

Geoscience 분야 데이터 리포지터리 현황과 Coretrustseal 인증 획득 방안에 관한 연구

- re3data.org와 Coretrustseal 인증 모범사례를 중심으로 -

A Study on the Status of Data Repositories in the Field of Geoscience
and Ways to Obtain Coretrustseal Certification:
Focusing on re3data.org and Coretrustseal best practices

김 주 섭 (Juseop Kim)*

< 목 차 >

I. 서론	IV. Coretrustseal 인증 획득 방안
II. 이론적 배경	V. 논의 및 결론
III. 데이터 리포지터리 현황	

요약: 연구데이터 공유 및 보존을 위한 도구로 데이터 리포지터리가 활용되고 있다. 본 연구는 데이터 리포지터리 현황을 조사하고 Coretrustseal 인증 획득 방안을 제안하기 위하여 수행되었다. 연구 목적을 달성하기 위하여 re3data.org에서 Geoscience 분야의 데이터 리포지터리 현황에 대해 조사하였고 해당 리포지터리 중 Coretrustseal 인증을 획득한 10개의 모범사례를 분석하였다. 먼저, 리포지터리 현황을 파악한 결과, 국가별로 미국, 독일, 캐나다가 전체 리포지터리 중 66%를 차지하고 있으며 CTS 인증 현황 경우 미국, 유럽 그리고 독일의 비중이 58%로 나타났다. 다음으로 CTS 인증 확보를 위한 방안으로서 데이터 정책, 조직 인프라 그리고 기술 인프라 측면에서의 필요한 세부내용 및 요구사항을 제시하였다. 추후 본 연구 결과가 국내의 데이터 리포지터리 구축 및 운영 그리고 CTS 인증 획득에 도움이 되길 기대해본다.

주제어: 데이터 리포지터리, 리포지터리 인증, 데이터 공유 및 재사용, re3data, Coretrustseal

ABSTRACT: A data repository is being used as a tool for sharing and preserving research data. This study was conducted to investigate the current status of data repositories and to propose methods for obtaining Coretrustseal certification. In order to achieve the purpose of the study, the current status of data repositories in the field of Geoscience was investigated at re3data.org, and 10 best practices that obtained Coretrustseal certification were analyzed among the repositories. First, as a result of identifying the current status of repositories, the United States, Germany, and Canada accounted for 66% of the total repositories by country, and in the case of CTS certification, the proportion of the United States, Europe, and Germany was 58%. Next, as a way to secure CTS certification, necessary details and requirements in terms of data policy, organizational infrastructure, and technical infrastructure were presented. It is hoped that the results of this study will be helpful in establishing and operating domestic data repositories and obtaining CTS certification.

KEYWORDS: Data Repository, Repository Authentication, Data Sharing and Reuse, re3data, Coretrustseal

* 전북대학교 문헌정보학과 강사, 연구데이터융복합연구소 전임연구원
(kimjuseop@jbnu.ac.kr / ISNI 0000 0004 7492 1806)

• 논문접수: 2023년 5월 27일 • 최초심사: 2023년 5월 28일 • 게재확정: 2023년 6월 17일
• 한국도서관·정보학회지, 54(2), 89-110, 2023. <http://dx.doi.org/10.16981/kliss.54.2.202306.89>

※ Copyright © 2023 Korean Library and Information Science Society
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

I. 서론

1. 연구 목적 및 필요성

데이터 기반 연구가 중요해짐에 따라 데이터 공유, 데이터 보존 그리고 데이터 재사용 등과 같은 데이터 관리에 관한 요구사항이 다양하게 도출되고 있다. 이러한 데이터 관리를 위한 도구로서 데이터 리포지터리가 다양한 학문 및 커뮤니티 그리고 연구기관 등에서 활용되고 있다. 일반적으로 데이터 리포지터리는 데이터의 기탁, 공유, 관리, 보존 그리고 재사용을 위한 플랫폼으로 개발된다. 이러한 데이터 리포지터리는 다양한 주제, 커뮤니티 그리고 국가를 중심으로 데이터 리포지터리가 구축 및 운영되고 있으며 이러한 리포지터리 현황을 일반적으로 확인할 수 있는 사이트가 바로 re3data.org이다. 기존의 오픈 액세스 리포지터리의 글로벌 디렉토리인 OpenDOAR이 존재하지만 re3data.org는 이와는 다르게 연구데이터에 초점이 맞추어져 있다.

re3data.org는 데이터 리포지터리 레지스트리를 만들기 위해 2012년부터 2015년까지 독일 연구재단(DFG)이 자금을 지원한 연구 프로젝트로 출발하였다. 이러한 re3data.org는 데이터 리포지터리, 포털 및 기타 서비스 제공업체와 같은 연구데이터의 관리, 저장, 접근 및 이용을 지원하는 서비스를 찾고 평가하기 위한 연구자, 편저, 출판사 및 기타 이해관계자 등 다양한 대상을 지원하고 있다. 2023년 5월 현재, re3data.org에 3,116개가 등록되어 있으며, 해당 사이트에서는 리포지터리의 품질 정보를 나타내는 인증 여부를 확인할 수 있다(Kindling et al., 2017).

데이터 공유 및 재사용에 있어 핵심은 데이터 품질이다. 특히 데이터 공유 및 재사용은 융합연구를 가능하게 하며 연구의 재현성을 확보할 수 있는 수단이다. 결국은 좋은 데이터 품질은 융합연구와 재현성 확보하는 점에 있어 근거가 될 수 있다. 하지만 모든 데이터의 품질을 관리하고 검증하기에는 인력, 여건 및 예산 등에 한계가 있을 수밖에 없다. 따라서 데이터를 담는 그릇 즉, 리포지터리의 품질을 논의하는 게 현실적으로 적절하다고 할 수 있다. 본 연구는 re3data.org에 등록된 데이터 리포지터리의 현황을 분석하고 그 중 CTS 인증을 획득한 Geoscience 분야의 모범사례를 분석하기 위해 설계되었다. 연구의 대상으로 Geoscience 분야를 선정한 이유는 해당 분야가 데이터를 관리하기 위한 노력이 타 학문보다 일찍 시작되었으며 학문 특성상 과거 데이터의 보존 및 활용이 중요하기 때문이다(한중규, 2019). 본 연구의 목적을 달성하기 위해 수집한 모범사례는 CTS 인증 획득 방안을 도출하기 위한 근거자료로서 사용되었으며 해당 내용을 중심으로 인증 신청서를 작성하기 위한 3가지 차원의 세부 내용을 도출하고자 한다.

2. 연구 방법

본 연구의 목적은 연구데이터 리포지터리의 현황을 분석하고 데이터 리포지터리 체계 중의

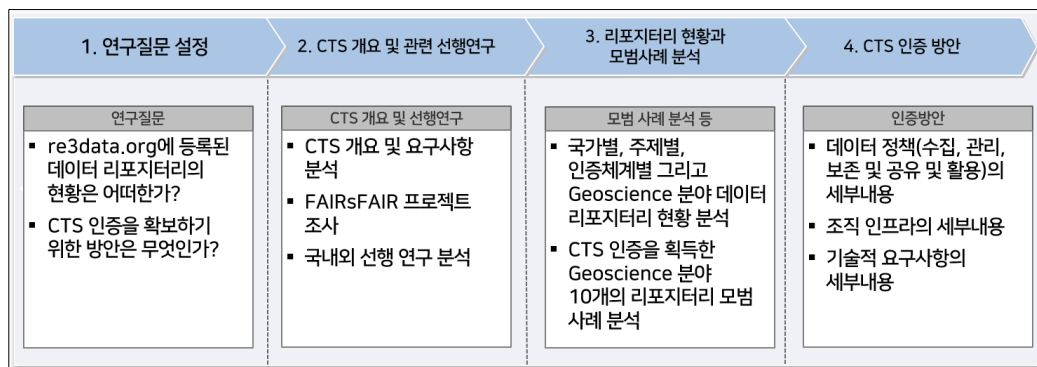
하나인 Coretrustseal(이하 CTS)의 인증을 획득하기 위한 방안을 제시하기 위함이다.

연구 목적을 달성하기 위해 제시한 연구 질문은 다음과 같다.

- 1) re3data.org에 등록된 데이터 리포지터리의 현황은 어떠한가?
 - 국가별 데이터 리포지터리의 현황은?
 - 분야별 데이터 리포지터리의 현황은?
 - CTS 인증을 획득한 리포지터리의 현황은?

- 2) CTS 인증을 확보하기 위한 방안은 무엇인가?
 - CTS 인증 확보를 위한 데이터 정책의 세부 내용은?
 - CTS 인증 확보를 위한 조직 인프라의 세부 내용은?
 - CTS 인증 확보를 위한 기술적 요구사항은?

연구 질문의 답을 제시하기 위하여 연구데이터 리포지터리 레지스트리인 re3data.org를 분석하였다. 분석의 세부내용은 국가별, 분야별 그리고 인증을 받은 리포지터리 현황을 포함하고 있다. 다음으로 국내 데이터 리포지터리가 CTS 인증을 부여받기 위한 방안은 무엇인지를 제시하고자 Geoscience 분야의 모범사례 10개를 선택하여 분석의 대상으로 선정하였다. 모범사례는 CTS 인증을 획득한 해당 리포지터리의 인증 신청서를 말하며 해당 신청서는 CTS 홈페이지에서 수집하였다. 다음의 <그림 1>은 단계별 연구내용을 나타낸 것이다.



<그림 1> 연구 프로세스

본 연구는 크게 4단계로 구성하였다. 먼저, 1단계에서는 본 연구의 목적을 달성하기 위한 연구 질문을 설정하였다. 2단계에서는 CTS의 개요와 해당 인증을 획득하기 위한 요구사항 그리고

FAIRsFAIR와 같은 CTS 인증을 지원하기 위한 사례를 조사하였다. 또한, 국내외 선행연구를 분석하여 본 연구와의 차이점을 제시하였다. 3단계에서는 re3data.org에서 국가별, 주제별, 인증체계별 그리고 본 연구의 핵심 대상인 Geoscience 분야의 데이터 리포지터리 현황을 조사하였으며, CTS 인증 방안을 제시하기 위하여 10개의 Geoscience 분야 CTS 인증 모범사례를 분석하였다. 마지막으로 CTS 인증 획득 방안으로서 데이터 정책, 조직 인프라 그리고 기술적 요구사항으로 나누어 각각의 세부 내용을 제시하였다.

본 연구의 제한점으로는 10개의 Geoscience 분야의 CTS 인증 모범사례를 분석한 내용이 일반화될 수 있는지 여부이다. 국내에는 아직 CTS 인증을 부여받은 곳이 없고 해외에 비해 리포지터리 인프라가 부족한 게 현실이다. 국내 여건에서는 조금은 시기상조이지만 CTS 인증과 같은 리포지터리의 품질을 미리 고심하고 구축 및 운영단계에서 적용한다면 데이터와 해당 리포지터리에 대한 신뢰도를 확보할 수 있다고 판단된다. 아울러, 본 연구결과로서 제시된 데이터 정책, 조직 인프라 그리고 기술 인프라의 세부 내용은 CTS 인증 준비 시 고려사항으로 활용될 수 있을 것이다.

II. 이론적 배경

이번 장에서는 연구데이터 리포지터리 인증 체계 중 하나인 CTS와 해당 인증 체계를 획득하기 위한 관련된 지원 사례로서 FAIRsFAIR에 대해서 기술하고자 한다. 마지막으로 관련된 선행연구를 통해 본 연구와의 차별성에 대해서 살펴보하고자 한다.

1. Coretrustseal 개요 및 인증 요구사항

CTS는 WDS(World Data System)와 DSA(Data Seal of Approval) 협력으로 출범한 데이터 리포지터리 인증 기관으로서 WDS-DSA 데이터 리포지터리의 요구사항을 토대로 데이터 리포지터리에 핵심 수준의 인증을 제공한다. 이러한 CTS는 국제적인 커뮤니티 기반의 비영리 단체인 네덜란드 법률(CTS 재단 정관 및 절차 규칙)에 따른 법인으로 운영되고 있으며 검토자 총회를 대표하는 12명의 선출직 위원으로 구성된 표준 및 인증 위원회의 통제를 받고 있다 (coretrustseal, 2023).

기존 DSA와 WDS 인증을 갱신하려면 현재는 CTS 인증으로 전환해야 한다. CTS의 요구사항은 총 17개로 구성되어 있으며 요구사항별 가중치는 동일하다. 다음의 <표 1>은 CTS 요구사항을 정리한 것이다.

〈표 1〉 CoreTrustSeal(2023-2025) 요구사항

CoreTrustSeal(2023-2025) 요구사항		
카테고리	요구사항	내용
배경 정보 (R0)	R0	•배경정보 - 리포지터리 전체적인 설명(리포지터리 유형 선택 등)
조직 및 인프라 (R1-R6)	R1	•임무 및 범위 - 리포지터리는 '디지털 객체'에 대한 접근을 제공하고 보존할 명확한 임무가 있음
	R2	•권한 관리 - 리포지터리는 적용 가능한 모든 권한을 유지하고 규정 준수를 모니터링함
	R3	•서비스의 영속성 - 리포지터리에는 데이터 및 메타데이터에 대한 지속적인 접근과 보존을 보장할 계획이 있음
	R4	•법률 및 윤리 - 리포지터리는 가능한 데이터 및 메타데이터가 법적 및 윤리적 규범을 준수하여 생성, 선별, 보존, 접근 및 사용되도록 보장됨
	R5	•거버넌스 및 정보자원 - 리포지터리에는 임무를 효과적으로 수행하기 위해 명확한 거버넌스 시스템을 통해 관리되는 적절한 예산과 충분한 수의 직원이 있음
	R6	•전문지식 및 지침 - 리포지터리는 기관 내외부에서 지속적인 전문 지식, 지침 및 피드백을 확보하기 위한 메커니즘을 채택함
디지털 객체 관리 (R7-R13)	R7	•출처 및 신뢰성 - 리포지터리는 디지털 객체의 진위를 보장하고 출처 정보를 제공함
	R8	•기타 및 평가 - 리포지터리는 이용자의 관련성과 이해성을 보장하기 위해 정의된 기준에 따라 데이터와 메타데이터를 허용함
	R9	•보존 계획 - 리포지터리는 장기 보존에 대한 책임을 지고 계획되고 문서화된 방식으로 보존 계획을 관리함
	R10	•품질 보장 - 리포지터리는 기술 품질(technical quality) 및 표준 준수를 다루고 최종 사용자가 품질 관련 평가를 수행할 수 있도록 충분한 정보를 제공함
	R11	•워크플로우 (Workflows) - 디지털 객체 관리는 기탁에서 접근까지 정의된 워크플로우에 따라 이루어짐
	R12	•검색 및 식별 - 리포지터리는 이용자가 디지털 객체를 검색하고 적절한 인용을 통해 지속적으로 참조할 수 있도록 함
	R13	•재사용 - 리포지터리를 사용하면 시간이 지남에 따라 디지털 객체를 재사용할 수 있으므로 이해 및 사용을 지원하는 데 적합한 정보를 사용할 수 있음
정보 기술 및 보안 (R14-R16)	R14	•스토리지 및 무결성 - 리포지터리는 문서화 된 프로세스를 적용하여 데이터 및 메타데이터 저장 및 무결성을 보장함
	R15	•기술 인프라 - 리포지터리는 잘 지원되는 운영 체제와 지정 커뮤니티에 제공하는 서비스에 적합한 기타 핵심 인프라 소프트웨어 및 하드웨어에서 관리됨
	R16	•보안 - 리포지터리는 시설과 해당 데이터, 메타데이터, 제품, 서비스 및 이용자를 보호함

CTS의 요구사항은 17개 항목으로 배경정보(1), 조직 및 인프라(6), 디지털 객체 관리(7) 그리고 정보 기술 및 보안(3)으로 구성되어 있다. 먼저, 배경정보는 리포지터리의 개요를 나타내고 조직 및 인프라에는 임무 및 범위, 권한 관리, 서비스의 영속성, 법률 및 윤리, 거버넌스 및 정보자원 그리고 전문지식 및 지침 등 6개의 요구사항으로 리포지터리를 운영하기 위한 인력, 예산 그리고 전문지식을 보유하고 있는지를 확인하고 있다. 다음으로 디지털 객체 관리에는 출처 및 신뢰성, 보존 계획, 품질 보장, 워크플로우, 검색 및 식별에 관한 요구사항으로 데이터 관리와 관련된 전반적인 내용을 담고 있다. 마지막으로 정보 기술 및 보안 카테고리에는 스토리지 및 무결성, 기술 인프라 및 보안과 관련한 요구사항으로 데이터의 저장 및 무결성 그리고 보안을 지원하기 위한 기술적 인프라 측면에 대한 내용을 담고 있다. CTS 인증은 3년마다 재갱신을 요구하고 있으며 일반적으로 리포지터리를 운영하는 커뮤니티 또는 기관에서 신청서를 작성한 후 제출되는 프로세스로 진행된다. 제출된 신청서를 전 세계 75명의 CTS 검토자 중 2명이 해당 신청서에 대한 최소 3회 이상 검토를 진행한 후 수정 및 보완 그리고 최종 인증 여부 등을 담당한다.

따라서, 인증 신청서의 기술된 내용이 CTS 인증 획득 여부에 핵심이라고 볼 수 있다. CTS의 요구사항은 정기적으로 3년마다 갱신되고 있다. 위에서 제시한 2023-2025 버전은 2022년 말에 발표한 요구사항으로 이전 버전과의 차이점은 다음의 <표 2>와 같다.

<표 2> CTS 요구사항 변경 전/후

변경 전 (2020-2022)			변경 후 (2023-2025)		
카테고리	요구사항	내용	카테고리	요구사항	내용
배경 정보	R0	•리포지터리 유형	배경 정보	R0	•리포지터리 유형
조직 및 인프라	R1	•임무 및 범위	조직 및 인프라	R1	•임무 및 범위
	R2	•라이선스		R2	•권한 관리
	R3	•접근의 영속성		R3	•서비스의 영속성
	R4	•기밀/윤리		R4	•법률 및 윤리
	R5	•조직 인프라		R5	•거버넌스 및 정보자원
	R6	•전문가 지침		R6	•전문지식 및 지침
디지털 객체 관리	R7	•데이터 무결성 및 신뢰성	디지털 객체 관리	R7	•출처 및 신뢰성
	R8	•평가		R8	•기탁 및 평가
	R9	•문서화된 보관절차		R9	•보존 계획
	R10	•보존 계획		R10	•품질 보증
	R11	•데이터 품질		R11	•워크플로우
	R12	•워크플로우		R12	•검색 및 식별
	R13	•데이터 검색 및 식별		R13	•재사용
기술	R14	•재사용	정보 기술 및 보안	R14	•스토리지 및 무결성
	R15	•기술 인프라		R15	•기술 인프라
	R16	•보안		R16	•보안

먼저, CTS 요구사항의 카테고리의 요구사항이 변경된 것을 확인할 수 있다. 최신 버전(2023-2025)에서는 ‘조직 및 인프라’ 카테고리의 경우 ‘R2 라이선스’가 ‘R2 권한 관리’로 ‘R3 접근의 영속성’은 ‘R3 서비스의 영속성’으로 ‘R4 기밀/윤리’는 ‘R4 법률 및 윤리’로 ‘R5 조직 인프라’는 ‘R5 거버넌스 및 정보자원’으로 ‘R6 전문가 지침’은 ‘R6 전문지식 및 지침’으로 요구사항의 세부 내용을 나타내는 용어가 변경되었다. 다음으로 ‘디지털 객체 관리’에서는 ‘R7 데이터 무결성 및 신뢰성’은 ‘R7 출처 및 신뢰성’으로 ‘R8 평가’는 ‘R8 기탁 및 평가’로 ‘R9 문서화된 보관절차’는 ‘R10 보존 계획’과 함께 ‘R9 보존 계획’으로 통합되었음을 알 수 있다. 이어서 이전 버전의 ‘R11 데이터 품질’은 ‘R10 품질 보증’으로 ‘R13 데이터 검색 및 식별’은 ‘R12 검색 및 식별’로 요구사항 변경이 이루어졌다. 마지막 카테고리인 ‘기술’은 카테고리명이 ‘정보기술 및 보안’으로 바뀌었고 기존의 ‘R14 재사용’이 ‘R13’ 동일한 요구사항으로 이동하였으며 해당 ‘R14’는 ‘스토리지 및 무결성’이라는 요구사항으로 정보기술 및 보안 카테고리에 추가되었다. 따라서 최신 버전의 CTS 요구사항은 배경정보(1), 조직 및 인프라(6), 디지털 객체 관리(7) 그리고 정보기술 및 보안(3)으로 구성되었다.

2. FAIRsFAIR

FAIRsFAIR는 FAIR 원칙을 기반으로 데이터 관리, 절차, 표준 등 관련 문제에 대한 전반적인 인프라의 개발 및 구체적인 실현을 다루기 위한 프로젝트이다. 해당 프로젝트는 시맨틱 상호운용성을 지속 가능한 방식으로 지원하는 기술과 FAIR 원칙 준수를 지원하는 모범사례에 대한 권장사항을 도출하기 위하여 적극적으로 노력하고 있다. 이 프로젝트는 다음의 <표 3>과 같은 목적과 기능을 가지고 있다.

<표 3> FAIRsFAIR 목적 및 기능

목적	기능
<ul style="list-style-type: none"> • FAIR 메타데이터 스키마, 어휘, 프로토콜 및 온톨로지를 지정하여 연구 리소스의 시맨틱 상호운용성을 개선함 • FAIR 조정(FAIR-aligned) 리포지토리에 대한 상호운용성 요구사항 및 시스템 액세스 가능성에 대한 솔루션 제공 • FAIR 지원 서비스에 대한 가이드라인 및 권장사항 공식화(Formulate) • FAIR 원칙이 연구 소프트웨어에 어느 정도 적용될 수 있는지에 대한 평가 	<ul style="list-style-type: none"> • FAIR 원칙 사용을 위한 솔루션 제공 • FAIR 원칙을 준수하는 데이터 모범사례 채택 • FAIR 인증 및 EOSC(European Open Science Cloud), 이하 EOSC) 기능 인프라 전환 • FAIR 원칙 사용 및 구현을 위한 플랫폼 제공 • EOSC 참여를 위한 필수 FAIR 원칙 제공 및 준수 • 인프라 구성 요소 작동 검증

FAIRsFAIR에서는 이러한 목적을 달성하고 제시된 기능을 구현하기 위해 ‘정책 현황 파악 및 개선 권장사항’, ‘FAIR 데이터 원칙을 지원하는 모범사례에 대한 권장사항’ 그리고 ‘데이터 리포지터리를 위한 지원 프로그램’을 제공하고 있다. 이 중 세 번째 서비스인 ‘데이터 리포지터리를 위한 지원 프로그램’에서는 ‘CTS 인증 획득을 위한 데이터 리포지터리 지원’ 그리고 ‘리포지터리의 FAIRification¹⁾ 지원 및 데이터의 FAIR 요소 충족 및 상호운용성 향상’이라는 2가지의 목표를 가지고 시작되었다. 실제 해당 서비스를 통해 2019년 7월과 8월 사이에 오픈 콜이 진행되어 2그룹이 선정되었다. 먼저, 73개의 애플리케이션 중에서 10개의 리포지터리가 CTS 인증 지원에 선정되었고 12개의 리포지터리가 상호운용성 개선 지원에 포함되었다. 다음의 <표 4>는 해당 프로그램의 내용과 선택된 리포지터리를 기술한 것이다(FAIRsFAIR, 2023).

1단계에서는 해당 리포지터리가 FAIR 데이터 원칙과 CTS 요구사항을 준수하고 있는지 함께 리뷰를 진행한다. 2단계에서는 CTS 요구사항 체크리스트를 바탕으로 해당 리포지터리가 신청서 초안을 작성한다. 3단계에서 작성된 신청서 초안을 바탕으로 각 2인으로 구성된 5개의 팀이 1차 검토, 동료 평가로 2차 검토를 수행한다. 4단계에서는 검토를 통해 수정된 신청서를 CTS에 제출

1) FAIRification 프로세스는 FAIR 데이터 접근, 데이터 분석, 시맨틱 모델 정의, 데이터 연결, 라이선스 할당, 데이터셋에 대한 메타데이터 정의 그리고 FAIR 데이터 리소스 배포 등을 포함한다.

하고, CTS 위원회에서 공식 신청, 검토, 승인 절차를 수행한다.

〈표 4〉 FAIRsFAIR 데이터 리포지터리 CTS 인증 프로그램 개요

구분	소요기간	내용	지원 리포지터리명
1단계	2019.09.-2020.04.	FAIR 데이터 원칙과 CTS 리포지터리 요구사항 간 매핑 진행	• Apollo(영국) • DaSCH(스위스)
2단계	2020.05.-2020.09.	CTS 신청서 초안 작성	• DASS-BiH(보스니아-헤르체코비나) • DASSH(영국)
3단계	2020.10.-2021.02.	신청서 초안 검토 및 동료 평가	• ESRF Data Repository(프랑스) • IAGOS Data Center(프랑스) • ICOS Data Portal(스웨덴)
4단계	2021.02.-2021.05	신청서 제출 및 인증 프로세스 진행	• The Movebank Data Repository(독일) • SOCIB(스페인) • Tárki Data Archive(헝가리)

3. 선행연구

다음의 〈표 5〉는 re3data 및 CTS 인증과 관련된 국내외 연구를 연도순으로 정리한 것으로 연구의 핵심내용을 포함하고 있다.

〈표 5〉 국내외 선행연구

저자(연도)	논문명	내용
Khan et al. (2023)	연구데이터 리포지터리의 글로벌 개요: re3data 레지스트리 분석	• 언어, 소프트웨어, 메타데이터 표준 및 콘텐츠 범위 그리고 리포지터리 유형 등으로 분석
Khan, Thelwall, & Kousha (2022)	데이터 리포지터리가 제한되어 있는가? 현재 관행, 과제 및 미래 기술에 대한 조사	• 데이터 리포지터리 관리자를 대상으로 온라인 설문 조사 실시
Frank (2022)	CoreTrustSeal 데이터 리포지터리 인증 프로세스의 위험	• '리포지터리 잠재적 위험 Risk'에 대한 인식 차이에 대해 논의
Peng, Gross, & Edmunds (2022)	신뢰할 수 있는 FAIR 데이터 및 리포지터리를 촉진하는 관리 성숙도 평가 접근 방식 간 매핑	• CTS 요구사항, DSMM 및 FAIR 원칙을 비교 분석
Downs (2021)	오픈 데이터의 새로운 가치를 위한 기회 개선: 연구 데이터 리포지터리 평가 및 인증	• 리포지터리 인증 제도의 활용 및 기여 그리고 발전 방향에 대한 논의
Donaldson (2020)	신뢰할 수 있는 디지털 리포지터리 웹 사이트의 인증 정보: 내용 분석	• 91개의 TDR 인증 리포지터리 웹사이트의 콘텐츠 분석
Misgar, Bhat, & Wani (2020)	브릭스 국가들이 개발한 오픈엑세스 연구데이터 리포지터리에 관한 연구	• BRICS 국가의 리포지터리 강점, 내용 유형, 언어 인터페이스의 다양성 그리고 주제에 대한 조사
Cho (2019)	re3data를 기반으로 한 아시아 RDR 연구	• re3data에 등록된 아시아 데이터 리포지터리 152개의 전반적인 운영수준 및 국가간 격차 분석
Lindlar & Rudnik (2019)	CoreTrustSeal에 주목하다: 디지털 보존 및 연구자료 관리 관점에서 본 R0 기준에 대한 권고사항	• CTS R0 기준에 대한 권장사항 도출
Pisharoti & Petters (2019)	CoreTrustSeal 리포지터리 인증 요구 사항과 관련된 VTechData 평가	• VTechData의 자체 감사 보고서로, CoreTrustSeal 리포지터리 인증 획득의 일환으로서 진행됨

Khan et al.(2023)은 RDR(Research Data Repositories)와 RDR의 특징에 대한 전체적인 개요를 제공하였으며 해당 연구를 통해 re3data 레지스트리에 기여한 상위 15개국 중 유럽 10개국, 아시아 3개국, 북미와 오세아니아가 1개국이 차지하고 있음을 가장 많이 사용되는 언어는 영어, 소프트웨어는 DataVerse와 Dspace가 메타데이터 표준은 Dublin Core와 DataCite가 데이터 유형은 과학 및 통계 데이터 유형이 차지하고 있음을 기술하였다.

Khan, Thelwall, Kousha(2022)는 re3data.org에 등록된 연락처 정보를 수집하여 데이터 리포지터리 관리자를 대상으로 온라인 설문조사를 실시하였다. 그 결과, FAIR 데이터, 연구에 대한 이용자 지원 제공, 모범사례 개발이 우선순위로 도출되었으며, 향후 리포지터리 시스템에 대한 주요 권장사항으로 데이터와 시스템 간의 통합 및 상호운용성, 더 나은 연구데이터 관리도구, 데이터셋을 다운로드하지 않고 계산하는 도구, 자동화 시스템이 제시되었다.

Frank(2022)는 CTS 인증에 검토자로 참여했던 리포지터리 직원과 검토자로 참여한 경험이 없는 리포지터리 직원 간 '리포지터리 잠재적 위험'에 대한 인식 차이를 탐구하였다. 그 결과, CTS 인증에 검토자로 참여한 리포지터리 직원은 데이터 리포지터리의 잠재적 위험요소를 더 명확히 파악하고 있으며, 이는 검토자로서 공유한 경험과 교육 수료 등의 결과로 보이는 것으로 파악하였다.

Peng, Gross, Edmunds(2022)는 CTS 데이터 리포지터리 요구사항과 데이터 관리 성숙 매트릭스 그리고 FAIR 원칙을 비교 분석하였다. 그 결과, TDR의 주요 목표가 CTS 데이터 리포지터리 요구사항과 같은 인증 요구사항을 충족하고 데이터를 장기적으로 관리하기 위한 조직 거버넌스 및 관리 계획을 성문화하는 것임을 연구를 통해 도출하였다.

Downs(2021)는 오픈 데이터 및 데이터 재사용의 촉진을 위해 연구데이터 리포지터리의 활용이 필요하며 이러한 리포지터리의 평가 및 인증 제도가 데이터 기탁자의 신뢰를 얻기 위해 지속적으로 발전해오고 있음을 나타내고, 데이터 리포지터리의 개선과 발전을 위해서는 여러 요소가 필요함을 기술하였다.

Donaldson(2020)은 91개의 TDR 인증 리포지터리 웹사이트의 콘텐츠 분석 결과를 사용하여 TDR 상태, TDR 인증의 존재 및 위치, 추가 인증 정보에 대한 하이퍼링크 여부, 인증 프로세스 설명 그리고 감사 보고서의 공유 여부를 조사하였다. 조사 결과, 대부분의 데이터 리포지터리는 TDR 인증에 대한 정보를 제공하고 있는 것으로 파악되었다.

Misgar, Bhat, Wani(2020)는 re3data.org에 등록된 BRICS 국가의 리포지터리 강점, 리포지터리 내용 유형, 언어 인터페이스의 다양성 및 주제에 관하여 조사하였다. 브릭스 국가 중에서는 인도와 중국의 비중이 가장 높으며, BRICS 리포지터리의 비율은 전체 2,264개 저장소의 3.62%에 불과하였다. 또한, 대부분의 리포지터리는 과학 및 통계 데이터 콘텐츠를 제공하며, 가장 선호되는 언어는 영어이며 주제는 자연과학 분야가 가장 많은 비중을 차지하고 있음을 밝혀냈다.

Cho(2019)는 re3data.org에 등록된 아시아 데이터 리포지터리 152개의 전반적인 운영수준 및

국가간 격차를 분석하였다. 해당 분석은 콘텐츠 유형, 주제, 키워드, 데이터 라이선스, PID, 저장소 시스템, 메타데이터 및 API와 같은 RDR의 속성을 중심으로 진행하였으며, 그 결과 아시아권 리포지터리의 일반적인 유형은 라이선스 또는 정책은 개방되어 있으나 영구식별자가 부여되지 않고 인증되지 않은 상태로 운영되고 있음을 확인하였다.

Lindlar와 Rudnik(2019)은 CTS의 R0 기준에 대하여 데이터 보존 커뮤니티와 연구데이터 관리 커뮤니티의 관점에서 심층적으로 분석하고 권장사항을 도출하였다. 해당 연구 결과, 리포지터리 인증 관련 단어에 있어서 디지털 보존과 연구데이터 관리 간의 관점 차이를 확인할 수 있었다.

Pisharoti와 Petters(2019)는 버지니아 공대의 기관 리포지터리인 VTechData의 CTS 리포지터리 인증 획득을 위해 CTS 인증 요구사항을 검토하였으나 해당 요구사항을 충족하지 못했음을 결과로 제시하였다. 해당 연구는 CTS 인증 요구사항에 대한 기관 리포지터리의 관점을 확인할 수 있다는 의의가 있다.

이상의 10편의 해외 논문을 살펴본 결과, re3data.org에 등록된 4편의 데이터 리포지터리 관련 연구와 6편의 CTS 인증과 관련한 연구를 확인할 수 있었다. 데이터 리포지터리 현황 관련한 연구의 경우 언어, 소프트웨어, 메타데이터 표준 및 콘텐츠 범위 등을 나타내는 re3data의 속성을 중심으로 조사되었으며 CTS 인증 관련한 연구는 CTS 인증에 대한 인식 차이, FAIR 원칙 등과의 비교, 타 인증 획득을 받은 웹사이트 분석 그리고 요구사항에 대한 권장사항 등에 대하여 논의하였다. 본 연구의 핵심은 CTS 인증을 획득하기 위한 방안을 제시하기 위함이다. 특히, 타 연구에서 수행되지 않은 CTS 인증 모범사례 분석을 통해 해당 인증 확보를 위한 신청서 작성 가이드라인을 제시한다는 점에서 타 연구와의 차별성을 가지고 있다고 판단된다.

Ⅲ. 데이터 리포지터리 현황

이번 장에서는 re3data.org에 등록된 데이터 리포지터리 현황에 대해 조사 및 분석하고자 한다. 현황 조사는 주제별, 국가별, 전체 인증체계별 마지막으로 국가별 CTS 인증으로 나누어 수행하였다.

1. 주제별 분석

데이터 리포지터리 등록 레지스트리인 re3data.org에 등록된 리포지터리의 주제는 크게 '인문학 및 사회과학', '생명과학', '자연과학' 그리고 '공학' 4개로 구성되어 있다. 각 주제는 리포지터리 별 1개가 배정되는 것이 아니라 등록 시 주제 분야를 다양하게 지정할 수 있어 개수의 합이 전체의 총합과 다를 수밖에 없다. 다음의 <표 6>은 re3data.org에 등록된 데이터 리포지터리의 주제별

현황을 나타낸 것이다.

〈표 6〉 주제별 데이터 리포지터리 현황

대분류	중분류	소분류	대분류	중분류	소분류
인문학 및 사회 과학 (1,158)	인문학 (351)	고대문화(75)	자연 과학 (552)	화학 (52)	분자화학(4)
		역사(118)			화학적 고체 및 표면 연구(4)
		미술, 음악, 연극 및 미디어 연구(80)			물리 및 이론 화학(3)
		언어학(116)			분석화학, 방법개발(화학)(4)
		문학연구(33)			생화학 및 식품화학(4)
		비 유럽 언어 및 문화, 사회 및 문화 인류학, 유대인 연구 및 종교 연구(51)			폴리머 연구(3)
		신학(22)		물리학 (45)	응축 물질 물리학(1)
	철학(14)	광학, 양자 광학과 원자물리학, 분자 및 플 라즈마(1)			
	사회 및 행동 과학 (492)	교육학(78)			입자, 핵 및 필드(1)
		심리학(51)			통계물리, 연성물질, 생물물리, 비선형동 역학(1)
		사회과학(314)			천체 물리학 및 천문학(13)
		경제학(234)		수학(29)	
	법학(58)	지구 과학 (148)		대기과학 및 해양학(26)	
	생물학 (109)			기초 생물학 및 의학 연구(32)	지질학 및 고생물학(12)
식물학(23)			지구 물리학 및 측지학(20)		
동물학(25)			지구화학, 광물학 및 결정학(9)		
의학 (173)			미생물학, 바이러스학 및 면역학(21)	지리학(60)	
		의학(115)	수자원 연구(17)		
	신경과학(17)	공학 (512)	기계 및 산업공학(4)	기계공학 및 건설기계공학(2)	
농업, 임업, 원예 및 수의학(77)			열 공학/ 공정 공학(7)	공정공학, 공업화학(2)	
생명 과학 (638)	X			열에너지 기술, 열 기계, 유체역학(1)	
			재료 과학 및 공학(12)	재료공학(4)	
				재료과학(6)	
			컴퓨터 공학, 전기 및 시스템 공학 (96)	시스템공학(7)	
				전기공학(6)	
			건설 공학 및 건축(35)	컴퓨터과학(62)	

re3data.org에 등록된 데이터 리포지터리는 2023년 5월 기준으로 3,116개가 등록되어 있다. 주
제별 리포지터리로 확인한 결과, 인문 및 사회과학이 1,158개로 가장 많으며 다음으로 자연과학이
552개, 생명과학이 538개 그리고 공학이 512개로 나타났다. 대분류로 살펴보면, 인문학 및 사회과
학 중 사회 및 행동과학이 492개로 가장 많으며 생명과학에서는 173개로 의학이 자연과학에서는
148개로 지구과학이 마지막 공학에서는 컴퓨터 공학, 전기 및 시스템 공학이 96개로 가장 많은
규모로 확인되었다.

2. 국가별 분석

다음의 <표 7>은 re3data.org에 등록된 국가별 데이터 리포지터리 현황을 나타낸 것이다. 국가별 리포지터리는 1개 국가 당 1개의 데이터 리포지터리만을 등록할 수 있는 것이 아니며 국가 간 협력 리포지터리가 허용되어 중복된 개수가 포함되어 있다.²⁾

<표 7> 국가별 데이터 리포지터리 현황

No.	국가	개수	No.	국가	개수
1	미국	1,170	15	오스트리아	44
2	독일	498	16	벨기에	38
3	캐나다	393	17	노르웨이	31
4	영국	314	18	스웨덴	29
5	유럽 연합	287	19	덴마크	27
6	프랑스	126	20	러시아 연방	23
7	오스트레일리아	102	21	브라질	18
8	스위스	84	22	멕시코	18
9	중국	81	23	남아프리카 공화국	16
10	네덜란드	74	24	폴란드	16
11	일본	66	25	포르투갈	15
12	스페인	53	26	뉴질랜드	14
13	인도	51	27	핀란드	13
14	이탈리아	50	28	대한민국	13

국가별 데이터 리포지터리 현황을 살펴본 결과, 미국이 1,170개로 가장 많으며 뒤를 이어 독일, 캐나다, 영국, 유럽연합 순으로 나타났다. 미국과 캐나다를 중심으로 한 북미 지역과 독일, 영국, 유럽연합 그리고 프랑스 등 유럽 지역이 대다수를 차지하고 있는 것으로 나타났다. 우리나라는 현재, 13개가 등록되어 있으며 이마저도 현재 운영되고 있지 않은 리포지터리를 제외하면 11개로 확인되었다. 또한, 국가간 협력 리포지터리를 제외하면 순수하게 한국이 운영중인 데이터 리포지터리는 8개로 확인되었다. 8개 중 Geoscience 분야에 해당하는 리포지터리는 관할해역 해양정보 공동활용 시스템(Jurisdictional Ocean Information Sharing System), 국제 우주 환경 서비스(International Space Environment Service), 지오 빅데이터 오픈 플랫폼(Geo Big Data Open Platform) 그리고 한국 극지 데이터 센터(Korea Polar Data Center) 등 4개가 re3data.org에 등록되어 있다. 한중일을 리포지터리 등록 개수로 비교하면 중국 81개 및 일본 66개로 차이가 많음을 알 수 있다.

3. 데이터 리포지터리 인증 체계별 분석

본 절에서는 데이터 리포지터리 인증 체계별로 현황을 제시하고 연구의 핵심 대상인 Geoscience

2) 국가로 등록된 'International'의 경우 본 조사에서 제외하였다.

분야의 인증 현황과 그리고 해당 분야에서 CTS 인증이 얼마나 되는지 살펴보고자 한다. 다음의 <표 8>은 전체 데이터 리포지터리 중 인증 현황과 Geoscience 분야 데이터 리포지터리 중 인증 현황을 나타낸 것이다.

<표 8> 전체 인증 현황 및 Geoscience 분야 인증 현황

전체 인증 현황	현황	비고	Geoscience 분야 인증 현황	현황
Coretrustseal	139개	45.8%	Coretrustseal	53
WDS	43개	14.1%	WDS	3
RatSWD	41개	13.5%	DSA	2
CLARIN Certificate B	26개	8.5%	DIN 31644	1
DSA	15개	4.9%	Trusted Digital Repository	1
DINI Certificate	10개	3.3%	Toal	59
Trusted Digital Repository	2개	0.6%		
DIN 31644	1개	0.3%		
ISO 16363	1개	0.3%		
기타	25개	8.2%		
Total	303개			

전체 데이터 리포지터리 중 인증 현황을 살펴본 결과, CTS가 139개로 전체 약 46%로 차지하였으며 다음으로 WDS가 43개로 14.1%, RatSWD가 41개로 13.5% 그리고 CLARIN Certificate B가 26개로 8.5%로 나타났다. 특히, WDS와 DSA는 현재 CTS로 통합되었으며 WDS와 DSA는 재인증시 CTS 인증을 받아야 하므로 전체적으로는 CTS 인증이 197개로 산술적으로 65% 이상 차지하고 있음을 알 수 있다. 또한 Geoscience 분야 데이터 리포지터리 인증 현황을 살펴본 결과 등록된 868개 중 59개가 인증 받은 리포지터리로서 그 중 CTS가 53개로 약 90%를 차지하고 있으며 나머지 WDS와 DSA가 3개 그리고 2개로 그 뒤를 이었다.

다음의 <표 9>는 국가별 CTS 인증 현황을 나타낸 것이다.

<표 9> 국가별 CTS 인증 현황

No.	국가	개수	No.	국가	개수
1	미국	35	9	일본	3
2	유럽연합	26		러시아연방	3
3	독일	20		스페인	3
4	네덜란드	18		스위스	3
5	영국	16	16	체코 공화국	2
6	중국	6		덴마크	2
7	호주	5		핀란드	2
8	노르웨이	4		이탈리아	2
9	오스트리아	3		포르투갈	2
	프랑스	3		슬로바니아	2
	아일랜드	3			

국가별 CTS 인증 현황은 CTS 인증을 2개 이상 보유한 국가를 중심으로 표를 작성하였다. 전체 CTS 인증은 139개로서 미국이 35개로 가장 많고, 다음으로 유럽연합, 독일, 네덜란드 그리고 영국 순으로 각각 26개, 20개, 18개, 16개로 나타났다. 차이가 나지만 중국의 경우 5개 그리고 일본의 경우에도 3개의 CTS 인증을 받은 리포지터리를 보유하고 있지만 한국은 0개로 확인되었다.

데이터 리포지터리 현황을 파악한 결과, 국가별로 미국, 독일, 캐나다가 전체 리포지터리 중 66%를 차지하고 있으며 리포지터리 인증 체계 중 CTS가 차지하는 비중이 45%이며 이마저도 기존 인증을 대체하는 WDS와 DSA를 제외한 비율이다. 그만큼 CTS 인증 체계가 압도적으로 많고 영향력이 커지고 있는 것을 확인할 수 있다. 또한, CTS 인증 현황을 국가별로 확인해본 결과 미국, 유럽 그리고 독일의 비중이 58%로 나타났다. 이러한 현황을 비추어볼 때 아직 국내에서는 데이터 리포지터리의 등록 비중이 상대적으로 낮을 뿐만 아니라 인증받은 리포지터리도 전혀 없는 것으로 나타났다. 물론, 국내의 경우 KISTI의 DataON을 중심으로 국가 연구데이터 공유를 위한 적극적인 모습을 보이고 있으며 해당 연구기관이 개발한 데이터 리포지터리 소프트웨어인 NaRDA를 국내 각 기관에 무료로 배포 및 설치³⁾하고 있어 앞으로 국내 여건은 밝다고 할 수 있다. 이에 따라 전략적으로 국내 리포지터리를 re3data.org에 등록하여 홍보하는 것도 국내 기관의 위상에도 도움이 될 수 있을 것이다.

IV. Coretrustseal 인증 획득 방안

이번 장에서는 Geoscience 분야의 데이터 리포지터리가 CTS 인증을 확보하기 위한 방안을 제시하고자 한다. 방안을 제시하기에 앞서 먼저, Geoscience 분야 중 CTS 인증을 부여받은 데이터 리포지터리를 선택하고 해당 리포지터리의 CTS 신청서를 분석하였다.

1. 분석 개요

다음의 <표 10>은 본 연구의 핵심 대상인 Geoscience 분야 데이터 리포지터리 중 CTS 인증을 획득한 리포지터리의 개요를 나타낸 것이다.

CTS 최신 버전이 2023년 초에 공개된 것을 감안하여 10개의 리포지터리 중 '2020-2022' 버전이 8개 그리고 '2017-2019' 버전이 2개로 나타났다. 본 연구에서는 10개의 리포지터리 신청서⁴⁾를 CTS 홈페이지에서 수집하여 R0을 제외하고 R1~R16까지의 기술 내용을 모두 분석하였다.

3) 2022년 10월 현재, 국내 25개 기관이 데이터 리포지터리를 구축 및 운영하고 있으며, 이 중 22개 기관이 DataON 리포지터리 호스팅을 이용하거나 NaRDA를 도입하였다.

4) CTS 위원회 검토를 거친 최종 버전의 신청서

〈표 10〉 Geoscience 분야 CTS 인증 리포지터리

리포지터리	웹페이지	인증일	CTS 버전
Norwegian marine data centre (NMD)	• https://www.hi.no/en/hi/forskning/research-groups-1/the-norwegian-marine-data-centre-nmd	2022년 09월 30일	2020-2022
TOAR Database Infrastructure	• https://toar-data.fz-juelich.de/	2022년 05월 31일	2020-2022
ICTS SOCIB Data Repository	• https://socib.es/	2022년 03월 30일	2020-2022
Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC)	• https://daac.ornl.gov/	2022년 03월 30일	2020-2022
National geoscience data centre (NGDC)	• http://www.bgs.ac.uk/services/ngdc/	2022년 03월 04일	2020-2022
Centre for Environmental Data Analysis (CEDA)	• https://www.ceda.ac.uk/	2021년 06월 29일	2020-2022
International Council for the Exploration of the Sea (ICES)	• https://ecosystemdata.ices.dk/	2021년 03월 22일	2020-2022
UK Polar Data Centre (UK PDC)	• https://www.bas.ac.uk/data/uk-pdc/	2021년 03월 18일	2020-2022
PANGAEA	• https://www.pangaea.de/	2019년 06월 17일	2017-2019
Earth Resources Observation and Science Center (EROS)	• https://eros.usgs.gov/	2018년 11월 21일	2017-2019

2. CTS 인증 모범사례 분석

다음의 〈표 11〉은 Geoscience 분야 CTS 인증을 획득한 10개의 리포지터리 신청서 내용을 CTS 요구사항을 기준으로 상위요소 및 하위요소로 구분하여 나열한 것이다. 각 모범사례의 상세한 세부 내용은 본 연구에 모두 담을 수 없어 핵심 키워드 및 문장을 중심으로 제시하였다. 다만, CTS 요구사항인 R1~R16까지의 항목이 독립적이지 않고 상호관련성이 높아 연계되는 항목들이 있으며 해당되는 세부내용이 유사하게 기술되어 있는 경우도 있다. 실제 본 연구에서 분석한 10개의 모범사례의 신청서에도 다수의 CTS 요구사항에 대하여 상호참조로 연계하여 기술한 경우를 다수 발견할 수 있었다.

〈표 11〉 CTS 인증 모범사례 분석

카테고리	요구사항	내용		
조직 및 인프라	R1	• 임무 및 범위	리포지터리 사명문	임무, 목표, 범위, 연혁
			호스트 기관	연혁, 펀딩
			데이터 정책 및 보안	수집, 관리, 보존 정책
	R2	• 권한 관리	데이터 정책	수집 및 기탁 정책
			이용약관	라이선스 유형
	R3	• 서비스의 연속성	호스트 기관 지속가능성	데이터 등록 절차
			데이터 연속성 계획	예산 지원 / 조직 연혁
	R4	• 법률 및 윤리	법적, 윤리적 기준	데이터 정책, 표준
			직원교육	승계 계획 및 문서 보존전략 -백업 등
	R5	• 거버넌스 및 자원	호스팅 기관(안정성)	개인정보보호법 등
데이터 보안, 연구윤리 교육 등				
			연혁 및 거버넌스	
			예산 지원	

카테고리	요구사항	내용		
디지털 객체 관리	R6	•전문지식 및 자원	조직	부서, 직원 수 등
			직원 교육	데이터 관리 교육 전문성 신장 교육
			관련 기관 연계	협회 및 회의 참여
	R7	•출처 및 신뢰성	전문가 거버넌스 (운영 위원회 등)	데이터 정책 승인, 행동 강령 등
			전문가 및 커뮤니티 교류	국내외 위원회, 협회 활동 및 자문
			수집/기탁 정책	수집/기탁 요건 및 절차 표준 메타데이터 제공 및 문서화
	R8	•기탁 및 평가	데이터 무결성 및 디지털 객체 손상 검사	체크섬, 메타데이터 품질 체크리스트 등
			버전 전략	감사 추적(변경 기록)
			영구식별자	DOI, ORCID 등
			품질 관리 절차	품질 검사 표준 및 제공
	R9	•보존 계획	메타데이터	검토 및 추적 프로세스
			선호 데이터 형식 목록	데이터 및 메타데이터 완전성 확인
			데이터 기탁 및 수집 정책	데이터 평가(데이터 값 체크리스트) 데이터 평가 기준 미달 시 조치
	R10	•품질 보장	문서화 절차 및 과정	수집 범위 / 편집 시스템 수집 관리 스토리지 시스템
수집 정책			보존 절차 표준 보존 계획/기간	
보존 정책			백업 복사본의 일관성 위험 관리 및 완화 보안	
			보존 정책 및 법령	마이그레이션
			복사, 이전, 저장	메타데이터 표준 백업
			보존계획 문서화	
R11	•워크플로우	데이터 품질 관리(메타데이터 기술/보장 절차)	메타데이터 기술/보장 절차 데이터/ 메타데이터 품질 검사 메타데이터 표준 준수	
		데이터/메타데이터 문의		
R12	•검색 및 식별	워크플로우/데이터 생명주기 프로세스	수집 및 기탁 정책 관리 정책 보안 및 보존 정책	
		워크플로우 도식, 문서화		
R13	•재사용	검색 기능, 방법		
		검색 가능한 메타데이터	메타데이터 표준	
		데이터 권장 인용		
정보 기술 및 보안	R14	영구식별자	DOI, ORCID 등	
		메타데이터 기준		
		메타데이터(스키마) 표준		
		선호 데이터 형식 목록		
		마이그레이션 계획		

카테고리	요구사항	내용		
	R15	• 기술 인프라	메타데이터 표준 및 준수	
			스토리지	
			기술 인프라	
			하드웨어 및 소프트웨어	
			위기관리 전략	재해복구계획/백업
	R16	• 보안	필수서비스 복구/백업	
			보안 정책 문서화	
			보안 시스템 및 리스크 분석	IT 보안 시스템
				보안 수준
				물리적 보안
접근 보안				
재해 및 연속성 계획				
		담당자 및 보안 교육		

조직 및 인프라의 경우, R1은 리포지터리 사명문, 호스트 기관 그리고 데이터 정책 및 법안에 관하여 R2에는 데이터 정책 R3에는 호스트 기관 지속가능성 및 데이터 연속성 계획이 R4에는 법적, 윤리적 기준과 리포지터리 운영 직원의 교육이 R5에는 운영의 안정을 위한 호스팅 기관의 예산, 조직, 전문직원의 교육 등이 마지막으로 R6에는 운영 위원회 등 전문가 거버넌스 및 커뮤니티 활동 등이 포함되어 있다. 다음으로 디지털 객체관리의 경우, R7은 수집/기탁 정책, 데이터 무결성, 데이터 버전 전략 그리고 영구식별자가 R8에는 품질 관리 절차, 메타데이터 등이 R9에는 보존 계획의 문서화 정착 및 과정, 보존 정책 등이 R10에는 마이그레이션, 백업 등 보존 관련한 품질보장이 R11에는 워크플로우 과정에 따른 품질 보장 내용이 R12에는 데이터 라이프 사이클에 따른 수집 및 기탁, 관리 그리고 보안 및 보존 정책이 마지막으로 R13에는 재사용을 위한 메타데이터 표준, 데이터 인용 그리고 영구식별자 관련 내용이 기술되었다.

3. CTS 인증 획득 방안

본 절에서는 4장 2절에서 도출된 CTS 인증 모범사례의 상위요소와 하위요소를 기준으로 CTS 인증을 대비하는 기관의 데이터 정책과 조직 인프라 그리고 기술적 요구사항에 초점을 맞추어 도출하였다. CTS 요구사항이 기본적으로 조직 및 인프라, 디지털 객체 관리 그리고 정보기술 및 보안으로 구성되어 있지만 본 연구는 CTS 인증을 획득하기 위한 기관의 전략을 구성하기 위함으므로 해당 내용을 정책, 조직 그리고 기술 등 세 가지 방향으로 방안을 제시하였다.

10개의 모범사례를 분석 및 종합한 결과 먼저, 데이터 정책은 데이터 수집, 데이터 관리, 데이터 보존 그리고 데이터 공유 및 활용 이렇게 4가지 정책으로 구분할 수 있었다. 다음의 <표 12>는 CTS 인증을 위한 기관 데이터 정책에 포함되어야 할 내용을 각각 수집, 관리, 보존 그리고 공유

및 활용으로 정리한 것이다. 해당 세부 내용은 각 정책 세부 내용으로 구분하여 기술하였다.

〈표 12〉 CTS 인증 모범사례 분석을 통한 데이터 정책 세부내용

정책 카테고리	정책 세부 내용	비고	정책 카테고리	정책 세부 내용	비고
데이터 수집	• 데이터 기탁 및 수집 범위	R1	데이터 보존	• 리포지터리 보존 계획 및 절차	R1
	• 기탁 절차 및 이용약관	R2		• 승계 계획 • 보존 전략 및 백업	R3
	• 수집 시 민감정보 처리 기준	R4		• 데이터 무결성 검사 • 메타데이터 정책: 표준 • 버전 전략 지침 • 영구 식별자	R14
	• 상위 기관 수집 정책	R6		• 데이터 평가 지침 • 메타데이터 정책	R8
	• 메타데이터 제공 및 관련 정책	R7		• 장기보존 기준 • 보존 절차 • 마이그레이션 • 백업 • 스토리지 시스템	R9
	• 선호 데이터 형식 목록 • 데이터 기탁 시 데이터 평가 기준	R8		• 메타데이터 표준 • 데이터 및 메타데이터 품질 검사	R10
	• 수집 정책의 문서화	R14		• 데이터 생명주기에 따른 워크플로우	R11
	데이터 관리	• 데이터 라이프 사이클에 따른 데이터 관리 지침		R1	데이터 공유 및 활용
• 라이선스 정책		R2	• 연구윤리 지침 • 공개위험 데이터 정책	R4	
• 저작권 정책		R2/R4	• 영구식별자	R7	
• 메타데이터 표준		R3	• 메타데이터 표준 준수	R8/R10	
• 연구 윤리 정책 • 개인정보보호 정책 • 공개 위험 데이터 정책: 익명화 등 • 데이터 보안 정책		R4	• 워크플로우에 따른 데이터 공유 및 활용 정책	R11	
• 데이터 관리 인프라 및 조직: 담당자, 위원회 구성 등 • DMP		R5/R6	• 검색 기능 및 방법 • 검색가능한 메타데이터 • 데이터 권장 인용 • 영구 식별자 • 재사용 정책	R12	
			• 메타데이터 표준 • 선호 데이터 형식 목록	R13	

데이터 수집 정책에는 기탁 및 기탁절차, 민감정보, 메타데이터, 선호 데이터 형식 그리고 기탁 시 평가 기준이 포함되어 있다. 다음으로 데이터 관리에는 데이터 라이프 사이클에 따른 관리 지침, 윤리, 저작권 및 라이선스, 메타데이터 표준 그리고 데이터 보안 정책 등이 나타나 있다. 이외 데이터 관리에는 데이터 관리 인프라 및 조직 그리고 DMP와 관련한 내용이 제시된

다. 다음으로 데이터 보존에는 보존 계획, 리포지터리 승계 계획, 데이터 버전 전략, 데이터 평가, 장기 보존 기준, 보존 절차 그리고 데이터 및 메타데이터 품질 검사 등 일반적인 데이터 보존에 따른 가이드와 품질 보장 등이 제시된다. 마지막으로 데이터 공유 및 활용에는 연구윤리, 검색 기능 및 방법, 데이터 인용, 재사용 정책 그리고 메타데이터 표준 등에 대한 내용을 확인할 수 있다.

다음의 <표 13>은 CTS 인증을 준비하기 위한 조직 인프라 관련한 내용을 정리한 것이다.

<표 13> CTS 인증 모범사례 분석을 통한 조직 인프라 세부내용

카테고리	세부 내용	비고
조직 인프라	<ul style="list-style-type: none"> • 연혁, 사명문(임무, 비전, 범위 등), 이사회 등 • 모기관과의 연계: 예산 지원, 인적·기술적 인프라 지원 	R1
	<ul style="list-style-type: none"> • 모기관의 지속 가능성: 연혁, 역할, 충분한 예산, 위원회 등 • 리포지터리: 직원의 전문성, 정기적인 예산 지원 	R3
	<ul style="list-style-type: none"> • 직원의 데이터 윤리 및 보안 교육 • 컨소시엄 구성(윤리적 문제 대응), 연구 윤리 위원회 등 	R4
	<ul style="list-style-type: none"> • 모기관과의 거버넌스 • 충분한 예산 지원 • 충분한 수의 전문성을 갖춘 직원 • 직원의 전문성 및 데이터 관리 교육 • 내외부 전문가 / 커뮤니티 교류 및 지원: 운영 위원회 등 	R5
	<ul style="list-style-type: none"> • 거버넌스: 데이터 관리 워킹그룹 등 • 직원의 전문성 신장 교육 • 국내외 관련 활동 참여: 국내외 기구 및 위원회 소속 회원 • 국내외 커뮤니티와 협력 및 교류: 자문 및 MOU체결 	R6
	<ul style="list-style-type: none"> • 보안 담당 전문 직원 • 정보 보안 관련 직원 교육 	R16

조직 인프라는 리포지터리를 운영하는 조직, 인력 그리고 예산 등과 관련한 내용을 담고 있다. 특히, 해당 리포지터리의 지속 가능성에 따른 예산, 위원회, 직원 및 직원의 전문성, 전문성을 위한 교육 여부 등과 관련한 내용과 내외부 커뮤니티 활동을 통한 위원회 참여 및 개최 여부 등에 대한 것도 포함되어 있다.

다음의 <표 14>는 기술 인프라 측면에서 CTS 인증 준비 시 필요한 리포지터리의 기술적 요구 사항을 정리한 것이다.

기술 인프라는 데이터 리포지터리의 시스템 요구사항을 기준으로 정리하였다. 모범사례의 신청서 내용을 살펴본 결과, 데이터 리포지터리 기술적 요구사항은 ‘검색’, ‘접근’, ‘메타데이터’, ‘품질보장’, ‘보안’, ‘시스템 지원 및 개발’ 그리고 ‘기타’ 등 7개의 카테고리로 구성하였다. ‘기타’에는 DOI와 같은 영구식별자의 발급, 워크플로우 그리고 인용에 대한 내용을 포함하고 있다.

〈표 14〉 CTS 인증 모범사례 분석을 통한 기술 인프라 세부내용

카테고리	요구사항	세부내용	
기술 인프라	검색	<ul style="list-style-type: none"> • 다양한 검색 방법 • 맞춤형 검색 인터페이스 또는 데이터 접근 방법 제공 	<ul style="list-style-type: none"> • 검색 API • 선호 데이터 형식 목록 제공
	접근	<ul style="list-style-type: none"> • 데이터 수집 시스템 • 기관 전용 라이선스 및 오픈 데이터 사용 시, 접근 권한에 관한 기술적 부분 지원 	<ul style="list-style-type: none"> • 접근 권한 및 인증 제한 시스템 • 데이터 형식 및 변환기 지원 • 문서화된 프로세스에 따른 데이터 저장
	메타 데이터	<ul style="list-style-type: none"> • 출판 후 데이터 및 메타데이터 변경 절차 • 메타데이터 표준 및 검토 프로세스 • 메타데이터 완전성 및 정확성 측정 검사 	<ul style="list-style-type: none"> • 메타데이터 품질 체크리스트 및 검사 • 데이터 및 메타데이터 카탈로그 지원 • 국제 표준 메타데이터 지원
	품질보장	<ul style="list-style-type: none"> • 데이터 업로드 및 다운로드 전후 데이터 손실 체크 • 데이터 변경 시 감사 추적 시스템(audit trail) • 데이터셋 출판 및 오류 발견 시 조치 • 레코드 평가 프로세스 및 평가 도구 • 품질 관리 프로세스 • 출판 프로세스 데이터셋 및 단계 변경사항 추적 시스템 	<ul style="list-style-type: none"> • 버전관리 시스템: 변경사항 발생 시, 새 버전 생성 • 데이터 파일 및 코드 무결성 검사 • 체크섬 및 고정성 검사 • 마이그레이션 방법 및 구현 • 형식 및 표준 변경 시 정기적 업데이트 시스템
	보안	<ul style="list-style-type: none"> • 데이터 검토 시스템 • 중요 정보 및 민감 데이터 익명화 • 보안 암호로 된 서비스를 통한 데이터 전송 • 데이터 담당자의 시스템 정기적 모니터링 • 보존과 보안에 관한 시스템적 관리: 위험 완화 도구 사용 	<ul style="list-style-type: none"> • 시스템 및 물리적 보안 • 데이터 관리 위험 및 영향 평가 • 장애/재해 복구 시스템 • 백업 및 미러링 • 보존 수준에 따른 접근 권한 제한 시스템 • 백업 시 온/오프라인에서 원본 파일 관리 시스템
	시스템 지원 및 개발	<ul style="list-style-type: none"> • 데이터 포털 구축 및 운영 • 컬렉션 복사, 변환, 저장, 접근 가능한 시스템 지원 • 보존 프로그램 및 리포지터리 구현 • OAIS 기반 보존 시스템 개발 • 국제 및 관련 도메인 표준으로 시스템 구축 	<ul style="list-style-type: none"> • 대역폭 지원 • 클라우드 스토리지 및 클라우드 컴퓨팅 구축 및 지원 • 스토리지 및 분산, 백업 시스템 • 수집 및 관리, 배포, 공유 기술과 관련된 인프라 제공
	기타	<ul style="list-style-type: none"> • DOI 등과 같은 영구식별자 발급 및 할당 프로세스 	<ul style="list-style-type: none"> • 접근 및 수집 지원하는 워크플로우 도식화 • 데이터셋 랜딩 페이지에 인용

V. 논의 및 결론

본 연구의 목적은 CTS 인증을 확보하기 위한 방안을 제시하는 것이다. 해당 목적을 달성하기 위하여 먼저 re3data.org와 같은 연구데이터 리포지터리 등록 사이트를 조사 및 분석하였다. 해당 내용은 주제별, 국가별, 인증체계별 등에 대한 것으로 가장 많이 차지하는 주제는 인문학 및 사회과학, 생명과학, 자연과학 그리고 공학 순으로 나타났다. 이 중 본 연구의 핵심 대상인 Geoscience 분야가 포함된 자연과학에서는 지구과학이 가장 많은 분포를 차지했다. 국가별 분석에서는 미국, 독일, 캐나다, 영국 등의 국가에서 가장 많은 리포지터리를 보유하고 있으며 특히, 중국 및 일본도 각각 81개, 66개 등으로 나타났지만 한국은 13개로 리포지터리 보유가 저조하며 이마저도 국가간 협력을 제외하면 운영하고 있는 리포지터리가 8개로 파악되었다. 또한, 인증 체계별 분석에서는 CTS가 전체 비중의 46%를 차지하고 있음을 알 수 있으며 Geoscience 분야에서도 CTS 인증이 약 90%의 분포를 보이고 있다. 이외에도 메타데이터, 콘텐츠 유형 그리고 소프트웨어에 대한

분석도 진행하였으며 상세 내용은 본 논문의 3장에서 기술하였다.

다음으로 CTS 인증 획득 방안을 제시하기 위하여 Geoscience 분야의 CTS 인증을 받은 10개의 리포지터리를 대상으로 모범사례를 분석하였다. 분석 결과는 CTS 인증을 획득하기 위한 방안으로서 데이터 정책, 조직 인프라 그리고 기술 인프라 3가지 차원으로 구성하였다. 우선, 데이터 정책 차원에서는 데이터 수집, 관리, 데이터 보존 그리고 공유 및 활용 정책을 중심으로 해당 정책의 세부 내용이 어떻게 기술되어야 하는지를 제시하였다. 두 번째는 리포지터리를 운영하는 조직 인프라 관련한 세부 내용을 제시하였다. 해당 내용에서는 예산, 직원, 직원의 전문성 그리고 각종 커뮤니티 활동에 대한 사항을 포함하고 있다. 마지막으로 리포지터리의 기술 인프라 관련하여 해당 리포지터리 구축 및 운영시 요구사항에 대하여 제시하였다. 요구사항은 크게 7가지로 구성되어 있으며 해당 내용은 '검색', '접근', '메타데이터', '품질보장', '보안', '시스템 지원 및 개발' 및 '기타'가 포함되었다.

데이터 공유와 재사용은 이제 필수가 되고 있다. 특히 데이터 품질은 데이터 공유를 성공적으로 이끌기 위한 동인이 되었으며 따라서 데이터 품질을 확보하기 위한 국내외 노력은 다양하게 구현되고 있다. 그 중 데이터를 담는 그릇이라고 할 수 있는 데이터 리포지터리는 데이터 공유와 재사용을 실현시키기 위한 중요한 도구라고 볼 수 있다. 모든 데이터의 품질을 담보할 수는 없지만 데이터 리포지터리에 대한 인증을 통해 해당 데이터의 신뢰성을 확보할 수 있을 것이다. 특히 CTS 인증이 대세가 되어 가고 있는 현실에 앞으로 국내 데이터 리포지터리의 정책, 조직 그리고 기술적 요구사항에 대한 기준을 해당 인증이 제시할 수 있을 것이라 판단된다. CTS 인증은 결국 데이터 리포지터리 품질 기준으로 이어질 수 있다. 따라서 앞으로 데이터 리포지터리의 품질 기준 또는 평가 기준과 관련한 연구가 이어지기를 기대해 본다.

참 고 문 헌

- 한종규 (2019). 연구데이터 리포지터리 구축 경험과 과제(지질자원분야). 19년 한국전문도서관협의회-국가정책정보협의회 공동 학술세미나, 2019년 6월 27일. 여수: 디오션리조트 호텔.
- Cho, J. (2019). Study of Asian RDR based on re3data. The Electronic Library, 37(2), 302-313. <https://doi.org/10.1108/EL-01-2019-0016>
- coretrustseal (2023, April 8). CORETRUSTSEAL. Available: <https://www.coretrustseal.org/>
- Donaldson, D. R. (2020). Certification information on trustworthy digital repository websites: A content analysis. PloS One, 15(12), e0242525. <https://doi.org/10.1371/journal.pone.0242525>
- Downs, R. (2021). Improving opportunities for new value of open data: assessing and certifying research data repositories. Data Science Journal, 20(1), 1-11. <https://doi.org/10.5334/dsj-2021-001>

- FAIRsFAIR (2023, April 15). FAIRsFAIR. Available: <https://www.fairsfair.eu/>
- Frank, R. D. (2022). Risk in the CoreTrustSeal data repository certification process. iConference 2022 Posters. <http://hdl.handle.net/2142/113742>
- Khan, A. M., Loan, F. A., Parray, U. Y., & Rashid, S. (2023). Global overview of research data repositories: an analysis of re3data registry. *Information Discovery and Delivery*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/IDD-07-2022-0069>
- Khan, N., Thelwall, M., & Kousha, K. (2022). Are data repositories fettered? a survey of current practices, challenges and future technologies. *Online Information Review*, 46(3), 483-502. <https://doi.org/10.1108/OIR-04-2021-0204>
- Kindling, M., Pampel, H., Sandt, S. van de, Rücknagel, J., Vierkant, P., Kloska, G., Witt, M., Schirmbacher, P., Bertelmann, R., & Scholze, F. (2017). The Landscape of research data repositories in 2015: A re3data analysis. *D-Lib Magazine*, 23(3/4). <https://doi.org/10.1045/march2017-kindling>.
- Lindlar, M. & Rudnik, P. (2019). Eye on CoreTrustSeal - Recommendations for criterion R0 from digital preservation and research data management perspectives. 16th International Conference on Digital Preservation iPRES 2019, Amsterdam, The Netherlands. <https://doi.org/10.5281/zenodo.3267689>
- Misgar, S. M., Bhat, A., & Wani, Z. A. (2020). A study of Open Access research data repositories developed by BRICS countries. *Digital Library Perspectives*, 38(1), 45-54. <https://doi.org/10.1108/DLP-02-2020-0012>
- Peng, G., Gross, W. S., & Edmunds, R. (2022). Crosswalks among stewardship maturity assessment approaches promoting trustworthy FAIR data and repositories. *Sci Data* 9, 576. <https://doi.org/10.1038/s41597-022-01683-x>
- Pisharoti, N. & Petters, J. L. (2019). An assessment of VTechData with respect to the CoreTrustSeal Repository Certification Requirements. <http://hdl.handle.net/10919/89054>
- re3data (2023, April 5). re3data.org. Available: <https://www.re3data.org/>

• 국한문 참고문헌의 영문 표기

(English translation / Romanization of references originally written in Korean)

- Han, Jong Kyu (2019). Research data repository building experience and challenges (Geoscience). 19 Korea Special Library Association-National Association for Policy Information Joint Academic Seminar, June 27, 2019. Yeosu: The Ocean Resort Hotel.