

국가 과학데이터 공유·활용 서비스를 위한 사례 연구

진영근, 이원구*
충남도립대학교 컴퓨터정보과

A Case Study on Sharing & Using of National Scientific Data

Young-Goun Jin, Won-Goo Lee*

Department of Computer Information, Chungnam Provincial University

요 약 모든 분야에서 과학데이터의 생산 및 수집에 많은 비용과 시간이 소요되고 있으며, 또한 공유와 재활용 정도가 매우 낮은 실정이다. 학제간 연구를 지원하기 위해서는 대량으로 발생하고 있는 과학데이터가 체계적으로 보존 및 공유되어야 하며, 또한 반복할 수 없거나 반복하는데 많은 비용이 발생하는 실험 및 관찰 데이터의 보존과 재사용 및 활용 환경 구축이 필수적이다. 이에, 본 연구에서는 국가 과학데이터를 수집·저장·관리·공유·활용할 수 있는 서비스와 국가 R&D 관련 기관, 국제적인 과학기술 단체 및 기관과의 다양한 과학데이터 연동 인터페이스를 제안한다. 이를 통해, 현재 개발되어 있는 기존 과학 데이터베이스들의 활용성을 증대시키고 산업적 파급 효과가 높은 과학 데이터를 더욱 개발하여 국가 연구개발 활동을 지원하는 과학 데이터 인프라를 확충하는데 기여하고자 한다.

주제어 : 과학데이터, 연구데이터, 데이터 공유, 데이터 재활용, 데이터 융합

Abstract Production and collection of scientific data in all areas is costly and time consuming. The level of sharing and recycling of scientific data is also very low. In order to support interdisciplinary research, massive scientific data should be systematically preserved and shared. In addition, it is essential to build an infrastructure to preserve and utilize the costly experiment and observation data. In this study, we propose a service that can collect, store, manage, share, and utilize national science data. It also suggests interfaces for various scientific data interactions with national R & D related organizations, international scientific and technological organizations and institutions. This will increase the availability of existing scientific databases that are currently being developed.

Key Words : Scientific Data, Research Data, Data Sharing, Data Re-use, Data Convergence

1. 서론

모든 분야에서 과학데이터의 생산 및 수집에 많은 비용과 시간이 소요되고 있으며, 또한 많은 비용과 시간이 투자된 과학데이터가 개인 연구자나 기관에 의해 산발적으로 축적되어 체계적 관리가 부족하거나, 특정 조직에 의해 독점되어 공유와 재활용 정도가 매우 낮은 실정이다[1,2].

일부 분야에서는 해외의 과학데이터에 대한 의존도가 높아 많은 비용이 소요되거나 국내 실정에 적합하지 않

는 데이터가 활용되는 경우가 있어 연구의 효율성 및 신뢰성을 저해시키며, 또한 중복 투자나 유사 연구 반복 등의 문제를 초래함으로써 궁극적으로 경제적 손실을 야기하고 있다[3].

대용량 데이터의 분석 및 응용을 위해서는 슈퍼컴퓨터 및 가상화 시스템 등의 하드웨어와 관련 기술들이 요구되지만, 이를 연구기관 또는 연구자가 개별적으로 보유하거나 효율적으로 활용할 수 있는 IT 기술이 부족하여 이는 정보의 가공 및 고부가가치화의 한계점으로 작용하고 있다[4,5].

*Corresponding Author : Won-Goo Lee (dryope19@gmail.com)

Received March 20, 2018

Accepted August 20, 2018

Revised June 27, 2018

Published August 28, 2018

학제간 연구를 지원하기 위해서는 대량으로 발생하고 있는 과학데이터가 체계적으로 보존 및 공유되어야 하며, 또한 반복할 수 없거나 반복하는데 많은 비용이 발생하는 실험 및 관찰 데이터의 보존과 재사용 및 활용 환경 구축이 필수적이다[6-8].

본 연구에서는 국가 과학데이터를 수집·저장·관리·공유·활용할 수 있는 서비스와 국가 R&D 관련 기관, 국제적인 과학기술 단체 및 기관과의 다양한 과학데이터 연동 인터페이스를 제안한다. 또한, 과학데이터 센터 간 IT 인프라스트럭처, 과학데이터를 공유·활용하기 위한 국가 과학데이터 플랫폼 기반의 과학데이터 연동 표준 인터페이스를 제시한다.

2. 해외의 공유·활용 플랫폼 분석

2.1 iRODS

2.1.1 개요 및 기능

iRODS(integrated Rule Oriented Data System)은 오픈 소스로 제공되고, 데이터 관리 소프트웨어로서 규모가 크고, 중요하고, 복잡한 데이터를 관리하고 있으며, 수백 개 기관이 자신들의 연구데이터를 관리하기 위해서 iRODS를 사용 중에 있다. iRODS는 수십억 개의 파일, 수백 페타바이트의 데이터를 관리하기 위해서 데이터 가상화, 데이터 복구, 워크플로우 자동화, 보안 협력과 같은 네 개의 주요 기능을 제공하고 있다[9,10].

iRODS 존(Zone)은 iRODS 서버 소프트웨어가 구동 중인 컴퓨터들의 네트워크로서, 존은 연결된 저장 장치들에 보관된 데이터와 메타데이터 목록에 적재된 메타데이터를 제공한다. 하나의 존에는 목록서버 역할을 하는 하나의 서버가 있으며, 목록서버는 목록정보를 보유하고 있는 RDBMS와 연결되어 있다. 또한, 하나의 존에서 목록서버를 제외한 다른 서버들은 목록 서비스를 제공받으며, 모든 iRODS 서버들은 iRODS 클라이언트들의 접속 요청을 받고, 모든 iRODS 서버들은 Zone에 의해 제공되는 파일들을 보유한 스토리지 자원들을 운영(host) 할 수 있다.

모든 iRODS 서버들은 워크플로우 자동화를 구현한 iRODS 규칙들(Rules)을 실행 할 수 있으며, 서버를 추가함으로써 성능, 보안을 향상시킬 수 있으며, 또한 다중화를 통한 Zone의 장애대응(resilience) 능력을 높일 수 있음. 한 지역 내에서 혹은 지리적으로 분산된 환경에서 모두 가능하다.

2.1.2 데이터 가상화 및 복구

Fig. 1과 같이 iRODS는 물리적인 저장소에 적재된 파일들을 논리적 표현으로 제공하며, 이것을 데이터 가상화 부르며, iRODS에 적재된 데이터는 iRODS 클라이언트를 통해서 접근가능하다. 또한, iRODS 클라이언트들은 파일들을 컬렉션에 조직된 데이터 객체들로 표현하고, 데이터 객체들과 파일들, 컬렉션들과 서버 디렉토리들은 큰 차이가 없으나 다음과 같이 중요한 구분점이 있으며, iRODS 서버들은 데이터 전송 동안에 데이터 브로커 역할을 수행한다[9,10].

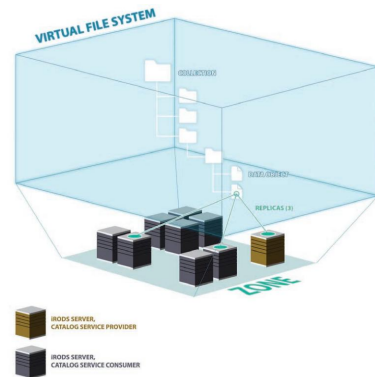


Fig. 1. iRODS Virtual File System

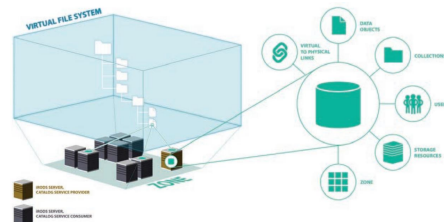


Fig. 2. iRODS VFS vs. Zone

Fig. 2와 같이 메타데이터 목록은 Zone의 데이터 객체들에 대한 정보를 보유하고 있으며, 또한, 컬렉션들과 사용자들, Storage Resources, Zone 자체에 대한 정보를 가지고 있다. 메타데이터들은 모두 관계형 데이터베이스에 있고, 메타데이터들은 PostgreSQL이나 MySQL 혹은 Oracle 데이터베이스 관리 시스템에 있어야 한다.

2.1.3 워크플로우 자동화 및 보안 협업

각각의 iRODS 서버는 이벤트에 의해 백그라운드에서 실행되는 규칙 엔진(Rule Engine)을 운영하고, 규칙 엔진은

iRODS Rules를 이용해서 프로그램 된다. 이때, iRODS Rules은 iRODS가 특정 시스템 행위를 했을 때 어떤 행위가 실행되어야 하는지(be triggered) 명시해 놓은 것이며, iRODS 규칙 엔진은 플러그인으로 확장 될 수 있다.

이러한 규칙은 프로그래밍 언어에 상관없이 플러그인이 작성된 프로그래밍 언어로 정의될 수 있다. 일반적으로 iRODS 순수(native base) Rule Engine은 도메인에 특화된 c 프로그래밍 언어(iRODS Rule Language)를 사용하며, Rule Language는 microservices라 불리는 외부 함수들(functions, 기능들)을 이용하고, 이미 iRODS에 구현되어 있다.

iRODS는 데이터를 출판하는 분야가 아니라도 여러 워크그룹 내에서 데이터를 공유해야 하는 것을 지원하기 위해 Tickets, Permissions, Federation 기술을 지원하고 있다.

우선, Tickets은 데이터 객체와 컬렉션으로의 통제된 공공 접근을 제공하며, non-iRODS 이용자들에게 티켓을 제공해 파일을 읽고 쓸 수 있게 해 준다. Permission은 유닉스 파일시스템과 유사함. 데이터 객체와 컬렉션의 소유자가 iRODS 이용자나 그룹에게 읽기, 쓰기 등 권한을 지정할 수 있으며, Federation은 데이터 공유 및 출판을 하나의 Zone 영역을 넘어서 가능하게 해준다.

이용자들은 iRODS의 데이터와 메타데이터에 iRODS 클라이언트들을 통해 접근하고, 클라이언트들은 API를 통해서 통신을 하며, 일부 클라이언트들은 다수의 API들을 이용하게 된다.

2.2 Fedora Commons

2.2.1 개요 및 기능

Fedora 4의 경우 Resources, Containers, Binaries 3개의 구조를 가지고 있으며, 이는 Fedora 3에서의 Objects and Datastreams, Objects, Datastreams와 같은 개념이다[11].

Fedora는 파일 시스템을 포함하고 데이터베이스, 메타데이터, 원시 데이터를 위한 다양한 스토리지 옵션을 제공하며, 검색, 발견 등의 서비스를 제공하기 위해 다양한 응용프로그램과 통합될 수 있도록 설계되어 있다. 또한, Native Linked Data를 지원하고, 표준적인 웹 표준을 통해서 리파지터리의 핵심 서비스를 제공하며, Hydra, Islandora와 같은 대형 어플리케이션의 프레임워크의 핵심 구성요소로 사용된다.

플러그 형태의 role-based 인증모듈, XACML 혹은 웹 접근 제한으로 디지털 자산을 보호하고, 데이터의 무결성, 데이터와 관련된 다양한 로그, 데이터 버전 정보, 백

업 및 복구 기능을 제공하며, 수 백 만개의 파일과 메타 데이터 레코드를 처리할 수 있다.

Fedora는핵심 서비스를 RESTful APIs를 통해서 제공하고, OAI-PMH 배포를 지원하며, SWORD 프로토콜을 이용한 데이터 제출 또한 지원한다.

내장되어 있는 메시징 프레임워크를 통해서 메시지 기반의 워크플로우 기능을 지원하고, 확장 가능한 워크플로우를 제공한다. WAR 파일을 서블릿 컨테이너에 추가함으로써 손쉽게 운영환경 구축이 가능하고, 여러 프로세스들을 리파지터리의 하나의 이벤트로 묶음으로써 일관성과 성능향상을 가져올 수 있으며, Fedora에 의존적이지 않은 Open RDF 형식으로 데이터를 반출할 수도 있다.

2.2.2 저장 및 연계

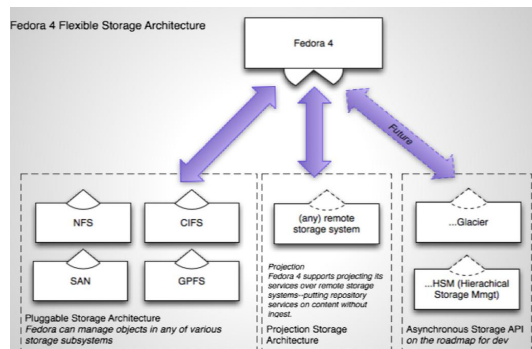


Fig. 3. Architecture on Fedora Commons Storage

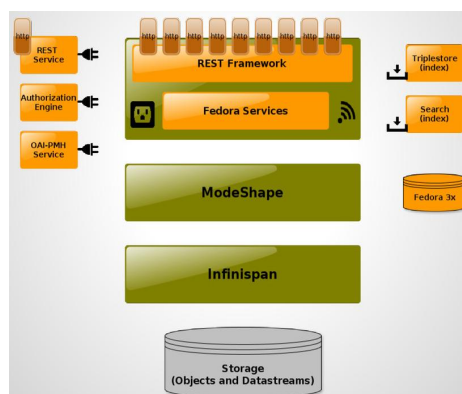


Fig. 4. Related Module on Fedora Commons

Fig. 3과 같이 플러그 형식의 저장소 아키텍처를 갖는다. NFS, CIFS, SAN, GPFS와 같은 다양한 저장소 서버 시스템에 담긴 객체들을 처리할 수도 있고, 원격의 저장

소 시스템에게 Fedora 4 서비스의 반영이 가능하다.

Fig. 4에서 보여지는 바와 같이 OAI Provider, SWORD Server와 같은 선택적, 플러그인 형식의 컴포넌트와 Audit Service, Triplestore, Indexing(Solr), Camel과 같은 외부 컴포넌트로 구성되어 있다.

2.3 B2Share

2.3.1 개요 및 기능

장기적 연구 데이터셋의 출판과 연구자 및 시민과학자의 작은 데이터의 장기간 사용이 가능하도록 데이터 생산 및 관리를 지원하기 위해 EUDAT가 제공하는 단순 스토리지 서비스이며, B2Share는 범 유럽 데이터 인프라 스트럭처로 사용되고 있다[12,13].

B2Share는 Invenio 디지털 도서관 프레임워크를 기반으로 하며, Invenio는 CERN 문서서버를 구동하기 위해 디지털 도서관 소프트웨어 솔루션으로 개발되었다. 2002년 이후로 고에너지 물리학 분야의 1백만 건 이상의 서지 레코드를 관리하고 있으며, 논문, 단행본, 저널, 사진, 비디오 등 다양한 콘텐츠 정보를 처리하고 있다[14].

또한, 서지레코드 관리, 자동화된 레코드 품질 체크, 수정 가능한 브라우징 기능을 포함한 대출 모듈, 데이터 셋 수집 기능, 도서관 상호대차 프로세스를 지원하고, 연구 결과물의 획득, 보존 및 데이터셋, 분석 코드, 가상 머신 환경, 환경 설정 및 지식정보를 수집한다. 브라우저 기반의 데이터 가상화 및 클라우드에 저장된 코드의 재실행 또한 지원한다.

2.3.2 리퍼지터리 및 데이터 모델

Fedora는 오디오, 사진, 비디오 자료 관리, 썸네일 생성과 포맷 추출, 포트폴리오 검색 결과 커스터마이징, 앨범 및 재생리스트 제작, 멀티미디어 콘텐츠 태깅 등의 다양한 형태의 멀티미디어 아카이브를 지원하고 있다.

또한, 기관의 논문, 리포트, 학위논문 출판, 컬렉션 구성 및 운영, 수집 프로세스 및 승인을 위한 환경설정, 영구적 식별자 생성, OAI-PMH를 통한 자원 배포와 같은 기관형 리퍼지터리 구동도 가능하다.

Fedora는 articles, books, photos, videos, data, and software를 기술하기 위해서 JSON 스키마를 사용한다던지, MARC21 with BibTeX, DataCite, Dublin Core, EndNote, RefWorks 메타데이터 형식을 지원하는 등 유연한 데이터 모델을 제공하고 있다. 다양한 과학연구 분야의 메타데이터 기술 또한 가능하다.

Invenio 3 메타데이터 저장소는 JSON 형태로 메타데이터를 관리한다. 따라서 어떠한 형식의 구조도 가능하며, dojson 라이브러리를 통해서 JSON 형식과 Marc 21 형식 변환 또한 가능하다. 이러한 Invenio 3 메타데이터의 검증은 Fig. 5와 같이 레코드에 포함된 \$schema 이름의 URL 값으로 지정된 JSON 스키마를 통해 레코드가 생성되거나 갱신될 때 형식을 검증하게 된다.

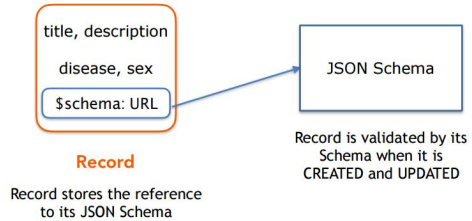


Fig. 5. Invenio 3 metadata validation

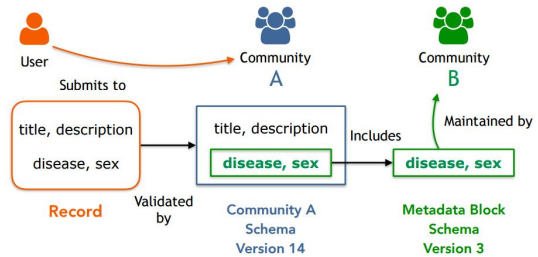


Fig. 6. Dynamic Flow on Invenio 3 metadata

Fig. 6과 같이 커뮤니티가 사용하는 one versioned schema는 공통 필드와 다른 커뮤니티에서 관리하는 ‘메타데이터 블록 스키마들로 구성되어 있으며, 메타데이터 저장 공간으로 PostgreSQL JSONB, Elasticsearch가 사용되고, MySQL도 여전히 사용 가능하다.

2.4 Dataverse

2.4.1 개요 및 기능

Dataverse는 연구데이터의 공유, 보존, 인용, 탐색, 분석하기 위한 오픈소스 웹 응용프로그램이다. 자동생성되는 데이터를 인용할 수 있고, 다 단계 출판 프로세스를 지원하며, 메타데이터 필드 기준의 패킷 검색이 가능하다. 또한, 검색 및 데이터 제출을 위한 API를 제공하고, Shibboleth를 통한 기관의 계정과 연동을 통해 Single Sing On을 지원한다.

description/citation, domain-specific 또는 custom fields, file metadata의 세 단계 메타데이터 수준을 지원하고, 미리 정의된 이용자의 역할에 따른 접근 제어를 지원하고, 커스터마이징이 가능하도록 설계되었으며, TwoRavens 연동으로 테이블 형식의 데이터 분석 및 재포매팅이 가능하고, WorldMap 연동으로 지리공간 파일 매핑을 지원하다[15,16].

2.4.2 아키텍처와 구성요소

Dataverse는 자바 엔터프라이즈 에디션으로써 웹 아카이브 파일인 war 파일로 배포되는 웹 응용 프로그램으로써, 리눅스는 RHEL/CentOS 사용을 권장하고 있으며, 동시에 war파일이 설치 될 Java EE 어플리케이션 서버로 Glassfish 사용을 권장하고 있다.

관계형 데이터베이스로 PostgreSQL 사용을 권장하며, 패스워드 재설정 및 공지기능을 사용하기 위해서 SMTP 메일서버가 필요하다. 또한, DOI 출판지원을 위해 EZID 라이선스(subscription) 혹은 DataCite 계정 필요하며, Dropbox와의 연계가 가능하다. Shibboleth 인증시스템을 사용할 수 있으며, 이 경우에는 아파치 서버가 별도로 필요하다.

시스템적으로는 2.8 GHz processor 두 개, 8 GB 메모리, 50 GB 디스크면 가상 머신 하나를 운영할 수 있고, 하나의 로드밸런스, 여러대의 프락시서버, 웹서버, 데이터베이스 서버 구동이 가능하다.

3. 제안된 국가 공유·활용 플랫폼

3.1 플랫폼 계층 및 구조

국가 과학데이터 플랫폼은 분야별 과학데이터 연구기관에서 생산한 과학데이터들을 등록, 저장, 관리 및 공유·활용할 수 있는 환경을 제공하는 국가 과학데이터 센터 구축을 위한 기반 기술들을 제공해야 한다.

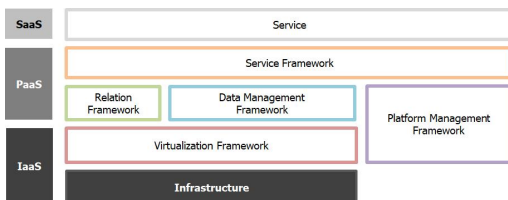


Fig. 7. The Structure on Sharing Platform of Nation Scientific Data

이를 위해, 국가 과학데이터 플랫폼은 과학데이터 서비스, 플랫폼 및 인프라스트럭처의 3계층으로 구성하였고, 플랫폼 계층은 제공하는 기능의 공통성에 따라 연계, 데이터 관리, 플랫폼 관리 및 가상화 등 Fig. 7에서와 같이 4개의 프레임워크로 구성하였다.

우선, 플랫폼 계층은 인프라스트럭처를 기반으로 과학데이터 서비스를 제공하기 위해 연계 프레임워크를 포함하여 SW공통 기반 프레임워크들을 제공한다. 플랫폼 계층은 아래와 같은 5개의 프레임워크로 구성(연계 프레임워크, 서비스 프레임워크, 데이터 프레임워크, 플랫폼 관리 프레임워크, 가상화 프레임워크)되며, 각각의 프레임워크는 독립적으로 정의되는 기능성에 따른 개발 및 실행 환경을 지원한다[17,18].

특히, 연계 프레임워크는 플랫폼 계층의 하부 계층부터 상부 계층의 기능들을 제공하기 위한 외부 기관 또는 시스템들과의 연동 환경 제공하고 관리하는 기능을 지원한다. 연계 프레임워크에서 제공되는 상호연동 인터페이스들은 연동 대상이 되는 시스템 환경에 따라 정의하여 제공한다.

3.2 프레임워크별 플랫폼 기능

Fig. 8과 같이 연계 프레임워크, 데이터 관리 프레임워크, 플랫폼 관리 프레임워크, 서비스 프레임워크, 가상화 프레임워크 등 5개의 프레임워크별 기능 명세는 다음과 같다.

첫째, 연계 프레임워크는 2가지 측면을 고려하여야 한다. 우선 자체 비관리 기관의 경우, 국가 과학데이터 공유·활용 플랫폼은 클라우드를 기반으로 하여 설계하고 있기 때문에 클라우드 관계에 따라 Inter/Intra 클라우드 연동 규격을 기초로 하여 연계하여야 한다. 자체 관리기관의 경우, 자체 구축된 Legacy 시스템을 활용하여 과학데이터를 관리하는 기관과의 연계는 OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting) 규격을 준용하여 연계하여야 한다.

둘째, 데이터 관리 프레임워크는 3가지 측면에서 고찰할 필요가 있다. 우선 메타데이터의 경우, 서비스 프레임워크의 접근 인터페이스를 통해 호출되는 메타데이터 표준 서비스 컴포넌트들로 구성되고, 관리 서브계층에 저장된 관련 메타데이터를 저장, 조회하여 사용자의 서비스 호출에 대한 결과를 제공한다. 각종 메타데이터 저장 및 이에 접근하기 위한 접근 인터페이스를 제공하는 컴포넌트로 구성되고, 각종 메타데이터 저장을 위한 DB가 제공되며, 이에 접근하여 저장, 조회 및 변경을 수행할 수 있

는 인터페이스를 제공하여야 한다. 또한, 등록된 메타데이터를 국가 과학데이터 플랫폼 기반의 원격 시스템과 공유하기 위한 메시지 서비스 기능을 제공할 필요가 있다.

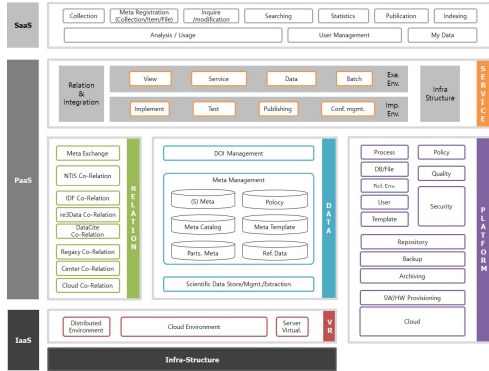


Fig. 8. The Functional Structure of Platform

과학데이터의 경우, 등록하는 과학데이터의 파일 내용을 확인할 수 있도록 일부 파일의 내용을 추출하여 메타데이터에 등록하는 기능을 제공하고, 과학데이터를 저장하기 위한 영구 식별자 생성, 발급을 위한 컴포넌트, 과학데이터에 대한 영구 식별자로 DOI를 발급한 이후, IDF에 발급된 DOI를 등록하는 절차까지 수행하여야 한다. I/O 인터페이스의 경우, 분산 환경 및 스토리지 환경에 파일을 저장하고 조회하기 위한 표준 I/O API를 제공할 필요가 있다.

셋째, 플랫폼 관리 프레임워크는 4가지 측면에서 바랄 필요가 있다. 메타데이터 관리의 경우, 과학/메타데이터를 저장·관리하는 데이터 프레임워크를 지원하기 위한 공통 기반 기능들로 구성되고, 표준 메타데이터 환경을 제공하며, 표준 메타데이터 관리 요소 및 템플릿을 제공 관리하는 기능을 제공하여야 한다. 서비스 관리의 경우, 과학데이터 서비스 개발환경, 실행환경, 운영 및 관리환경을 지원하기 위한 공통 기능들로 구성되며, 국가 과학데이터 서비스 운영자와 과학데이터 관리기관 관리자에게 실행되는 과학데이터 서비스 상태 및 서비스 자원의 운영 현황에 대한 관리 정보를 제공할 필요가 있다.

포털 운영의 경우, 과학데이터 서비스 운영자에게는 실행되는 과학데이터 서비스에 대한 서비스 상태 모니터링 및 운영 정보를 제공하며, 국가 과학데이터 관리기관 관리자에게는 클라우드 센터의 인프라 자원부터 서비스 단위로 실행되는 상태에 이르는 모든 관리 정보를 제공 받는다. 또한, 서비스 관리, 메타데이터 관리 및 가상화

관리 컴포넌트에서 제공하는 개별 기능 컴포넌트들을 호출하여 서비스 운영자와 센터 관리자에 필요한 기능을 제공하여야 한다. 가상화 관리의 경우, 국가 과학데이터 센터의 물리적, 논리적 인프라스트럭처와 클라우드 서비스 및 SW 관리 기능을 제공할 필요가 있다.

넷째, 서비스 프레임워크는 전자정부 표준 프레임워크 구조를 수용하여 구성되며, 기 정의된 전자정부 표준 프레임워크의 세부 기능을 준용 또는 확장 적용하여 클라우드 기반 과학데이터 서비스를 제공한다.

다섯째, 가상화 프레임워크는 3가지 측면에서 살펴봐야 한다. 우선, IaaS의 경우, 컴퓨팅, 스토리지, 네트워크 등 다양한 리소스*를 표준화된 인터페이스를 통해 제공되어야 한다. PaaS의 경우, 과학데이터 어플리케이션 개발자와 관리자는 과학데이터 플랫폼 클라우드에 어플리케이션을 배포하고 관리하며, PaaS 플랫폼에 분산저장, 과학데이터 및 메타데이터 검색 기능을 포함하여야 한다. 끝으로, SaaS의 경우, 국가 과학데이터 플랫폼 서비스 이용자는 접근하는 과학데이터 플랫폼 및 이용자별 서비스 권한 현황에 따라 서로 다른 접근 가능한 UI 서비스를 제공받아야 하며, 이를 지원 하는 멀티테넌트 기능을 제공할 필요가 있다.

4. 결론 및 제언

본 연구에서는 데이터 리파지터리로 글로벌하게 사용되고 있는 iRODS, Fedra Commons, Invenio 기반의 B2SHARE, Dataverse 공유·활용 플랫폼을 검토하였다. 각 소프트웨어에서 제공되는 이용자 인터페이스 기능은 큰 차이가 없다. 각 소프트웨어의 특징적 기능을 살펴보면 다음과 같다.

iRODS의 Rules은 iRODS가 특정 시스템 행위를 했을 때 어떤 행위들이 실행되어야 하는지(be triggered) 명시해 놓은 것으로서, 규칙 기반으로 시스템적 행위를 제어할 수 있다. Fedora Commons는 수 백 만개의 파일과 메타데이터 레코드를 처리할 수 있어 Hydra, Islandora와 같은 대형 어플리케이션의 프레임워크의 핵심 구성요소로 사용되고 있으며, Native Linked Data를 지원한다.

B2Share는 브라우저 기반의 데이터 가상화 및 클라우드에 저장된 코드의 재실행을 지원한다. JSON 스키마를 사용하고 있어 유연한 데이터 모델 설계가 가능하다.

Dataverse는 EZID 라이선스(subscription) 혹은 DataCite 계정을 이용해 DOI 출판을 지원하고 있으며,

Shibboleth를 통한 Single Sign On 기능을 지원한다. 또한 TwoRavens 연동으로 테이블 형식의 데이터 분석 및 재 포맷팅을 지원하고 있다.

본 연구에서 분석한 플랫폼은 모두 오픈 소스로서 데이터 리포지토리를 구축하고자 하는 기관이 손쉽게 획득하여 설치 가능하다. 구축하고자 하는 시스템의 목적과 최종적인 서비스 방향을 고려하여 소프트웨어를 선택하고 커스터마이징을 통해 데이터 리포지토리를 구축할 수 있을 것이다.

향후, 디지털 아카이빙 OAIS 참조모델에서 요구되는 기능적 요구사항들을 각 소프트웨어가 어떤 수준으로 구현하고 있는지? 어떤 방식으로 구현하고 있는지? 를 분석, 제시함으로써 기관의 환경에 적합한 데이터 리포지토리를 선정하는데 참고 자료로 활용될 수 있도록 보충이 필요하다.

REFERENCES

[1] S. H. Lee (2018). A Distributed Method for Efficient Analysis of Large-scale Scientific Data : Focusing on the Ocean Color Satellite Data, *University of Seoul*

[2] J. M. Yoon. (2016) Policy direction for efficient management and utilization of scientific data, *The Proceeding of Korea Technology Innovation Society*, 521-531

[3] S. H. Lee (2017). *Research and Development on Governance System, Core Technologies and Applications for Scientific Data* : KISTI

[4] S. T. Kim. (2010). A Study on a Model for Using and Preserving Scientific Data. *Journal of the Korean Biblia Society for Library and Information Science*, 21(4), 81-93

[5] S. U. Park. (2011) An Economic Ripple Effect Analysis of National Scientific Data Center Construction, *Journal of information science theory and practice*, 42(3), 55-69

[6] S. H. Lee (2013). *A Development on the Scientific Data Sharing and Utilizing System* : KISTI

[7] S. H. Lee (2011). *A Study on Establishing the Scientific Data Management and Application Framework to Build Up the National Scientific Data Center* : KISTI

[8] Research Data Management Strategy Requirements. <https://admire.jiscinvolve.org/wp/files/2013/05/ADMI-Re-RDM-Repository-Strategy-Requirements.pdf>

[9] *iRODS Technical Overview* <http://bit.ly/2dd7Yoa>

[10] *iRODS Architecture* <http://bit.ly/2d02B8u>

[11] *Introduction to Fedora 4*
<http://fedorarepository.org/presentations>

[12] *B2Share* <http://bit.ly/2dHdTQX>

[13] *B2Share on Invenio 3* <http://bit.ly/2e9GDVE>

[14] *Invenio Digital Library Framework*
<http://invenio-software.org/>

[15] *Dataverse* <http://dataverse.org/about>

[16] *Dataverse Features*
<http://dataverse.org/software-features>

[17] S. K. Kim et al.' (2016). A Study on the Development of Phased Big Data Distribution Model Based on Big Data Distribution Ecology, *Journal of Digital Convergence*, 95-106

[18] H. S. Byeon. (2017). The Status and Suggestions for Big Data Adaptation in the Government and the Public Agency. *Journal of Digital Convergence*, 13-25.

진 영 근(Jin Young Goun)

[정회원]



- 1986년 2월 : 한국과학기술원 전 전자과 석사
- 1998년 8월 : 충남대학교 컴퓨터 공학과 박사
- 1989년 12월 ~ 1999년 2월 : 한국 항공우주연구원 선임연구원
- 1999년 3월 ~ 현재 : 충남도립대학 컴퓨터정보과 교수
- 관심분야 : 오픈소스, 인공지능, 가상현실, 임베디드
- E-Mail : ygjinstar@gmail.com

이 원 구(Lee, Won Goo)

[정회원]



- 2002년 2월 : 한남대학교 컴퓨터 공학과(공학석사)
- 2005년 8월 : 한남대학교 컴퓨터 공학과(공학박사)
- 2005년 2월 ~ 2015년 3월 : KISTI 선임연구원
- 2015년 4월 ~ 현재 : 충남도립대학 컴퓨터정보과 교수
- 관심분야 : 빅데이터, 지식서비스
- E-Mail : dryope19@gmail.com