다른 사람들을 온라인 공간에서 검색하여 온라인 인맥을 구축하고 유지할 수 있는 온라인 공간을 의미함 • 사용자들은 기존 커뮤니티에 가입하거나 사용자 스스로 대화 커뮤니티를 형성하여 운영할 수 있음 • 특정 사용자나 분야의 제한 없이 누구나 참여 가능한 서비스
콘텐츠 커뮤니티 (Contents Communities)	• 사진, 동영상 등 특정한 종류의 콘텐츠를 제작하고 공유하는 사이트로 유튜브, 인스타그램, 핀터레스트, 플리커 등이 대표적 • 사진, 비즈니스, 게임, 음악, 레스토랑 등 특정 관심 분야를 공유하는 서비스
위키 (Wiki)	• 사용자들이 인터넷상 공동 문서를 작성하고 내용을 수정·편집하여 서로 의견을 공유하는 사이트 • 편집 가능한 웹 페이지로 웹사이트 상에서 콘텐츠를 추가하고 정보를 편집하여 마치 communal 문서나 데이터베이스처럼 운영됨 • 공동 창작, 협업 기반의 서비스로 위키피디아 서비스가 대표적
마이크로 블로그 (MicroBlog)	• 140~250자 이내의 짧은 텍스트로 관계형성 및 정보공유를 위한 서비스로 트위터, 미투데이 등이 대표적 • 짧은 단문형 서비스로 대형 소셜네트워크 서비스 시장의 틈새를 공략하는 서비스 • 간단한 인스턴트 메신저나 핸드폰의 문자 메시지와 블로그의 중간 형태로 사용자들이 주기적으로 자신이 관심 있는 전문 주제에 생각과 정보의 수집 공유가 가능한 공간임

출처: FKII조사연구팀, 2006; 도서관연구소, 2012; 정영철 외, 2012; 이윤희, 2014 재구성.

hosted), 내부 호스팅(internally hosted), 비즈니스 시스템 및 기기에 내장된 어플리케이션 (applications built into business systems and devices) 등 소셜미디어가 활용되는 컴퓨팅 환경을 기준으로 소셜미디어 유형을 구분(Public Record Office Victoria, 2017; 정보인권, 2018) 하기도 하였다.

## 2.2 소셜미디어 아카이빙 전략 유형

소셜미디어 아카이빙 전략은 크게 API를 활용한 데이터 아카이빙, 써드파티(third-party) 서비스 활용, 플랫폼 셀프 아카이빙 서비스 활

용, 오픈 소스 코드 활용, 데이터 리셀러(data reseller) 활용, 연구 데이터세트 활용, 연구 도큐멘테이션 및 메타데이터 활용 등 7가지로 구분할 수 있으며, 각 전략에 대한 세부 내용은 <표 3>과 같다.

## 2.3 선행연구 분석

### 2.3.1 소셜미디어 아카이빙 프로젝트

소셜미디어에 대한 아카이빙 프로젝트를 진행한 대표적인 사례에 대해 살펴보고, 실제로 어떤 아카이빙 도구 및 전략을 사용하였으며, 주요 시사점은 무엇인지에 대해 분석하였다.

<표 3> 소셜미디어 아카이빙 전략 유형

소셜미디어 아카이빙 전략		세부 내용
①	API를 활용한 데이터 아카이빙	<ul style="list-style-type: none"> <li>• 각 소셜미디어 플랫폼들은 API를 통해 소셜미디어 데이터를 주고받음</li> <li>• Twitter API, Graph API 등 서비스 제공자별 API를 이용하여 소셜미디어의 데이터를 아카이빙할 수 있음</li> <li>• Twitter의 경우 연구 목적에 한해 데이터의 1%를 Streaming API로 제공</li> </ul>
②	써드파티 서비스 활용	<ul style="list-style-type: none"> <li>• 일부 정부 및 문화유산 기관들은 ArchiveSocial, MirrorWeb, Erado 등 상용 소셜미디어 콘텐츠 서비스 및 하베스팅 서비스를 활용하기도 함</li> </ul>
③	플랫폼 셀프 아카이빙 서비스 활용	<ul style="list-style-type: none"> <li>• 각 플랫폼에서 제공하는 셀프 아카이빙 기능을 이용하여 전체 소셜미디어 데이터의 아카이빙 가능</li> <li>• Facebook은 관리자가 업로드한 콘텐츠에 한해 이메일 요청 후 수일 내에 Zip파일로 다운로드를 제공하고, Twitter는 개인 계정에 한해 json, csv 호환 가능한 형태로 데이터 다운로드를 제공함</li> </ul>
④	오픈 소스 코드 활용	<ul style="list-style-type: none"> <li>• 인터넷에 공개된 오픈 소스 코드를 활용하여 소셜미디어 데이터의 아카이빙이 가능함</li> <li>• 오픈 소스 코드를 활용하므로 재정적으로 자유로울 수 있으나 문제 발생 시에 대처할 수 있는 책임자가 없음</li> </ul>
⑤	데이터 리셀러 활용	<ul style="list-style-type: none"> <li>• GNIP는 독자적인 데이터 포맷(Gnip Data Format)을 이용하여 소셜미디어 데이터를 실시간으로 렌더링하는 'Power Track' 서비스(주로 학술연구 목적으로 활용)를 2011년부터 유료로 제공</li> <li>• DataSift는 페이스북 토픽 데이터(Facebook Topic Data) 등 소셜미디어 데이터 수집하여 재판매</li> <li>• 데이터 리셀러가 제공하는 소셜미디어 데이터는 역사적 데이터의 경우 100% 스트리밍 데이터를 제공하지만 리셀러의 데이터는 API를 이용하므로 플랫폼 제공자의 정책에 따라야 하며 역사적 데이터 제공보다 업체의 이익을 우선함</li> </ul>
⑥	연구 데이터세트 활용	<ul style="list-style-type: none"> <li>• 소셜미디어 연구를 위해 수집 및 구축된 데이터세트를 연구팀에게 입수하여 재활용</li> <li>• Twitter의 경우 5만 건의 트윗에 한해 기계비가독형(PDF 등)으로 제공</li> </ul>
⑦	연구 도큐멘테이션 및 메타데이터 활용	<ul style="list-style-type: none"> <li>• 소셜미디어 관련 연구의 결과물인 도큐멘테이션이나 메타데이터를 입수하여 아카이빙 전략 및 정책수립의 자료로 활용</li> </ul>

출처: 네이버시스템, 2017 수정·보완.

첫 번째 소셜미디어 아카이빙 프로젝트는 미국 의회도서관의 트위터 아카이브이다. 미국 의회도서관에서는 2010년에 트위터(Twitter)와 협정을 맺어 2006년 트위터 출범 당시부터 2010년도까지 작성된 모든 트윗 데이터를 기증받았다. 그 이후로도 계속적으로 공개된 트윗을 아카이브하여 첫 문서화된 트윗 컬렉션(2006-2017)을 소장하게 되었으며(도서관연구소, 2013; 국립중앙도서관, 2018), 트윗 데이터에는 트윗 위치, 설명 등 50여개의 메타데이터 필드가 포함되어 있다. 그러나 2018년도부터 미국 의회도서관은 트윗 아카이빙을 전체 수집에서 선택적 수집으로 정책을 변경(국립중앙도서관, 2018)하게 되었다. 수집 정책 변경에 대한 사유로는 생성되는 트윗 개수가 기하급수적으로 늘어났고, 기존에는 텍스트 형태로만 트윗을 수집하였기 때문에 이미지나 동영상, 링크 콘텐츠 등이 제거된 트윗은 그 수집 가치가 떨어지며, 트위터가 애초 설명한 것과는 달리 글자 수 제한을 확대하였다(허완, 2017)는 점을 들고 있다.

두 번째 소셜미디어 아카이빙 프로젝트는 호주 국립도서관(NLA)의 PANDORA 아카이브이다. PANDORA는 호주의 웹 아카이브 프로젝트로, 호주 온라인 출판물의 장기보존과 활용을 목적으로 시작되었다. PANDORA에는 현재 호주 총리의 웹기록물을 수집 및 관리하고 있으며, 관리되는 웹기록물에는 전·현직 총리들의 트위터 계정 또한 포함되었다. PANDORA에서는 전·현직 총리의 트위터 페이지를 14일~30일 주기로 수집하고 있으며, 수집 시점의 해당 트위터 페이지 전체를 일괄 수집하는 방식을 취하고 있다(최두원, 이수진, 윤은하, 오효정, 2018). 현재 PANDORA에서는 수집된 트위터 페이지 전체

를 단순히 보여주는 형태로 아카이빙 서비스를 제공하고 있으며, 검색 도구를 제공하지 않아 이용자 접근이 불편하다. 그리고 수집된 트위터 페이지 내의 게시물 원본 및 이미지를 식별하기 위해서는 PANDORA 페이지를 필수적으로 거쳐야 하기 때문에 이용 편의성이 떨어진다.

세 번째 소셜미디어 아카이빙 프로젝트는 영국 국립기록관(TNA)의 트위터와 동영상 아카이브이다. 영국 국립기록관에서는 2003년부터 영국 정부가 온라인으로 생산한 30억 건이 넘는 웹기록물을 수집하는 'UK Government Web Archive'를 운영하고 있으며(최두원 외, 2018), 그 일환으로 트위터, 유튜브와 같은 소셜미디어까지 수집 영역을 확대하여 아카이브 서비스를 제공하고 있다. TNA에서는 현재 Mirror Web 솔루션을 도입하여 트위터와 유튜브를 아카이빙하고 있으며, 해당 솔루션에는 기존 컬렉션의 마이그레이션 뿐만 아니라 콘텐츠 누락가능성을 줄이기 위한 일일캡처, 메타데이터 표시 방식의 향상, 날짜범위 통합 등도 포함되어 있다(한상호, 2018). 트위터와 유튜브 아카이브의 문제점으로는 키워드 검색 기능을 제공하지 않아 정보 접근을 제한함, 소셜미디어에 대한 메타데이터 등 세부 정보가 부족함, 2017년 이후로 업데이트되지 않음, 게시물 내용 자체만 수집하고 있어 댓글이나 '좋아요' 등 기록물에 대한 상호작용 정보를 알 수 없음 등이 제시되고 있다(최두원 외, 2018).

### 2.3.2 소셜미디어 아카이빙을 위한 시사점

미국 트위터 아카이브는 API(Ginp)을 활용하여 직접 아카이빙할 소셜미디어 데이터를 제공받는 전략을 취하고 있으며, 최근 소셜미디어

활용 동향에 맞춰 단순 텍스트뿐만 아니라 이미지, 동영상, 외부 링크 등을 포괄하여 데이터를 수집할 방안 모색이 필요한 상태이다. 호주 PANDORA 아카이브는 API를 활용하여 특정 시점의 트위터 페이지를 크롤링한 후, 재구성하여 아카이빙하는 전략을 취하고 있으며, 해당 아카이브는 소셜미디어 수집 자체에 의의를 두고 있기 때문에 이용자 정보 제공 측면은 미흡한 상태이므로 이에 대한 보완이 필요한 상태이다. 영국 트위터와 동영상 아카이브는 씨드파티 서비스를 통해 API를 활용하여 아카이빙하는 전략을 취하고 있으며, 메타데이터 정보 부족으로 인해 검색 및 정보 제공 기능이 다소 미흡한 상태이므로 이에 대한 보완이 필요한 상태이다.

이상의 소셜미디어 아카이빙 프로젝트 사례에 대한 내용을 바탕으로 도출된 시사점을 살펴보면, 먼저 소셜미디어 수집 시에는 공개 소프트웨어나 전용 API를 활용하는 것이 효율적이다. 특히, 공개 소프트웨어를 사용하면 소셜미디어 수집 시에 재정적인 부담이 경감되기 때문에 보다 유용한 방법이다. 둘째, 텍스트 이외에 이미지, 동영상 등 다양한 콘텐츠에 대한 멀티미디어 수집정책을 수립해야 한다. 소셜미디어 상에서 생산되거나 공유되는 콘텐츠 중 기관에서 자체 제작한 콘텐츠가 아닌 경우는 수집 및 보존 시, 저작권 문제를 해결해야만 한다. 셋째, 소셜미디어에 대한 이용자와의 상호작용 정보를 함께 수집할 필요가 있다. 기관에서 소셜미디어를 사용하는 목적 자체가 국민과의 소통을 위해서인 경우가 대부분이기 때문에 소셜미디어 수집 시에는 자체 내용뿐만 아니라 댓글이나 '좋아요' 등 이용자와의 상호작용

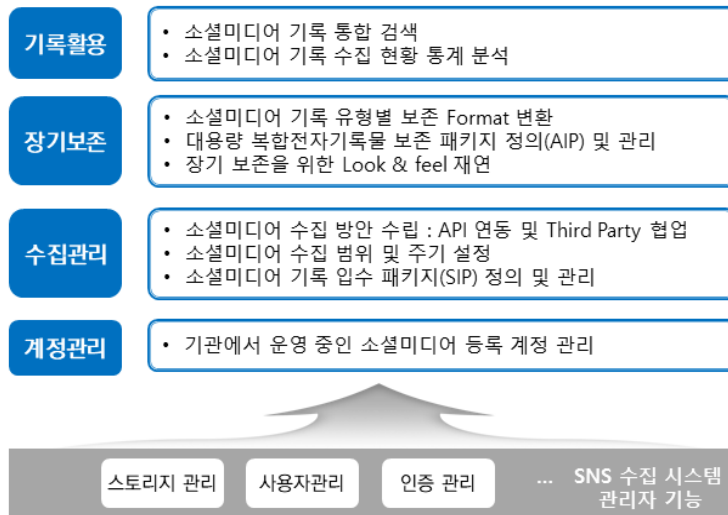
정보도 함께 수집할 필요가 있다. 넷째, 수집 이외에도 이용자 정보 제공 측면에 대한 고려가 필요하다. 해외 사례 대부분이 수집에만 초점을 맞추고 있어 수집한 결과를 이용자에게 제공하는 측면에 대한 고려는 다소 미흡한 편이며, 서비스 제공을 위한 기반 마련을 위해서는 상세 메타데이터를 수집해야 한다.

본 연구에서 분석한 해외의 소셜미디어 아카이빙 프로젝트들은 대부분 국가적인 차원에서 직접적이고 주도적으로 소셜미디어를 수집 및 보존하고 있다는 점에서 모범적인 실무 사례라고 할 수 있다. 이에 비해 우리나라는 공공영역의 소셜미디어를 수집하고 보존하기 위한 논의를 이제 시작하고 있는 단계에 있다. 따라서 해외 소셜미디어 아카이빙 프로젝트 사례로부터 도출된 시사점들을 종합적으로 검토하여 국가 차원에서 공공영역의 소셜미디어를 수집 및 보존하기 위한 실행방안을 마련할 필요가 있다. 다만, 해외 소셜미디어 아카이빙 프로젝트들 역시 아직까지는 대부분 검색 및 활용 측면에서 미흡한 점이 많고, 트위터나 유튜브 같은 일부 소셜미디어에 대해서만 프로젝트를 진행했다는 점, 그리고 이용자와의 상호작용 정보를 수집하지 않았다는 점 등을 개선하기 위한 방안 마련에 본 연구의 초점을 맞추고자 한다.

### 3. 소셜미디어 수집 방안

#### 3.1 소셜미디어 수집 및 관리 시스템

〈그림 1〉은 본 연구에서 제안하는 소셜미디어 수집 및 관리 시스템의 기본 구조를 시각화



〈그림 1〉 소셜미디어 수집 및 관리 시스템

한 것이다. 소셜미디어 수집 방법 및 전략은 소셜미디어의 유형별 특징에 따라 결정되며, 본 연구에서는 수집 및 관리시스템을 위한 기본 컴포넌트를 설계함으로써 구체적인 시스템 구축 방법을 제안하였다. 소셜미디어 수집 및 관리 시스템은 우선 스토리지 관리, 사용자 관리, 인증 관리 등 기본적인 시스템 관리를 위한 공통 기능을 바탕으로 소셜미디어에 특화된 계정관리, 수집관리, 장기보존, 기록활용 기능으로 구성된다. 계정관리를 통해 기관에서 운영 중인 소셜미디어 계정이 시스템에 등록되면 등록된 소셜미디어의 특징에 따라 수집 방안과 주기, 입수 패키지를 설정하여 주기적으로 소셜미디어를 수집하게 된다. 수집관리 단계에서 수집된 소셜미디어를 보존 포맷으로 변환하고 장기보존을 위한 재현 방안을 선택함으로써 장기보존 기능이 수행되고, 이후 기록활용 단계에서 보존된 소셜미디어에 대한 검색 및 수집 현황에 대한 통계 분석 등의 기능 등이 제공된다.

### 3.2 유형별 수집 구조

소셜미디어에는 텍스트를 비롯하여 이미지, 동영상 등 다양한 유형의 기록이 포함되어 있으며, 이들을 모두 포괄하여 수집할 수 있는 방안 마련이 최근 주요 연구 방향이다. 특히, 댓글이나 '리트윗', '좋아요' 등 기록에 대한 이용자의 상호작용 정보가 맥락 정보로서 중요해짐에 따라 각 소셜미디어에서 어떤 종류의 콘텐츠와 메타데이터를 기록으로 수집할 것인가에 대한 논의가 필요하다. 이에 본 연구에서는 소셜미디어 플랫폼별로 기록 구조가 서로 상이하기 때문에 우리나라에서 대표적으로 사용되는 소셜미디어들을 중심으로 유형별 수집 방안을 제시하고자 한다. 기본적인 소셜미디어 수집 및 관리 체계는 모든 소셜미디어에 동일하게 적용될 수 있으며, 세부적인 수집 방안 및 제출 정보 패키지(SIP: Submission Information Package) 구성은 소셜미디어 유형별로 별도 제시가 가능

하다. 수집된 소셜미디어는 차후 json이나 xml과 같은 '데이터세트형'으로 저장될 수 있다.

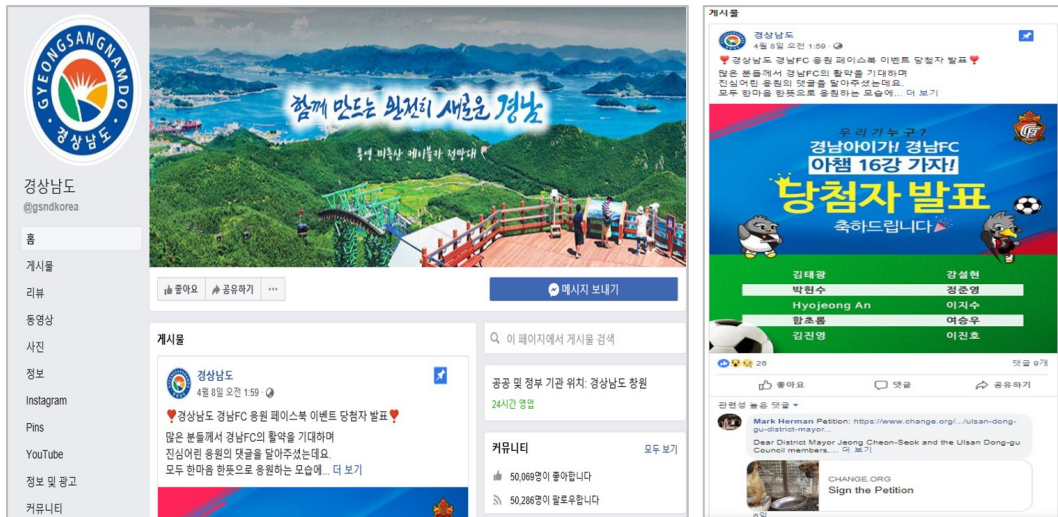
본 연구에서는 소셜미디어 유형에 따라 그 구조를 살펴보고 이를 수집하기 위한 방안을 제시하고자 한다. 이를 위해 실제 운영 중인 소셜미디어 계정을 시범 선정, 구체적인 예시를 통해 그 특성을 비교하고자 한다. 시범 선정 계정은 경남도청으로 선정하였는데, 그 이유는 앞서 언급했다시피 지방기록물관리기관 중 최초로 설립된 경상남도기록원에서 최근 경상남도 내 공공기관에서 운영 중인 소셜미디어 기록을 정부 차원에서 적극적으로 수집 및 보존할 의지를 표명하였고(경상남도기록원, 2018), 이 중 가장 대표적인 공공기관이 경남도청이기 때문이다. 사전 조사 결과, 현재 경남도청에서 운영하고 있는 소셜미디어 계정은 총 7개로, 본 연구에서는 가장 대표적인 소셜미디어 서비스인 페이스북(Facebook)과 트위터(Twitter), 인스타그램(Instagram), 유튜브(Youtube) 등 총 4가지

유형을 살펴보기로 한다.

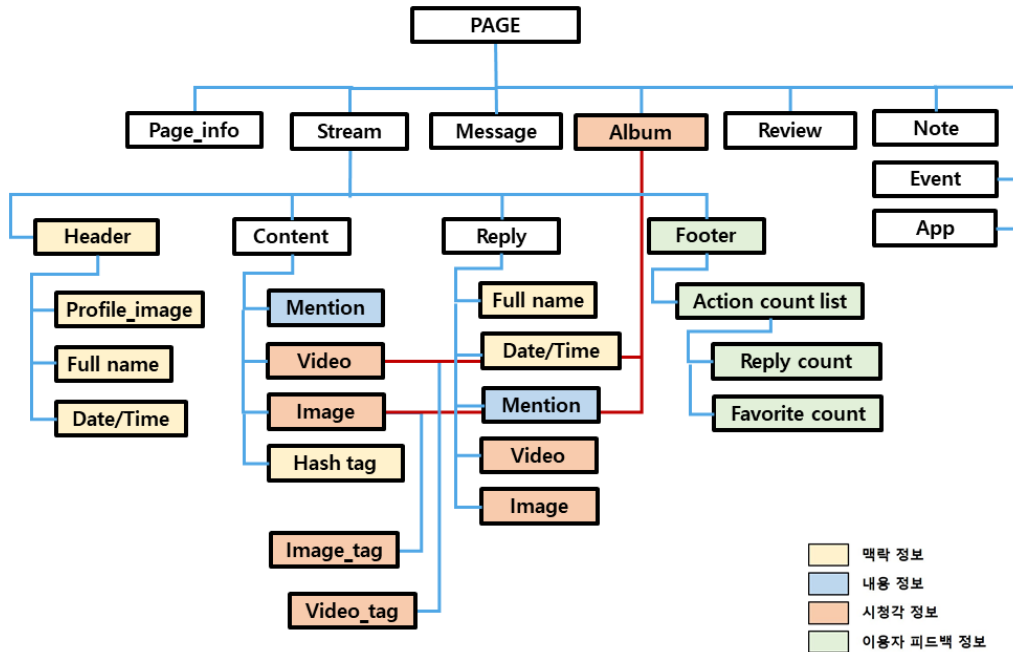
### 3.2.1 페이스북(Facebook)

페이스북은 개인 사용자를 위한 프로필, 기관이나 단체를 위한 페이지, 커뮤니티를 위한 그룹으로 구분(장인호, 황윤영, 이규철, 2014)되며, 본 논문에서는 기관이나 단체를 위한 페이지를 대상으로 데이터 구조를 분석하였다. <그림 2>를 통해 경남도청 페이스북 페이지를 대략적으로 살펴보면 왼쪽은 프로필과 메뉴로 구성되어 있고 오른쪽은 메뉴를 클릭할 시, 메뉴에 해당되는 정보를 보여주는 공간으로 주로 페이지 게시물과 프로필 정보로 구성되어 있다. 페이스북 페이지는 기본적으로 제공되는 기관 정보 및 게시물을 중심으로, 이를 재가공하여 다른 화면으로 보여주거나 애플리케이션 추가가 가능하여 다른 소셜미디어에 비해 복잡한 구조를 보인다.

페이스북 페이지를 구조화하면 <그림 3>과 같으며 페이지에 대한 상세 정보, 페이지 게시



<그림 2> 경남도청 페이스북 페이지 화면



〈그림 3〉 페이스북 페이지 구조

물인 스트림, 커뮤니티와 주고받는 메시지, 멀티미디어 게시물을 묶어서 보여주는 앨범, 페이지에 대한 리뷰, 노트, 기관과 관련된 이벤트 안내, 페이지에서 추가한 인스타그램, 유튜브 등과 같은 애플리케이션들이 제공되고 있다. 또한 〈그림 3〉의 오른쪽 하단을 보면 색깔별로 각 속성의 유형을 구분하였는데 노란색은 프로필(생산자)이나 일시(생산일), 해시태그와 같은 맥락 정보, 파란색은 소셜미디어 멘션에서의 내용 정보, 주황색은 시청각 관련 정보, 초록색은 이용자 피드백 정보를 표시한 것이다.

현재 페이스북에서는 Open API를 제공하지 않기 때문에 무료로 자동수집이 불가능하다. 따라서 페이스북을 수집할 시에는 별도의 상용수집 도구를 사용해야 하며, 페이스북의 경우 페이스북 전용 상용 수집 도구가 존재하기 때

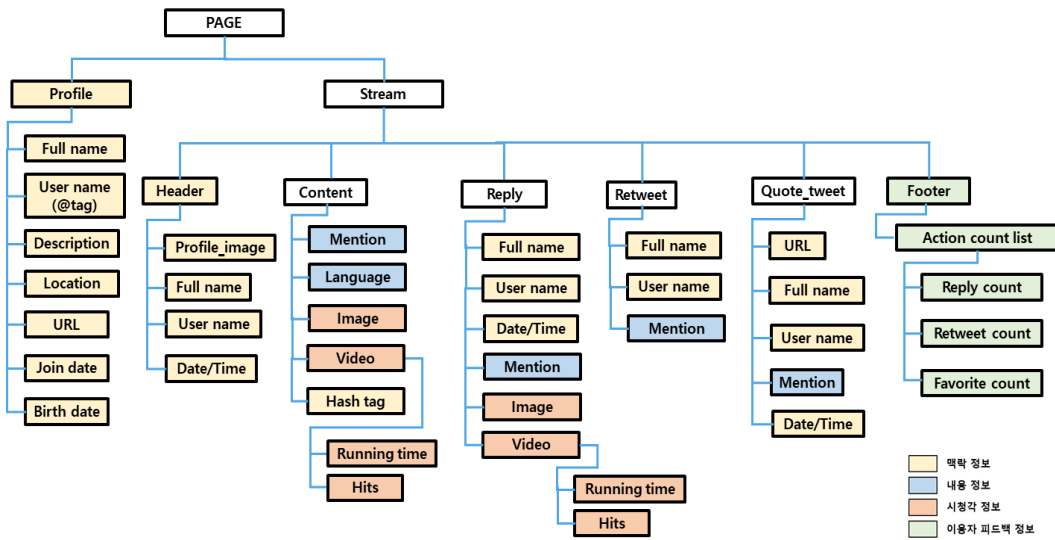
문에 이를 활용하거나 써드파티 서비스를 활용하여 아카이빙을 하는 것이 보다 효율적이다.

### 3.2.2 트위터(Twitter)

트위터는 대표적인 마이크로 블로그(microblog)로서 짧은 단문형 멘션을 통해 관계를 형성하거나 정보를 공유하기 때문에 전체적인 구조는 비교적 간단하다. 〈그림 4〉를 통해 경남도청 트위터 페이지를 대략적으로 살펴보면 왼쪽은 프로필 정보, 오른쪽은 해당 페이지의 게시물로만 구성되어 있어, 전체적으로는 비교적 간단한 구조로 보이지만 내부적으로 게시물 자체는 복잡한 다층적 구조로 구성되어 있다. 트위터 페이지를 구조화하면 〈그림 5〉와 같으며 크게 계정 프로필 정보와 페이지 게시물인 스트림으로 구성되어 있다. 트위터는 댓글 이외에도 트윗과



〈그림 4〉 경상도청 트위터 페이지 화면



〈그림 5〉 트위터 페이지 구조

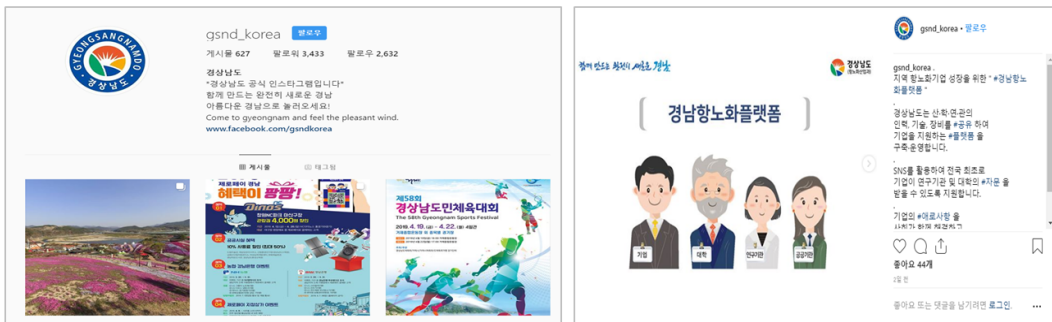
리트윗을 통해 이용자 피드백이 가능하기 때문에 해당 기능을 반영함으로써 스트림 구조가 비교적 복잡하게 구성된다. 트위터는 현재 다양한 유형의 Open API를 다수의 플랫폼에서 제공하

고 있기 때문에 API를 활용한 데이터 아카이빙 전략 적용이 가능하며, 미국 의회도서관 사례를 볼 때 협약을 통해 트위터를 통한 직접 이관도 가능하다.

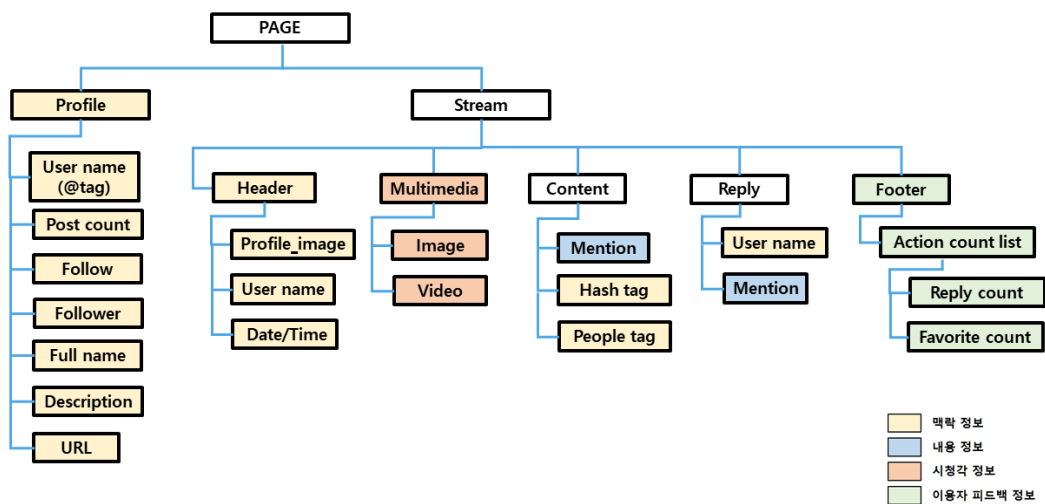
### 3.2.3 인스타그램(Instagram)

인스타그램은 사진이나 동영상을 공유할 수 있는 소셜미디어로서 게시물에 사진이나 동영상이 필수적으로 삽입되어야 한다. <그림 6>을 통해 경남도청 인스타그램 페이지를 대략적으로 살펴보면 상단에는 계정 프로필 정보, 하단에는 해당 페이지의 게시물만 구성되어 있어 전체적으로 비교적 간단한 구조로 구성되어 있다. <그림 7>을 통해 게시물 자체도 헤더, 멀티미디어, 내용, 댓글 등 비교적 간단한 구조로 구

성되어 있고, 인스타그램만의 특성으로 인해 게시물에는 사진, 동영상 등의 멀티미디어 자료가 필수적으로 삽입되어야 한다는 특징을 지니고 있다. 인스타그램 역시 아직까지는 Open API를 제공하지 않기 때문에 무료로 자동수집이 불가능하다. 따라서 인스타그램을 수집할 시에는 별도의 상용 수집 도구를 사용하거나 써드 파티 서비스를 활용하여 아카이빙을 하는 것이 보다 효율적이다.



<그림 6> 경남도청 인스타그램 페이지 화면



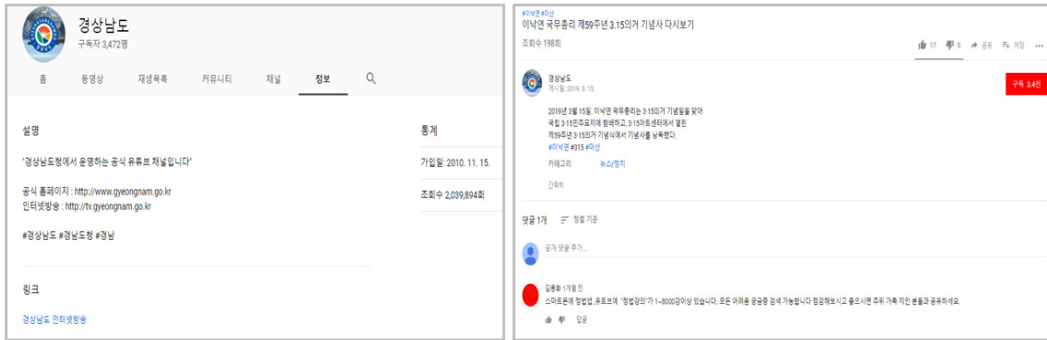
<그림 7> 인스타그램 페이지 구조

### 3.2.4 유튜브(Youyube)

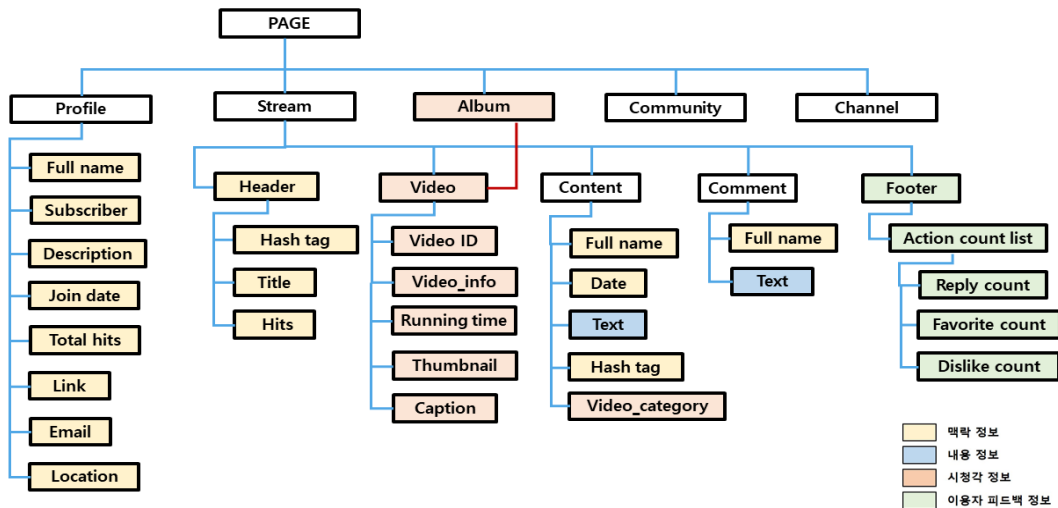
유튜브는 동영상 공유 서비스 플랫폼으로서 인스타그램과 마찬가지로 게시물에 동영상이 필수적으로 삽입되어야 한다. <그림 8>을 통해 경남도청 유튜브 페이지를 대략적으로 살펴보면, 상단에는 기본 프로필과 메뉴로 구성되어 있고, 하단에는 동영상 재생 목록을 중심으로 메뉴를 클릭할 시, 메뉴에 해당되는 정보를 보여주는 공간으로 구성되어 있다. 유튜브의 페이지는 기본

적으로 제공되는 기관 정보 및 동영상 게시물을 중심으로, 이를 재가공하여 제공하거나 별도의 커뮤니티 대화 화면 등으로 구성되어 있다. 특히, 동영상 특화 플랫폼이므로 게시물에는 동영상이 필수적으로 삽입되어야 하며 이에 대한 포맷 정보가 다수 존재한다.

유튜브 페이지를 구조화하면 <그림 9>와 같으며 계정 프로필 정보, 페이지 게시물인 스트림, 동영상 재생 목록인 앨범, 커뮤니티와 주고



<그림 8> 경남도청 유튜브 페이지 화면



<그림 9> 유튜브 페이지 구조

받는 대화 페이지, 추천 채널 페이지로 구성되어 있다. 유튜브 게시물에는 동영상이 필수적으로 삽입되어야 하므로 동영상과 관련된 보다 다양한 메타데이터가 존재한다. 유튜브 역시 Open API를 제공하지 않으며 무엇보다도 콘텐츠 저작권에 대한 법적, 제도적 문제가 선결되어야 한다.

### 3.3 소셜미디어 수집 도구

3.2절에서 기술한 바와 같이 소셜미디어는 다양한 유형마다 각각의 고유한 형태의 구조를 가지고 저장, 서비스되고 있다. 이러한 구조화된 기록물은 다양한 크롤링 도구를 활용해 수집 가능하다. 페이스북과 트위터는 자체적으로 부분적인 기능의 수집 도구를 제작하여 제공하고 있으며(이규철, 2007), 이 외에도 Archive-It, ArchiveSocial, Arkovi, MirrorWeb, SocialSafe 등 다양한 수집 도구들이 활용되고 있다. 그러나 현재 제공되는 수집 도구는 대부분 페이스북, 트위터, 유튜브 등의 일부 소셜미디어 기록에 대한 저장 방법만을 지원하며, 영구보존을 위한 수집 방법은 제공하고 있지 않다. 현재 제공되고 있는 소셜미디어 수집 도구는 크게 캡처(capture)와 데이터 저장 및 다운로드 방식

으로 구분할 수 있으며, 이 외에도 기타 기능을 제공하는 수집 도구가 일부 존재한다(〈표 4〉 참조).

캡처 방식은 소셜미디어 플랫폼에 특정 콘텐츠가 업로드되면 자동 또는 주기적으로 소셜미디어 스트림과 브라우저 활동을 캡처하여, 모든 장치에서 검색 및 접근할 수 있도록 하는 기능을 주로 제공한다. 데이터 저장 및 다운로드 방식은 소셜미디어 콘텐츠를 개인 백업장치에 저장하거나 스토리지 클라우드에 복사하여, 이를 다운로드할 수 있도록 하는 기능을 제공하는 방식이다. 이 외에도 기관이 자체 웹 아카이브를 구축하고 관리할 수 있도록 하는 기능을 제공하거나 여러 소셜미디어 계정을 모니터링할 수 있도록 대시보드를 제공하는 수집 도구가 일부 있다.

## 4. 소셜미디어 보존 방안

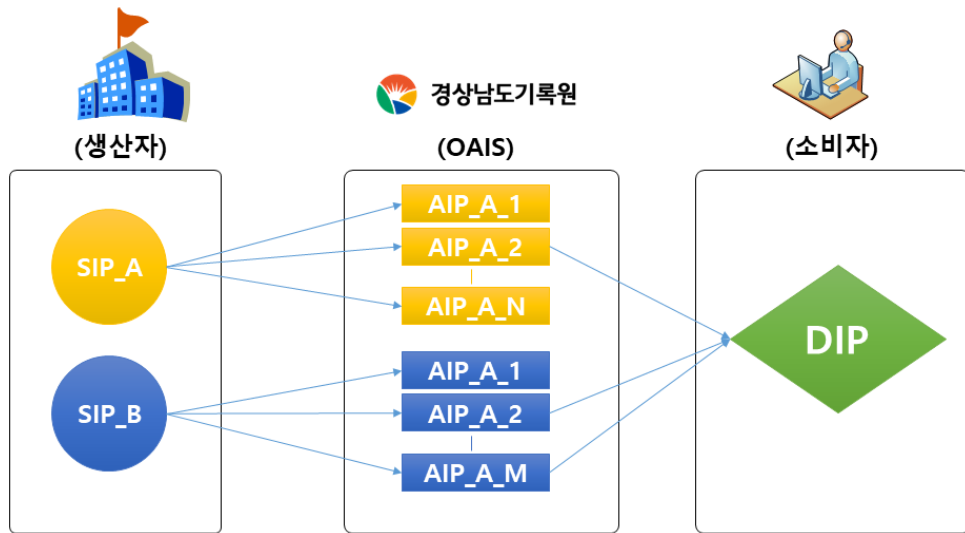
### 4.1 소셜미디어 보존 모형

본 연구에서는 소셜미디어에 적합한 문서보존포맷과 장기보존포맷을 도출함으로써 소셜미디어에 대한 보다 구체적인 보존 방안을 제

〈표 4〉 유형별 소셜미디어 수집 도구

유형 구분	수집 도구
캡처	Aleph Archives, Alfresco, Archify, Arkovi RegEd, Erado, Global Relay, Hearsay Social, Iterasi, LiveOffice, Cloud Preservation by Nextpoint, Ownbackup, Smarsh
데이터 저장 및 다운로드	ArchiveFacebook, Backupify for Personal Apps, Flickr Downloadr, FacebookArchive, Free YouTube Download, MirrorWeb, Parallel-Flickr, SocialSafe
기타	Archive-It, Hootsuite

출처: National Archives and Records Administration, 2013; 장인호, 황윤영, 이규철, 2014.



〈그림 10〉 소셜미디어 보존 모형: SIP, AIP, DIP 관계

안하고자 한다. 기본적으로 소셜미디어에 대한 장기보존포맷은 폴더형으로 설계하되, OAS 참조모형을 준수하기 위해 본 연구에서는 〈그림 10〉과 같은 보존 모형을 제안하였다. 다음 절에서는 소셜미디어 보존 모형에 활용된 OAS 참조모형의 3가지 정보 패키지인 제출 정보 패키지(이하, SIP), 보존 정보 패키지(이하, AIP: Archival Information Package), 정보 제공 패키지(이하, DIP: Dissemination Information Package)를 정의하고 각각의 특성에 맞게 폴더형 장기보존포맷을 제안하고자 한다.

#### 4.2 소셜미디어 문서보존포맷

Open API를 활용하여 소셜미디어 데이터를 수집하면 대부분 텍스트 위주의 데이터세트 형태로 수집될 것으로 예상되며 XML, JSON, CSV, ODS, SIARD 등의 '데이터세트형' 포맷으로 수집·이관될 것으로 판단된다. 그러므로, 데이터세

트형 포맷으로 수집되는 웹기록물들의 문서보존포맷으로 제안된 CSV, ODS, JSON, XML, SIARD의 5개 전자파일 포맷들이 소셜미디어 문서보존포맷으로 적합하다.

'데이터세트형' 포맷에서는 태그와 같은 일정한 표식("<.>", ",", ":" 등)으로 데이터와 데이터를 구분하고 기술하는 방식으로 ASCII, Unicode, EBCDIC와 같은 문자 인코딩 방식으로 표현된다. 미국 NARA(National Archives and Records Administration), 캐나다 LAC(Library and Archives Canada), 호주 NAA(National Archives of Australia)의 데이터세트 관점에서 선호·허용되는 포맷을 조사한 결과 CSV, XLS, ODS가 3곳에서 이관 가능하고, JSON, XLSX, XML가 2곳만 이관 가능하다(〈표 5〉 참조).

외부로부터 이관받는데 아니라 기록관에서 직접 수집하는 경우, 기업 의존성이 큰 XLS(X)를 제외하고 CSV, ODS, JSON, XML의 4개의 전자 파일이 문서보존포맷으로 적합하다고 분

〈표 5〉 국외 데이터세트 문서보존포맷 선호 및 허용 포맷 현황

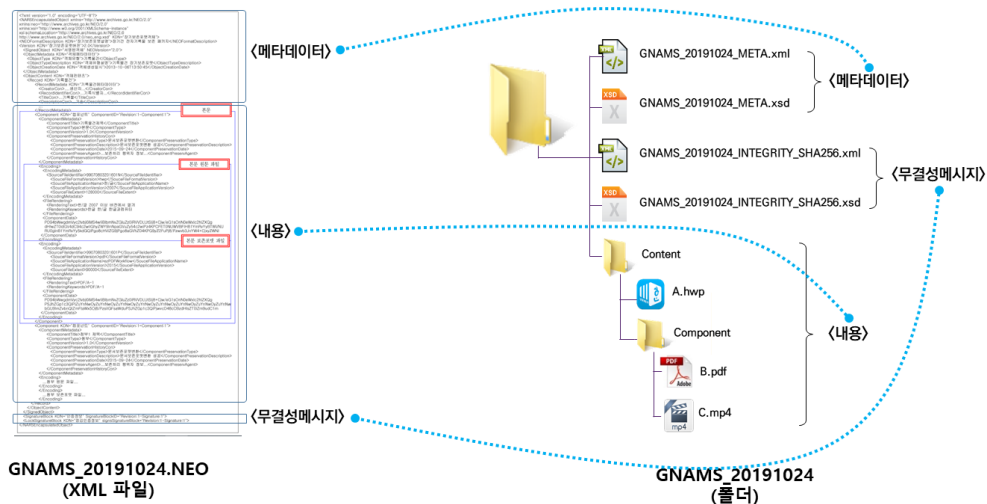
구분	미국 NARA		캐나다 LAC		호주 NAA	
	선호	허용	선호	허용	선호	허용
JSON	✓					✓
CSV	✓		✓			✓
XLS	XLS			✓		✓
	XLSX			✓		✓
ODS	✓			✓		✓
XML	✓					✓
DBF				✓		
SIARD						✓

석된다. 특히 XML은 적법성(well-formedness) 검사뿐만 아니라 유효성(validation) 기능까지 제공하기 때문에 가장 적절한 포맷으로 판단되며, 홈페이지 데이터베이스의 형태를 유지하면서 수집할 경우에는 특정 상용 SW에 의존성이 없는 SIARD의 형태가 적합하다. 무엇보다도 다수의 문서보존포맷을 선정할 수 있으며, 적절하지 않은 문서보존포맷은 제외할 수 있는 문서보존포맷 선정 기준 정의 및 선정 체계가 먼저 수

립되어야 한다.

### 4.3 소셜미디어 장기보존포맷

데이터세트형으로 수집된 소셜미디어 기록들은 파일형과 폴더형으로 관리할 수 있다. 경남도청을 예시로 한 〈그림 11〉을 통해 파일형과 폴더형 장기보존포맷 구조 및 서로 대응되는 요소들의 확인이 가능하다. XML 파일 내의 요소



〈그림 11〉 파일형과 폴더형 장기보존포맷 구조 및 대응 요소

(element) 였던 메타데이터, 무결성 메시지는 각각 xml/xsd 파일로 보관되고, 내용(원문, 첨부) 파일들은 특정 폴더(content)에 원문과 문서보존포맷으로 변환된 파일로 보관된다.

연구 결과, 장기적으로는 시스템 성능 측면과 관리의 용이성 모두를 위해 파일형 장기보존포맷에서 폴더형 장기보존포맷으로 변경하는 것이 바람직하다고 판단된다. 두 개의 패키징 방식은 기능적으로 동일(functionally equivalent)하지만 서로 상충하는 단점이 있으며, 파일형의 단점(생성 시간 및 저장소 증가)은 전자기록물이 증

가할수록 시스템 부담이 가중되지만, 폴더형의 단점(다수 개체 관리 비효율성 및 직렬화·비직렬화 시간 증가)은 IT 기술 발전으로 의해서 해결이 가능한 상황이다. 만약, 특정 전자기록물 유형에 대해서는 파일형 장기보존포맷을 도입하였다면, 새롭게 수용하는 유형에 대해서만 폴더형을 도입하고 안정화되면 다른 유형에 대해서도 폴더형 전환을 고려해야 한다. 장기보존포맷은 하나의 형식으로 다양한 전자기록 유형을 인캡슐레이션할 수 있기 때문에 소셜미디어의 장기보존포맷은 <표 6>과 같이 패키징할 수 있다.

<표 6> 소셜미디어 기록 장기보존포맷 제안

장기보존포맷	구분	내용
SIP (제출정보 패키지)	정의	• 하나 이상의 AIP 와/또는 해당 서술 정보를 생성하거나 업데이트에 사용하기 위해 생산자가 OAIS에 전달하는 정보 패키지(CCSDS 650.0.-M-2, 2012)
	설계방향	• 소셜미디어 기록은 기록원에서 수집하기 때문에 각 기록관에서 전달하지는 않지만 기록물 관련 메타데이터는 기록관에서 입력해야 함 • 기록원으로서의 이관, 기록관리시스템에 대한 지식, 네트워크 접속 및 인증 방식 등에 대한 전문지식이 없어도, 생산자가 손쉽게 메타데이터를 빠짐없이 작성할 수 있어야 함
	소셜미디어 기록물 SIP	• 상시적으로 처리과 → 기록관 → 기록원으로 이관되는 문서 유형과 달리, 소셜미디어 기록의 경우는 (1) 각 기관과 협의하여 합의된 기간에 크롤링하여 수집된 기록물과, (2) 수집하는 과정에서 수집된 메타데이터, (3) 수집된 소셜미디어 기록과 메타데이터를 입력으로 생성된 무결성 메시지가 SIP를 구성함 • 수집된 후 단일개체화(serialization) 방안으로 zip을 사용하는 방안 제안함
AIP (보존정보 패키지)	정의	• 콘텐츠 정보(CI)와 해당 보존 서술 정보(PDI)로 구성된 정보 패키지(CCSDS 650.0.-M-2, 2012)
	설계방향	• 오랫동안 보존되어야 하는 정보 패키지로 4대 속성을 유지하면서 보관되어야 하며, 이를 검증하기 위해서 주기적으로 전수 검사를 수행하여야 함 • 하나의 정보패키지를 검사하는 것과 전수 검사하는 것은 상황 자체가 다르기 때문에 전수 검사하는 상황을 고려하여 설계되어야 함
	소셜미디어 기록물 AIP	• AIP의 폴더형 장기보존포맷을 단일개체화 방안으로 zip을 사용한다면, 전수검사 시 압축해제 후 검사하고 다시 압축해야 하기 때문에 장기간 소요 • 이를 위해, zip 이 아닌 ASN.1과 같이 플랫폼 독립적인 방식으로 데이터 구조를 만들어 단일개체화하는 방안을 제안
DIP (제공정보 패키지)	정의	• OAIS에 대한 요청에 응답하여 아카이브가 소비자에게 보내는 정보패키지로 하나 이상의 AIP에서 파생됨(CCSDS 650.0.-M-2, 2012)
	설계방향	• DIP는 소비자가 열람하거나 확인하기 위한 정보 패키지로 해당 시대에 확인할 수 있는 형태로 제공되어야 함
	소셜미디어 기록물 DIP	• 전수검사를 위해 ASN.1과 같은 방식으로 단일개체화 되었다면, 다시 폴더구조로 변경하기 위해 단일개체화 해제(deserialization)하고 각각의 전자파일들을 해당 시대에 열람이 가능한지 확인하고 필요시 변환해야 함

## 5. 시범구축을 통한 검증

### 5.1 운영 현황

앞서 설명한바와 같이 본 연구에서는 앞서 제시한 방안의 실효성을 검증하기 위해 경상남

도기록원 관리 대상 기관의 소셜미디어 개설·운영 현황을 조사하여 수집 대상을 선정하였다. 2019년 2월 기준으로, 경상남도기록원 관리 대상 기관 31개 중 소셜미디어를 개설·운영하는 기관은 28개(약 90%)이며(〈표 7〉 참조), 페이스북, 네이버 블로그, 트위터 순으로 운영하고

〈표 7〉 경상남도기록원 관리 대상 기관 소셜미디어 운영 현황(2019.02.14. 기준)

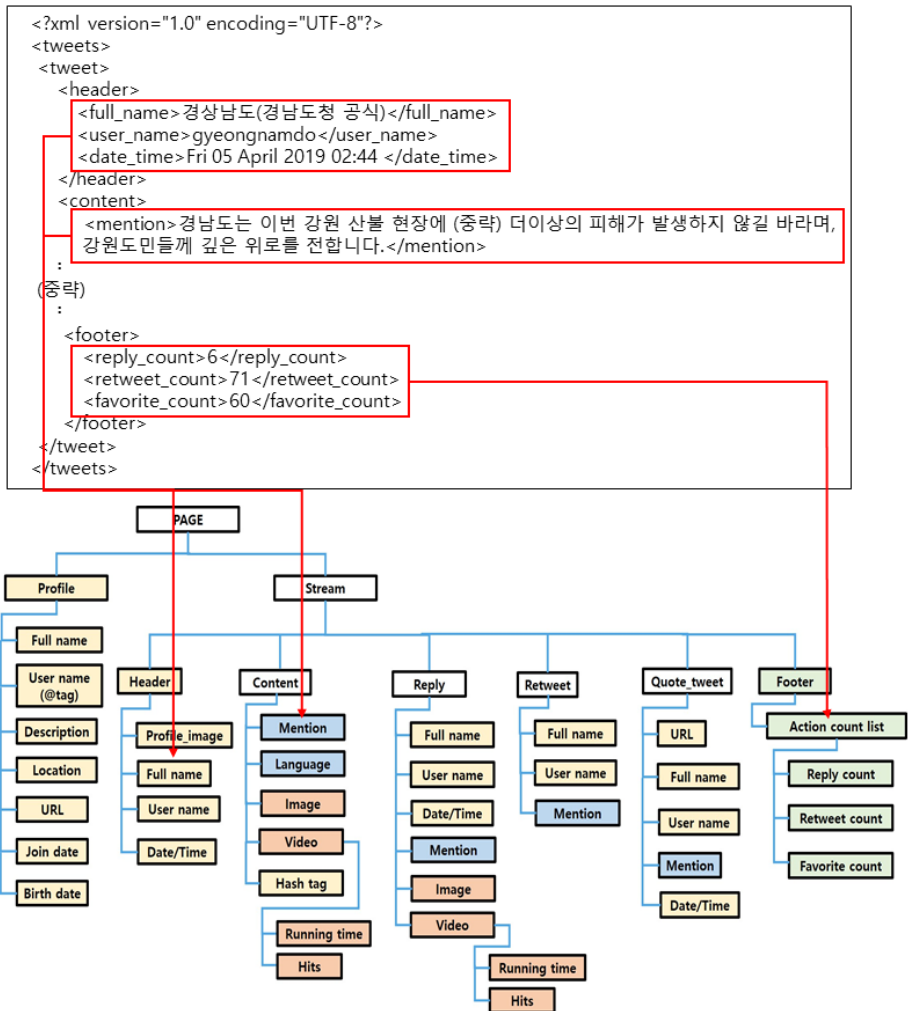
구분	기관명	SNS 수	소셜미디어 유형	
1	도	경상남도 본청	7	페이스북, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브, 핀터레스트
2	시	거제시	3	페이스북, 네이버 블로그, 인스타그램
3	시	김해시	6	페이스북, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브
4	시	밀양시	6	페이스북, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브
5	시	사천시	6	페이스북, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브
6	시	양산시	6	페이스북, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브
7	시	진주시	5	페이스북, 다음 블로그, 트위터, 인스타그램, 카카오톡
8	시	창원시	6	페이스북, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브
9	시	통영시	4	페이스북, 네이버 블로그, 트위터, 인스타그램
10	군	거창군	3	페이스북, 네이버 블로그, 트위터
11	군	고성군	4	페이스북, 네이버 블로그, 트위터, 밴드
12	군	남해군	4	페이스북(2), 네이버 블로그, 인스타그램
13	군	산청군	8	페이스북, 블로그, 네이버 블로그, 트위터, 인스타그램, 카카오톡, 유튜브, 핀터레스트
14	군	의령군	3	페이스북, 네이버 블로그, 인스타그램
15	군	창녕군	4	페이스북, 네이버 블로그, 트위터, 카카오톡
16	군	하동군	4	페이스북, 네이버 블로그, 트위터, 인스타그램
17	군	함안군	3	페이스북, 네이버 블로그, 트위터
18	군	함양군	4	페이스북, 네이버 블로그, 트위터, 카카오톡
19	군	합천군	4	페이스북, 네이버 블로그, 트위터, 인스타그램
20	공사	경남개발공사	1	페이스북
21	공사	거제해양관광개발공사	3	페이스북, 블로그, 다음 블로그
22	공사	김해시도시개발공사	1	페이스북
23	공사	밀양시시설관리공단	0	-
24	공사	양산시시설관리공단	1	카카오톡
25	공사	창녕군시설관리공단	1	페이스북
26	공사	창원경륜공단	3	페이스북, 네이버 블로그, 인스타그램
27	공사	창원시시설관리공단	1	페이스북
28	공사	통영관광개발공사	0	-
29	공사	함안지방공사	0	-
30	도립대학	경남도립 거창대학	3	페이스북, 네이버 블로그, 트위터
31	도립대학	경남도립 남해대학	3	페이스북, 네이버 블로그, 트위터
합계			107	

있다. 이 중 본 연구에서는 Open API를 통해 자동으로 수집 가능한 경상남도 본청 공식 트위터 계정(@gyeongnamdo)을 주요 연구 대상으로 선정하였다.

### 5.2 시범 수집

선정 대상인 경남도청 트윗 기록을 수집하기

위해 자체적으로 트위터 수집도구를 개발하였다. 수집도구는 트위터 수집 API인 'GetOld Tweets-python'와 'selenium', 'BeatifulSoup', 'urllib', 'json', 're', 'cookiejar' 등의 라이브러리를 활용하여 파이썬(Python)으로 개발하였다. 앞서 3.2장에서 예시한 <그림 5>의 트윗 페이지는 실제로는 <그림 12>와 같이 구성되어 있으며, 이를 xml로 수집하여 <표 8>과 같은 SIP로



<그림 12> 경남도청 트위터 XML과 페이지 구조

〈표 8〉 경남도청 트위터 SIP 예시

Metadata		Information
Header	Profile_image	
	Full name	경상남도(경남도청 공식)
	User name	@gyeongnamdo
	Date/Time	2019년 4월 5일 / 오전 2:44
Content	Mention	경남도는 이번 강원 산불 현장에 펌프차량 16대, 구급차 2대, 물탱크차량 2대 등 소방장비 포함 소방인력 55명을 긴급 지원했습니다. 또한 자원봉사자 280여 명을 추가 모집해 추후 수습 지원에도 적극 나설 예정입니다. 더이상의 피해가 발생하지 않길 바라며, 강원도민들께 깊은 위로를 전합니다.
	Language	Korean
	Photo	
Reply	Full name	오솔문
	User name	@hwangsji
	Date/Time	2019년 4월 5일 / 오전 3:18
	Mention	하아~ ~모두 감사합니다 ㅠ ㅠ
Retweet	Full name	민심은천심이다
	User name	@Spw1lmoajggjuF8
	Mention	공수처로 하나가 됩시다!! 공수처를 설치하라!!! 맞팔은 예의조잉^^ 잠깐! 낙지 무자게 싫어함!
Footer	Reply count	6
	Retweet count	71
	Favorite count	60

변환·저장된다. 제출 정보 패키지(SIP)는 수집 대상인 콘텐츠와 그에 대한 메타데이터로 구성되며, 소셜미디어 플랫폼에서 생성된 SIP가 기관으로 수집되면, 이를 검증한 다음에 보존 정보 패키지(AIP) 생성이 가능하다. 이 때 SIP 구성 시에 〈표 8〉과 같이 멘션에 대한 내용만 수집하는 것

이 아니라 주요 메타데이터도 함께 수집한다는 점이 중요하다. 메타데이터는 앞서 분석된 소셜 미디어의 페이지 구조를 토대로 기존에 기록관리 차원에서 중요시되는 필수 혹은 공통항목과 더불어 소셜미디어의 맥락을 반영하고 이용가능성을 높이기 위한 항목으로 구성되어 있다.

다음은 수집도구를 통해 수집된 트윗 기록의 세부 분석 결과이다. 트위터 개설일로부터 트윗이 마지막으로 게시된 기간을 기준(2010.01 ~ 2019.03)으로 트윗 수, '좋아요' 수, '리트윗' 횟수, 댓글 수를 <표 9>와 같이 수집하여 계량 분석을 수행하였다. 수집 결과, 약 4천여 개의 트윗 기록이 수집되었고, 이용자 반응 지수는 각각 '좋아요' 횟수가 15,965개, '리트윗' 횟수는 21,426으로 나타났다.

### 5.3 소셜미디어 수집 결과 활용

모든 기록의 관리의 궁극적인 목적은 활용에 있다. 소셜미디어 기록 역시 수집, 관리된 이후 이를 이용자에게 서비스하거나 수집된 기록으로부터 유의미한 현상을 파악하는 것이 해당 기록의 가치를 높이는데 매우 중요한 과정이다. 본 연구에서는 제안된 방법을 통해 수집·보존된 경남도청 트윗 기록을 통해 운영현황을 시계열 기반으로 살펴보고 주요 사건을 살펴봄으

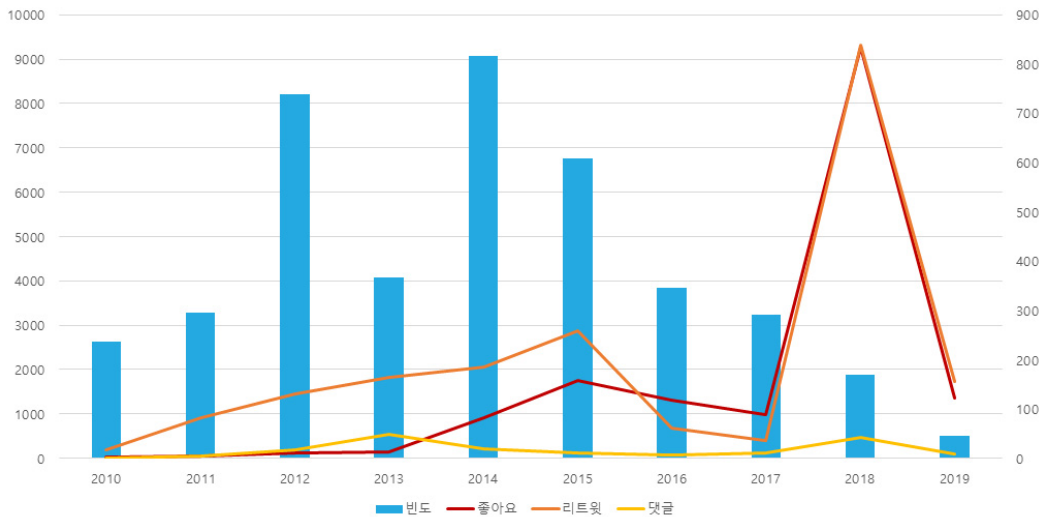
로써 소셜미디어의 아카이빙의 중요성을 상기시키고자 한다.

<표 9>의 경남도청 트윗 수집 결과의 연도별 추이를 <그림 13>과 같이 살펴보면 대체로 2014년 이후 트윗 수는 감소하는 반면 '리트윗' 횟수와 '좋아요' 수는 급증하는 추세이다. 2012년과 2014년에 트윗 수 대비 이용자 피드백이 낮았으나 2017년 이후 반전되는 급격한 변화를 보이고 있다. 이는 경남도청이 트위터를 통해 제공하는 정보의 양이 감소하였음에도 불구하고 '리트윗', '좋아요' 수 등 이용자 피드백이 증가한 것은 점차 시간이 지날수록 기관에서 이용자에게 보다 양질의 정보를 제공하고 있음을 의미한다. 트위터 운영 현황의 전반적인 추세를 살펴봤을 때, 경남도청에서는 점진적으로 이용자의 관심도가 높은 정보를 선별하여 트위터에서 제공하고 있으며, 이에 대한 이용자의 만족도가 점차 높아지고 있음을 시사한다.

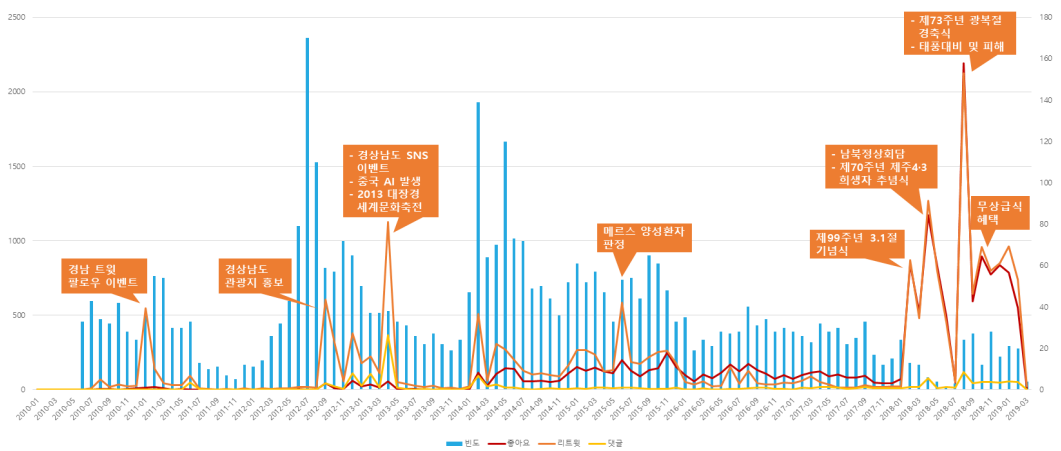
경남도청 트위터의 운영 현황을 보다 면밀하게 분석하기 위해 트윗 수와 이용자 피드백 지

<표 9> 경남도청 트윗 연도별 수집 결과

연도별 수집현황	수집 결과			
	트윗 수	좋아요 수	리트윗 수	댓글 수
2010	3,911	15,965	21,426	1,939
2011	296	57	925	59
2012	738	132	1,442	190
2013	366	151	1,823	552
2014	816	924	2,056	225
2015	608	1,756	2,873	119
2016	345	1,317	673	79
2017	291	984	409	130
2018	170	9,270	9,319	477
2019	45	1,347	1,724	108
합계	3,911	15,965	21,426	1,939



〈그림 13〉 경남도청 연도별 트위터 현황



〈그림 14〉 경남도청 월별 트위터 현황

수, 키워드에 대한 시계열 분석을 수행하였다. 시계열 분석을 통해 이용자 피드백 지수가 높은 정보 요인이 무엇인지 분석한 결과, 〈그림 14〉와 같이 경남도청 트위터에서는 경상남도에 특화된 뉴스 뿐 아니라 사회적 파장이 비교적 큰 특정 이슈 사건(AI 발생, 메르스 양성환자 관전, 태풍대비 및 피해 등)에 대한 이용자

의 관심이 높음을 확인할 수 있다. 특히 2018년도 이후부터는 생성된 트윗 수에 비해 이용자 반응 지수가 매우 상승한 것을 그래프를 통해 확인할 수 있다.

이상으로 경남도청 트윗 기록을 통해 운영현황을 시계열 기반으로 분석한 결과, 현재 경남도청 트위터에서는 주기별로 도출되는 특정 이

슈가 없으므로 별도의 특별한 수집 주기 설정은 불필요하며, 업데이트되는 트윗 수가 많은 편이고, 정보 유형이 다양하므로 월별로 수집 주기 설정하는 것이 바람직하다. 이 외에 특정한 사회적 이슈 발생 시에만 별도로 트윗을 수집한다. 이처럼 트위터 운영 현황 분석을 통해 도출된 수집 주기를 기반으로 소셜미디어 유형별로 정의된 수집 전략 및 SIP에 따라 소셜미디어 기록을 수집해야 하며, 이에 대한 검증 과정을 거친 후 AIP를 생성하여 아카이빙해야 한다. 차후에는 경상남도 유관기관에서 소셜미디어 계정을 새로 생성하거나 변경·삭제할 경우를 대비하여, 소셜미디어 계정의 변경 이력을 경상남도기록원에 보고하거나 등록하는 체계 마련이 필요하다.

## 6. 결론

본 연구는 소셜미디어의 유형별 특징과 그에 따른 수집 프로세스와 전략, 보존 모형 제안을 목적으로 소셜미디어 아카이빙 전략 및 프로젝트의 다양한 사례와 보존 포맷 적용 현황을 조사·분석하였으며, 실제 수집한 결과를 통해 활용 방안을 제시함으로써 소셜미디어 기록의 가치를 높이는데 일조하고자 했다.

소셜미디어 아카이빙 프로젝트 사례로부터 도출된 시사점은 다음과 같다. 첫째, 소셜미디어 수집 시에는 공개 소프트웨어나 전용 API를 활용하는 것이 효율적이다. 둘째, 텍스트 이외에 이미지, 동영상 등 다양한 콘텐츠에 대한 수집정책을 수립해야 한다. 셋째, 소셜미디어에 대한 이용자와의 상호작용 정보를 함께 수집할

필요가 있다. 넷째, 수집 이외에도 이용자 정보 제공 측면에 대한 고려가 필요하다.

이상의 시사점을 바탕으로 본 연구에서는 국가 차원에서 공공영역의 소셜미디어를 수집 및 보존하기 위한 실행 방안을 마련하고자 하였다. 그 결과로 소셜미디어 수집 및 관리 시스템의 기본 구조를 제안하였으며, 이를 토대로 페이스북, 트위터, 인스타그램, 유튜브 등 총 4가지 유형별로 소셜미디어의 페이지 구조와 적합한 수집 전략에 대해 개별적으로 제시하였다. 페이스북, 인스타그램, 유튜브의 경우에는 Open API를 제공하지 않기 때문에 별도의 상용 수집 도구를 사용해야 하고, 트위터의 경우에는 다수의 플랫폼에서 Open API를 제공하고 있기 때문에 API를 활용한 데이터 아카이빙 전략 적용이 가능하다. 이 외에도 멀티미디어 콘텐츠나 이용자와의 상호작용 정보 수집을 위한 메타데이터도 함께 제안하였다.

또한 본 연구에서는 소셜미디어에 적합한 문서보존포맷과 장기보존포맷을 도출함으로써 소셜미디어에 대한 구체적인 보존 방안을 제안하고자 하였다. 그 결과로 문서보존포맷으로는 데이터세트형 포맷 중 XML이나 SIARD의 형태가 적합하고, 장기보존포맷으로는 시스템 성능과 관리 측면에서 폴더형이 적합하다는 결과가 도출되었다. 더불어 제안방안의 실효성을 검증함으로써 연구에 구체성을 더하였고, 이를 통해 시계열 기반으로 운영 현황을 분석하여 주요 시점의 시대상을 포착함으로써 아카이빙의 활용성을 상기시켰다. 향후 본 연구는 국가적 차원에서 소셜미디어 아카이빙을 위한 보다 구체적인 수집 및 보존 모델을 수립할 때 도움이 될 것으로 기대한다.

## 참 고 문 헌

- 경상남도기록원 (2018). 경상남도기록원 영구기록관리시스템(AMS) 구축 제안요청서. 창원: 경상남도기록원.
- 김윤화 (2013). 미국의 소셜미디어 이용 현황. 정보통신정책연구, 25(1), 43-49.
- 네이버시스템 (2017). 경상남도기록원 정보화전략계획(ISP) 수립 용역 완료보고서 II. 창원: 경상남도행정국.
- 도서관연구소 (2012). 도서관 소셜미디어 서비스. 서울: 국립중앙도서관.
- 도서관연구소 (2013). 미 의회도서관 트위터 아카이브의 최신 현황 보고. 서울: 국립중앙도서관.
- 이귀영, 김지현 (2015). 중앙행정기관의 소셜미디어 아카이브 도입 방안에 관한 연구. 한국기록관리학회지, 15(3), 141-167. <http://dx.doi.org/10.14404/JKSARM.2015.15.3.141>
- 이규철 (2007). 행정기관 SNS 기록물 수집 보존 재현 도구 연구. 대전: 국가기록원.
- 이윤희 (2014). 국내 SNS의 이용 현황과 주요 이슈 분석. Internet & Security Focus, 56-78.
- 임진희 (2013). 전자기록관리론. 서울: 선인.
- 장인호, 황운영, 이규철 (2014). 정부부처의 페이스북 페이지 기록물 수집·보존·재현 방법. 한국기록관리학회지, 14(2), 117-128. <http://dx.doi.org/10.14404/JKSARM.2014.14.2.117>
- 정영철, 이기호, 배하석, 신철민 (2012). 보건복지 부문의 소셜미디어 활용 현황 및 정책과제. 세종: 한국보건사회연구원.
- 최두원, 이수진, 윤은하, 오효정 (2018). 대통령 SNS 기록물 관리방안에 관한 연구. 한국기록관리학회지, 18(2), 29-59. <http://dx.doi.org/10.14404/JKSARM.2018.18.2.029>
- 한상호 (2018). 영국(TNA)의 웹 아카이브 동향 조사. 기록인, 45, 24-29.
- FKII조사연구팀 (2006). 소셜미디어(Social Media)란 무엇인가. FKII Report, 242, 52-53.
- National Archives and Records Administration (2013). Best practices for the capture of social media records. Maryland: National Archives and Records Administration.

### [ 법률 ]

- “공공기록물 관리에 관한 법률”(법률 제14613호, 2017.9.22).
- “공공기록물 관리에 관한 법률 시행령”(대통령령 제29563호, 2019.2.26).

### [ 웹사이트 및 인터넷 자료 ]

- 국립중앙도서관 (2018). 미국의회도서관, 트위터 아카이빙 정책을 바꾸다. Retrieved from [https://wl.nl.go.kr/usr/com/prm/BBSDetail.do?bbsId=BBSMSTR\\_000000000456&nttId](https://wl.nl.go.kr/usr/com/prm/BBSDetail.do?bbsId=BBSMSTR_000000000456&nttId)

=5811&menuNo=12002&upperMenuId=12

- 정보인권 (2018). 소셜미디어 기록물: 소셜미디어 기록물 관리에 관한 제언. Retrieved from <https://act.jinbo.net/wp/38449/>
- 황준범 (2018). 트럼프의 트위터. Retrieved from <http://www.hani.co.kr/arti/opinion/column/842872.html>
- 허완 (2017). 미국 의회도서관이 야심찬 ‘역대 모든 트윗 아카이브’ 구축 프로젝트를 중단한 이유. Retrieved from [https://www.huffingtonpost.kr/2017/12/27/story\\_n\\_18909336.html](https://www.huffingtonpost.kr/2017/12/27/story_n_18909336.html)
- PANDORA (2019). Retrieved from <http://pandora.nla.gov.au/tep/162818>
- Public Record Office Victoria (2017). Social media records: Advice for managing social media records. Retrieved from <https://prov.vic.gov.au/recordkeeping-government/a-z-topics/social-media-records>

• 국문 참고문헌에 대한 영문 표기  
(English translation of references written in Korean)

- Choi, Doo-Won, Lee, Su-Jin, Youn, Eun Ha, & Oh, Hyo Jung (2018). A study on a presidential sns records management method. *Journal of Korean Society of Archives and Records Management*, 18(2), 29-59. <http://dx.doi.org/10.14404/JKSARM.2018.18.2.029>
- Chung, Youngchul, Lee, Gi-Ho, Bae, Hasuk, & Shin, Cheol Min (2012). A study of utilization of social media and policy implications in health and welfare. Sejong: Korea Institute for Health and Social Affairs.
- Digital Rights (2018). Social media records: Advice for managing social media records. Retrieved from <https://act.jinbo.net/wp/38449/>
- FKII Research Team (2006). What is social media?. *FKII Report*, 242, 52-53.
- Gyeongsangnam-do Archives (2018). Request for proposal of archives management system of gyeongsangnam-do archives. Changwon: Gyeongsangnam-do Archives.
- Han, Sang-Hyo (2018). Web archive trend survey of uk (tna). *Archival In*, 45, 24-29.
- Heo, Wan (2017). Reasons why the library of congress dropped the ‘All Tweets Archive’ project. Retrieved from [https://www.huffingtonpost.kr/2017/12/27/story\\_n\\_18909336.html](https://www.huffingtonpost.kr/2017/12/27/story_n_18909336.html)
- Hwang, Junbeom (2018). Trump’s twitter. Retrieved from <http://www.hani.co.kr/arti/opinion/column/842872.html>
- Jang, In-Ho, Hwang, Yun-Young, & Lee, Kyu-Chul (2014). The methods of collecting, preservation, reproduction for records of public sector’s facebook pages. *Journal of Korean Society of*

- Archives and Records Management, 14(2), 117-128.  
<http://dx.doi.org/10.14404/JKSARM.2014.14.2.117>
- Kim, Yunhwa (2013). Social media utilization status in united states. Korean Telecommunications Policy Review, 25(1), 43-49.
- Lee, Gwi-Yeong, & Kim, Jihyun (2015). A study on the introduction of a social media archive to the central government. Journal of Korean Society of Archives and Records Management, 15(3), 141-167. <http://dx.doi.org/10.14404/JKSARM.2015.15.3.141>
- Lee, Kyu-Chul (2007). The development of collection, preserving and reproducing tools for sns records in administrative organizations. Daejeon: National Archives of Korea.
- Lee, Yun Hee (2014). An analysis of the domestic utilization status and major issues of social network service. Internet & Security Focus, 56-78.
- Library Institute (2012). Library social media services. Seoul: National Library of Korea.
- Library Institute (2013). Report on the latest status of the library of congress twitter archives. Seoul: National Library of Korea.
- National Library of Korea (2018). Library of congress, twitter archiving policy change. Retrieved from  
[https://wl.nl.go.kr/usr/com/prm/BBSDetail.do?bbsId=BBSMSTR\\_00000000456&nttId=5811&menuNo=12002&upperMenuId=12](https://wl.nl.go.kr/usr/com/prm/BBSDetail.do?bbsId=BBSMSTR_00000000456&nttId=5811&menuNo=12002&upperMenuId=12)
- Neighbor System (2017). Completion report on establishing an information strategic plan (ISP) II. Changwon: Gyeongsangnam-do Administrative Bureau.
- Yim, Jin Hee (2013). Electronic records management. Seoul: Sunin.