

국가 데이터 통합 거버넌스 관점의 데이터 클리어링 하우스 시스템 아키텍처

이영주¹, 오대훈², 신신애³

¹한국지능정보사회진흥원 팀장

²한국지능정보사회진흥원 선임연구원

³한국지능정보사회진흥원 본부장

lyj@nia.or.kr, ohdang93@nia.or.kr, sashin@nia.or.kr(교신저자)

Designing a data clearinghouse system architecture for National Integrated Data Governance

Young-Joo Lee¹, Daehoon Oh², Shinae Shin³

^{1,2,3}Dept. of AI Data, National Information society Agency

⁴Corresponding Author(sashin@nia.or.kr)

요약

생성형 인공지능의 확산과 데이터 확보 경쟁이 치열해지면서 고품질, 고부가가치 데이터를 유통, 거래할 수 있는 생태계 조성의 필요성이 높아졌다. 그러나 데이터 소유자는 주권 보장, 수익 배분의 불확실성으로 인해 데이터 개방과 제공에 소극적인 것이 현실이다. 데이터 클리어링하우스는 참여자 간 수익 배분 합의 메커니즘을 설정하고 분산형 데이터 처리기술을 적용한 유통 이력관리를 통해 고품질 데이터의 안전한 거래와 신뢰성을 확보할 수 있는 기술적 기반을 제공한다. 본 연구에서는 국가 데이터 통합 거버넌스 내에서 데이터 클리어링하우스 설치 및 운영에 필요한 기능 요구사항과 비즈니스 시스템 아키텍처를 제안하고자 한다.

1. 서론

최근 ChatGPT의 등장과 함께 전세계적으로 촉발된 생성형 인공지능의 개발 경쟁에 따라, 고품질 학습용 데이터에 대한 수요가 증가하고 있다. 이와 더불어 국가 차원에서 데이터의 유통거래를 활성화하기 위한 다양한 정책들이 시행되고 있다[1]. 국내 데이터 시장은 연평균 5% 수준으로 성장세를 유지하고 있으나 막상 중요한 유통거래의 활성화는 지체되고 있는 상황이다[2]. 이에 대한 원인으로는 데이터에 대한 소유권과 영업 비밀 등 자산 침해 우려, 데이터 제공 이후 통제권의 상실 우려 등이 지목되고 있으며[3], 생성형 인공지능 개발에 필요한 뉴스 등의 컨텐츠는 저작권 침해 논란에 대해 언론사와 AI기업 간 서로 다른 주장을 하며 대립 중이다[4].

이와 같이 우리나라에서 데이터 유통거래를 활성화하고 데이터 경제를 창출하기 위해서는 데이터 시장에 참여하는 수요자와 공급자 간의 신뢰와 합의를 형성할 수 있는 제도와 거버넌스, 지원시스템 등이 필요한 상황이다. 세계 최대 단일시장인 EU는 테이

터 거래 시장을 포함한 역내 디지털 시장의 공정성, 투명성, 주권 보장을 위해 Gaia-X 프레임워크 하에 디지털 클리어링 하우스(이하 GXDCH)의 개념을 도입하였다[5].

본 연구에서는 이러한 배경에 따라 우리나라에서 데이터 유통거래 활성화에 필요한 공통 기술 기반으로 데이터 클리어링 하우스를 제안하고, 향후 국가 차원의 표준을 제정하고 분야별 데이터 플랫폼들이 참조할 수 있도록 주요 기능적 요구사항을 정의하고자 한다. 또한 최근 과기정통부 주도로 추진 중인 국가데이터 인프라 거버넌스 체계 하에서 유관 시스템 및 사용 주체들과 데이터 클리어링하우스 간 연계를 고려한 시스템 아키텍처를 제안한다.

2. 데이터 클리어링 하우스의 개념

본 연구에서는 데이터 클리어링 하우스를 ‘데이터의 유통거래 흐름에 따른 관련 정보를 생성, 관리, 감독하고 데이터의 이용 내역을 근거로 수익 정산 및 분배가 이루어지는 신뢰기반 운영체계’로 정의하고자 한다. 이는 데이터 컴플라이언스 준수를 검

증하는 EU의 GXDCH와 음반, 전자도서 등 디지털 콘텐츠 시장에서 저작권료 정산 및 분배를 수행하는 디지털 콘텐츠 클리어링 하우스의 개념을 데이터 유통거래 생태계에 적용한 것으로 이해할 수 있다. 데이터 클리어링 하우스는 국내 데이터 유통거래 시장 활성화 뿐 아니라 EU 데이터 규제에 대응하고 국제 데이터 시장 참여를 위해서도 필요성이 인정된다.

한국형 데이터 클리어링 하우스는 다음과 같은 구현 원칙을 가진다. 첫째로 데이터 거래에 대한 정보(가격, 거래조건, 데이터 사용범위 등)가 거래 당사자 간 투명하게 공개되어야 한다 두 번째로 상호 운용 가능한 플랫폼을 통해 표준화된 인터페이스를 제공해야 한다. 세 번째로 데이터에 대한 소유권과 사용 권리가 명확히 지켜질 수 있도록 보호 기술을 제공해야 한다. 마지막으로 개인정보보호법 등 민감한 데이터에 대한 법적 요구사항을 충족해야 한다[6].

3. 데이터 클리어링 하우스 시스템 아키텍처

본 연구에서는 선행 연구를 참고하여 특정 플랫폼에서 데이터 공급자와 수요자 사이의 데이터 거래 환경을 가정하여 구매 계약부터 데이터 교환, 이력 관리, 수익 분배 및 정산 과정에 걸쳐 데이터 클리어링 하우스가 제공해야 할 기능 요구사항을 정의한다.

3.1. 데이터 거래 관리기능

데이터 거래 관리기능은 데이터 제공자가 거래 가능한 데이터셋 정보와 메타데이터 및 가격 정책을 입력하는 데이터 등록 기능, 데이터 수요자가 거래 계약 완료 후 데이터 이용 목적, 기간, 범위 등 조건 정의 및 스마트 계약 기반 자동화 기능을 제공하는 계약 관리, 거래 승인 후 데이터 암호화 기반 전송을 인증하는 전송 인증기능을 포함한다.

3.2 수익 정산 및 분배 기능

정산 기능은 데이터 사용량 기반 정산 방식과 기여도 기반 가중치를 적용한 수익 분배 기준 설정 방법을 적용할 수 있다[7][8]. 이를 기반으로 클리어링 하우스에는 데이터 결합 시 각 데이터의 기여도를 계산하여 수익 비율을 산정하는 기여도 분석관리 기능, 거래 완료 시 제공자별 수익을 자동 계산하고 지급하는 자동 정산 기능, 거래 수수료 정산 기능 등이 포함된다.

3.3. 거래 투명성 및 감사 기능

디지털 워터마킹 기술과 해시 링크 추적 기술을 기반으로 한 데이터셋의 유통 및 이용 내역 추적 기능, 블록체인 또는 영지식증명에 기반한 거래 인증 기능, 관련 리포트를 제공하는 정산 대시보드와 감사 리포팅 기능 등이 포함된다.[9][10]

3.4 컴플라이언스 기능

국내 개인정보보호법 뿐 아니라 GDPR, CCPA 등 글로벌 데이터 규제 준수 여부를 확인할 수 있는 개인정보 보호 검증 기능[11], 개인정보 포함 여부 및 비식별 조치 여부를 점검하는 비식별조치 적합성 검사 기능, 데이터 소유권, 라이선스, 저작권 정보 관리를 검증하는 권리관계 검증 기능 등이 포함될 수 있다[12][13].

3.5 위험 관리 및 중재 기능

거래 조건, 데이터 품질, 법적 리스크 사전 검증 프로세스, 데이터 품질 문제, 계약 위반 발생 시 중재 요청 및 프로세스 지원, 데이터 거래 사고 발생 시 보상 처리 기능 등을 포함한다[14].

3.6. API 및 외부 연동 기능

타 데이터 거래소 및 기관의 데이터셋 검색·거래를 위한 외부 데이터 카탈로그 연동 기능, 분석 서비스 플랫폼과 API 연동으로 데이터를 직접 분석, 가공할 수 있는 분석 플랫폼 연동 기능이 포함된다.

이상으로 도출된 기능 요구사항을 시스템 아키텍처의 계층별 구성요소로 표현한 결과는 <표 1>과 같다. 각 구성요소의 타당성은 GDCH에서 정의한 기능과의 매핑을 통해 확인할 수 있다.

<표 1> 데이터 클리어링 하우스 아키텍처 계층별 구성요소

Layer	Components	
	Data Clearinghouse	Gaia-X Digital Clearinghouse
User Layer	data provider	data provider
	data user	data consumer
	system administrator	Federation Operator (Service Operator)
Application Layer	transaction mgt. service	Usage Control Service
	revenue distribution service	Usage Accounting Service
	contribution evaluation engine	Self-Description Validator

	data compliance evaluation	Compliance Verification
	audit log service	Service Credential Verifier
	risk mitigation service	Trust & Identity Framework
	api gateway	Federation API, Data Service API
Data Layer	transaction database	Transaction Record DB
	transaction clearing database	Usage History DB, Clearing Ledger
	metadata catalog	Gaia-X Federated Catalog (Self-Descriptions)
	data distribution history DB	Usage Log Repository
	data rights/agreement DB	Contract Store / Usage Policy Repository
	audit log DB	Compliance Audit DB
Interface Layer	blockchain and smart contract	Verifiable Credentials / Decentralized Identity (DID/VC)
	external data exchange	Inter-Federation API / Secure Connector
	transaction history api	Usage Logs API / Event Notification

4. 국가 데이터 인프라와 표준 데이터 카탈로그

과학기술정보통신부와 한국지능정보사회진흥원은 2024년부터 국내 각 분야 간 데이터의 자유로운 흐름을 위한 데이터 소재 정보를 통합 제공하고 데이터 산업 활성화 관련 정책의 원스톱 서비스 제공을 위해 국가 차원의 데이터 공통기준과 공통인프라 구축을 진행 중이다. 이의 일환으로 국내 민간, 공공의 각 데이터 플랫폼의 소재정보를 하나의 기준으로 관리하기 위해 국제표준을 준용한 국가 표준 데이터카탈로그를 정의하였다[15]. 향후 이를 기반으로 한 확장된 메타데이터를 연계하여 하나의 포털에서 국내 산재된 데이터를 손쉽게 검색하고 활용할 수 있는 정보를 제공할 예정이다. 현재 원-원도우에는 공공 데이터 포털을 포함한 국내 40개의 데이터플랫폼의 메타데이터가 연계되어 시범 운영 중이다[16].

표준 데이터 카탈로그는 DCAT-AP 2.1을 기반으로 한 공통 태그와 분야간 데이터 활용에 필요한 큐레이션 정보를 담은 확장 태그로 구성되어 있다. 또한 DCAT-AP 2.1에서 정의한 dataset 클래스와 data series 클래스에는 데이터 클리어링하우스에서 관리해야 할 개별 데이터셋의 식별자, 소유자, 이용권한, 사용조건 등을 정의할 수 있는 속성을 포함하고 있다.

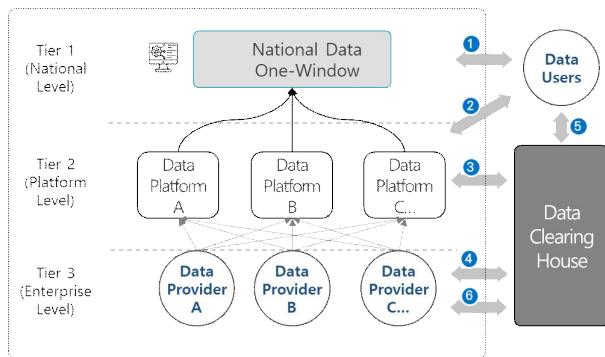
5. 국가 데이터 인프라 통합 거버넌스 내 데이터 클리어링 하우스 운영 모델

국가데이터인프라 거버넌스 체계 하에서 각 데이터 플랫폼들과 데이터 수요자와 공급자, 그리고 데이터 클리어링하우스를 구성하는 최상위 비즈니스 아키텍처는 <그림 1>과 같다.

국가데이터 원-원도우는 연계된 데이터 플랫폼이 제공하는 데이터 소재정보를 통합 검색할 수 있는 기능을 데이터 사용자에게 제공한다. 사용자는 원-원도우를 통해 필요한 데이터를 보유한 플랫폼을 식별할 수 있다. 각 플랫폼은 데이터를 유통할 수 있는 마켓플레이스를 운영하면서 직접 데이터를 보유하거나 결합, 가공을 위해 다시 원천데이터를 보유한 개별 기업을 공급자로 두게 된다. 이를 통해 실제 데이터 큐레이터 역할을 수행한다.

데이터 클리어링하우스의 구현 위치는 정책적 의사결정에 따라 다양하게 구성할 수 있지만, 기본적으로 플랫폼을 통해 기관(enterprise)과 기관 간 데이터 유통 및 거래를 가정하였을 때 클리어링 하우스는 별도의 중립적인 기관을 두어 그 역할을 수행하는 것이 바람직하다고 할 수 있다. 선행연구[17]를 기반으로 데이터 사용자 관점에서 데이터 검색부터 최종 활용 후 비용 지불 까지의 과정 중에 각 구성 요소 간 인터페이스 요건은 다음과 같다.

- ① 데이터 사용자는 원원도우 포털을 통해 데이터 카탈로그를 기반으로 필요한 데이터를 검색하고 식별
- ② 식별된 데이터에 대해 데이터 사용자와 플랫폼은 데이터 구매 협의 및 스마트 계약을 생성
- ③ 플랫폼 관리자는 클리어링 하우스를 통해 캡플라이언스 검증 결과에 따라 계약을 승인하고 클리어링 하우스에 계약 정보를 전송
- ④ 데이터 제공자는 데이터 플랫폼을 통해 암호화 또는 API 방식으로 데이터 제공 후 그 이력을 클리어링 하우스에 전송
- ⑤ 데이터 사용자는 데이터 취득 및 활용 후 계약에 따른 정산 및 수수료를 플랫폼에 지급하고 그 이력을 클리어링 하우스에 전송하고 공증
- ⑥ 데이터 제공자는 플랫폼을 통해 수익을 정산받고 클리어링하우스를 통해 그 이력을 확인



<그림 1> 국가데이터 인프라와 데이터클리어링 하우스 운영 모델

6. 결 론

공동의 이해관계를 가진 수요·공급자가 모여서 배타적 공간에서 데이터를 합의된 방식으로 공유하고 혁신적인 서비스를 창출하는 데이터 혁신 공간(데이터 스페이스)의 필요성이 강조되고 있다. 데이터 클리어링하우스는 이 데이터 혁신 커뮤니티를 신뢰 기반으로 운영할 때 필요한 가장 중요한 기술 기반이다. 민간의 데이터 유통 거래 뿐 아니라 공공 영역에서도 필요 시 데이터를 안전한 공간에서 공유, 활용할 수 있는 수단으로도 활용 가능하다.

우리나라는 현재 정부의 대규모 데이터 투자를 통해 다수의 데이터 플랫폼과 데이터 공급 기업을 육성하였으나, 실제 고품질 고부가가치 데이터의 경우 데이터 제공 이후의 통제 가능성과 수익 보상의 불확실성으로 인해 데이터 유통에 적극적이지 않은 상황이다. 데이터 클리어링하우스는 참여자 간 수익 배분 합의를 통해 고품질의 (학습용)데이터를 개방·공유·거래하고, 이력관리를 통해 정보의 비대칭이 해소된 Peach Market 형성을 통해 건전한 데이터 유통거래 질서를 확립할 수 있다. 향후 본 연구의 결과물을 활용한 개념검증(PoC)을 거쳐 국내 주요 분야에서 본격적 확산을 기대한다.

* 본 연구는 과학기술정보통신부와 한국지능정보사회진흥원의 ‘국가 데이터 산업 인프라 조성’ 사업의 지원을 받아 수행되었음

참고문헌

- [1] 관계부처 합동, “제1차(‘25년~’25년) 데이터산업 진흥 기본계획”, 2023
- [2] 한국데이터산업진흥원, “2023 데이터산업 현황조사”, 2024
- [3] 한국지능정보사회진흥원, 내부 조사자료
- [4] 연합뉴스, “지상파·네이버 AI 저작권 분쟁 파장…핵심은 ‘공정 이용’ 여부”, 2025.1.18 온라인기사
- [5] <https://gaia-x.eu/services-deliverables/digital-clearing-house/>, (access date, 2025.4.22.)
- [6] EU, Digital Clearinghouse 2.0, 2025
- [7] Luo, Xuan, et al. "On shapley value in data assemblage under independent utility."arXiv preprint arXiv:2208.01163(2022).
- [8] Zhao, Boxin, et al. "Addressing budget allocation and revenue allocation in data market environments using an adaptive sampling algorithm."International Conference on Machine Learning. PMLR, 2023.
- [9] Gujar, S. S. "Blockchain-Based Framework for Secure IoT Data Transmission," 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2024, pp. 1–6
- [10] Zhang B, Pan H, Li K, Xing Y, Wang J, Fan D, Zhang W. A Blockchain and Zero Knowledge Proof Based Data Security Transaction Method in Distributed Computing. Electronics. 2024; 13(21):4260
- [11] Cejas, O. A., Azeem, M. I., Abualhaija, S., & Briand, L. C. (2023). Nlp-based automated compliance checking of data processing agreements against gdpr. IEEE Transactions on Software Engineering, 49(9), 4282–4303.
- [12] Ahmad, Haris, Gagandeet Singh Aujla, GDPR compliance verification through a user-centric blockchain approach in multi-cloud environment, Computers and Electrical Engineering, Volume 109, Part B, 2023
- [13] Liu Y, Zhang Y, Yang Y, Ma Y. DOCS: A Data Ownership Confirmation Scheme for Distributed Data Trading.Systems. 2022; 10(6):226.
- [14] Zhou C, Yang Z, Yue S, Xuan B, Wang X. Securing Big Data Exchange: An Integrated Blockchain Framework for Full-Lifecycle Data Trading with Trust and Dispute Resolution.Electronics. 2023; 12(24):4894.
- [15] 한국지능정보사회진흥원, 내부자료
- [16] www.data1window.kr
- [17] 김근형, ”신뢰가 보장된 데이터 경제를 위한 자기 주권 데이터 유통플랫폼 설계“, 디지털콘텐츠학회논문지Journal of Digital Contents SocietyVol. 22, No. 3, pp. 483–490, Mar. 2021