



데이터 관리체계 구축방안 연구 : KOICA 개발협력 프로젝트 데이터 중심으로

2021. 12

**데이터 관리체계 구축방안 연구:
KOICA 개발협력 프로젝트 데이터 중심으로**

2021. 12

한국국제협력단

저자

책임연구원

윤 유 리

- (현) 한국국제협력단 ODA연구센터 연구원/과장
- 성균관대학교 경영학(무역학) 박사

주요 연구 및 사업 실적

- ▷ (공저) 2021년, “TOSSD Pillar II 의의와 STI 개발협력사업 확대방안.” no. 68
- ▷ (공저) 2021년, “개발협력 통합적 접근법의 의미와 사례.” 개발과 이슈. no. 67
- ▷ (공저) 2020년, 개발협력 프로젝트 데이터 활용방안 연구. 성남: 한국국제협력단
- ▷ (책임) 2020년, 4차 산업혁명 기반 혁신기술 활용을 통한 개발협력사업 발전방안 연구. 성남: 한국국제협력단
- ▷ 2019년, 신북방정책과 ODA전략 연계 방향 연구. 성남: 한국국제협력단
- ▷ 2019년, “UN 고위급정치포럼(HLPF)의 SDGs 이행과정검토Ⅳ.” 개발과 이슈. no. 59
- ▷ 2019년, 선진국의 중진국 원조 출구전략 연구. 성남: 한국국제협력단
- ▷ 2019년, “2030 어젠다 달성을 위한 GPEDC의 역할 및 한국의 기여 방안.” 개발과 이슈. no. 53

공동연구원

안 미 선

- (현) 한국국제협력단 ODA연구센터 연구원/전문관
- 서울대학교 국제대학원 국제지역학 석사

주요 연구 및 사업 실적

- ▷ (책임) 2021년, “TOSSD Pillar II 의의와 STI 개발협력사업 확대방안.” no. 68
- ▷ (공저) 2021년, “개발협력 통합적 접근법의 의미와 사례.” 개발과 이슈. no. 67
- ▷ (공저) 2020년, 4차 산업혁명 기반 혁신기술 활용을 통한 개발협력사업 발전방안 연구. 성남: 한국국제협력단
- ▷ (공저) 2020년, “개발협력의 관점에서 본 코로나19의 영향과 대응방안.” 개발과 이슈 no. 61

이 새 물

- (현) 한국국제협력단 ODA연구센터 전임
- 경희대학교 평화복지대학원 국제관계학 석사

본 연구보고서의 내용은 한국국제협력단의 공식적인 입장과 일치하지 않을 수 있음을 밝힙니다.

본 보고서에서는 제3자에게 권리(저작권)가 있는 저작물이 포함되어 있을 수 있습니다. 해당 내용 이용 허락은 사용자가 권리당사자에게 직접 확인해야 하며 그 사용에 관한 일체의 책임은 사용자에게 있습니다.

본 연구보고서의 내용은 출처와 집필자 명시 하에 인용이 가능하며, 아래와 같이 표기 바랍니다.

윤유리, 안미선, 이새물. 2021. 데이터 관리체계 구축방안 연구: KOICA 개발협력 프로젝트 데이터 중심으로. 성남: 한국국제협력단.

※ 본 자료는 KOICA 도서관 홈페이지(<http://lib.koica.go.kr>)를 통해서도 볼 수 있습니다.

국문 요약

4차 산업혁명 시대에 데이터의 가치는 빠르게 증대되고, 더욱이 코로나바이러스의 확산은 전 세계의 디지털화를 가속했다. 이에 따라 데이터 분석 및 활용이 더욱 중요해지고, 데이터 활용 역량은 개인 및 기관의 핵심 경쟁력으로 대두되고 있다. 이러한 추세를 반영하여 KOICA는 데이터의 원활한 수집·관리·활용을 위해 프로세스 및 시스템 개선, 인프라 최신화로 디지털 KOICA로 도약하고자 디지털 전환 전략을 수립하였다. 현재 KOICA에서 수집·생산되는 다양한 자료에 비해 데이터 관리체계는 다소 미흡하다. 데이터를 통해 개발협력 사업 현황 및 성과를 분석하고, 향후 사업 방향성 및 전략 수립 등에 적극 활용하기 위해서는 동일한 표준과 규정으로 관리되는 양질의 데이터가 매우 중요하며, 이를 가능하게 하는 데이터 거버넌스 정립이 필수적이다. 이에 본 연구에서는 데이터 거버넌스의 정의 및 핵심 요소를 살펴보고, 나아가 KOICA 개발협력 사업의 데이터 거버넌스 수립을 위한 근거자료로 활용될 수 있도록 데이터 현황을 분석하고 향후 방향성을 제시하고자 한다.

데이터의 폭발적인 증가와 그 가치로 인해 많은 기관에서 데이터 거버넌스의 필요성에 대해 강조하고 있으나 데이터 거버넌스에 대한 합의된 정의는 없는 것으로 보인다. 다수의 선행연구를 살펴보면 데이터 거버넌스에 대한 정의는 상이하나, 양질의 데이터를 수집하고 관리하여 조직에서의 다양한 가치 창출에 지속적으로 활용되도록 하는 데 그 목적을 둔다는 점에서는 유사하다. 이에 본 연구에서는 데이터 거버넌스를 데이터의 생성부터 활용까지의 전체 관리 과정에서 필요한 전략, 정책, 기준, 역할, 책임 등에 관한 체계라고 정의하고자 한다.

데이터의 자산으로서의 가치가 급증하고 있다고 하나, 기관에서의 데이터 가치는 아직까지는 저평가되거나 이에 대한 체계적인 관리를 위한 투자로 이어지기까지는 조직 내 상당한 견해 차이가 있는 것으로 사료된다. 데이터의 생성부터 활용까지 전체 사슬의 참여자들 간의 원활한 소통을 위해, 그리고 데이터를 잘 활용하는 동시에 위험과 비용을 줄이기 위해 데이터 거버넌스를 수립할 필요가 있다. 다양한 경로를 통해 생성 및 수집되는 데이터를 수많은 사람들이 필요에 따라 활용하기까지 여러 제약조건들이 존재한다. 또한 데이터는 많으나, 정작 필요한 데이터는 찾기 어렵거나 가공에 상당한 시간이 요구되는 경우가 많다. 이는 데이터의 공급과 수요(혹은 생성과 활용)가 원활하게 연결되어 있지 않기 때문이라고 볼 수 있다. 결국 데이터 거버넌스는

데이터 사슬체계 안에서의 각각의 역할자 간의 연결고리 역할을 하는 프레임워크다. 데이터 거버넌스는 데이터를 전체적으로 관리하는 하나의 운영 체계인 바, 기관의 성격에 따라 사업전략과 목표에 부합하도록 구성되어야 하며, 이를 위해 어떤 데이터가 어디서 생성되어 누가 권한을 갖는지, 누가 어떤 방식으로 활용하는지 살펴보는 것이 필요하다. 데이터 거버넌스의 구축에 정해진 순서는 없다. 다만, 기관의 특성과 상황에 적합한 데이터 거버넌스를 정의하여 작은 부분부터 구축하여 확대해 나아갈 것을 제시한다.

데이터 거버넌스의 주요 요소로 표준화, 품질, 생애주기를 들 수 있다. 데이터 표준화는 다양하게 존재하는 데이터를 특정 기준에 따라 정리하는 것으로, 데이터의 활용 가능성을 높이기 위한 사용자들 간의 또는 기관 내 공통의 언어 및 규칙을 설정하는 과정이다. 이렇게 마련된 기준은 기존의 데이터를 재정비하는 과정에서도, 나아가 일관성 있는 데이터 품질 관리를 위해서 반드시 필요한 절차이기도 하다. 또한 데이터를 공유하는 범위가 넓어지면서 데이터의 잘못된 활용을 방지하고 올바른 활용 및 확산을 위해 양질의 데이터로 관리하는 것이 중요하다. 데이터의 품질은 정해진 기준에 따라 정확히 입력이 되어 있는지, 입력된 데이터 값의 의미를 제대로 전달할 수 있도록 관리되고 있는지를 검토하고 관리하는 과정이다. 이를 위해 세부적인 품질 기준을 마련하고 품질 관리 전담 조직 구성이 필요하며, 표준화를 포함하여 데이터 저장 시 고려사항, 품질 진단 및 개선, 피드백을 통한 오류 관리 및 활용에 대한 성과 평가 등의 관리가 요구된다. 한편, 생성되는 데이터는 무한히 수집관리하여 저장하고 보관하는 것은 불가능하다. 또한 개인정보 등의 민감함 데이터에 대해서는 일정 기한이 지나면 오히려 폐기하고 장기간 저장을 방지하는 관리가 필요하다. 따라서 기관의 데이터는 일련의 기준에 따라 저장 및 관리하고, 그 활용도가 다하면 폐기하거나 별도의 보관장소로의 이관이 필요하다. 이를 위해서는 데이터가 어떻게 생성되고 활용되는지, 더 이상 활용되지 않는 시점은 언제인지 등의 데이터 생성부터 폐기까지의 기준, 그리고 데이터의 성격에 따른 등급체계를 마련하여야 한다.

KOICA에서 주로 생성되는 데이터는 개발협력사업 수행을 통해 얻게 되는 자료, 사업수행과정의 기록, 사업결과 및 평가결과 등을 중심으로 기관 운영과정에서 발생하는 다양한 자료 및 정보라고 볼 수 있다. KOICA의 문서 중심의 업무 프로세스를 고려했을 때 비정형 데이터가 상당 부분 차지하며, 무엇보다도 이러한 데이터의 디지털화는 아직 미흡한 수준이다. 이상적인 데이터 관리를 위해서는 각 데이터의 생애주기 및 단계별 업무 프로세스를 분석하여 업무 및 데이터 관리 시스템 구축이 요구된다. 하지만, 대부분의 기관이 그렇듯, 각 업무 분야별 관리시스템은 각기 다른 시기에 개별 구축되어 결과적으로 여러 시스템에 분산 관리되고 있다.

현재 KOICA의 상당 부분의 업무영역은 시스템을 통해 관리되고 있으며, 각 업무는 다수의 시스템과 연계되어 진행되고 있다. 하지만 시스템의 연계가 제한적이고, 데이터의 디지털화 및 표준화가 미흡하여 데이터 활용에 제약이 발생한다. 이에 따라 KOICA의 핵심이기도 한 프로젝트 데이터를 중심으로 데이터 거버넌스를 구축을 위한 표준화 등에 대한 논의가 절실하다. 통합사업관리 시스템을 중심으로 관리되고 있는 프로젝트 단계별 다양한 문서 및 데이터를 살펴본 결과, 데이터는 상당하지만 효율적인 관리의 필요성을 드러냈다. 특히 수집 및 관리할 데이터에 대한 명확한 가이드라인이 미흡한 바, 각 담당자별 통일되지 않은 기준에 따라 입력 및 수집되고 있다고 볼 수 있다. 이는 활용관점에서는 가치 있는 데이터가 거의 없는 결과를 초래하게 된다. 각 단계별, 업무별 입력 데이터의 정의 및 입력방식에 대한 규칙을 수립하는 것부터, 품질 기준 수립 등 전사적으로 데이터 관리의 중요성을 인식하고 데이터의 디지털화를 위해 노력해가며 디지털환경을 구축해나가야 한다.

개발도상국이라는 특수한 환경 속에 언어 장벽 및 시스템의 한계를 극복하고 양질의 데이터로 관리하는 것은 훨씬 더 많은 노력이 요구될 것이다. KOICA의 사업을 통해 얻어지는 개발협력 데이터는 특수한 만큼, 이에 대한 체계적인 관리 또한 큰 의미가 있을 것이다. KOICA의 데이터는 내부적으로는 사업 방향성 및 전략 수립에 활용 가능하며, 대외적으로는 개발도상국으로의 접근성 확대 및 국제개발협력 분야의 다양한 정보를 공유할 수 있는 계기를 마련할 수 있다. 단기간에 기관의 데이터 거버넌스 수립은 불가능에 가깝다. 데이터 거버넌스 수립은 일회성 프로젝트가 아닌, 지속적으로 관리하며 유동적으로 개선해 나아가야 하는 프레임워크다. 또한 전사적 확대를 위해서는 충분한 자원과 인력의 투입이 반드시 뒤따라야 할 것이다. 따라서 단기적으로는 현재 KOICA의 국별협력 프로젝트 데이터의 표준화를 위한 기준을 마련하는 것에서부터 시작하여 품질 및 생애주기 관리에 대한 전사적인 합의 및 내재화가 필요하며, 장기적으로는 개발도상국에서의 디지털 인프라 구축까지도 고려되어야 할 것이다.

Executive Summary

The value of data is rapidly increasing with the 4th industrial revolution, and the spread of Covid-19 is accelerating the digitization worldwide. Accordingly, data analysis are becoming more important, and data utilization capabilities are emerging as core competitiveness of individuals and institutions. Reflecting this trend, KOICA has established a digital transformation strategy improving processes and systems, and modernizing infrastructure, for better collection, management and utilization of data. Current data management system in KOICA is somewhat insufficient compared to the diversity of data which has been collected and produced by KOICA. Data governance is essential: policies and regulations are much necessary for high-quality data management, especially for analyzing the current status and performance of development cooperation projects through data thus utilizing them in establishing business directions and strategies. Therefore, this study examines the definition and key elements of data governance, and review the current data status. Based on the analysis, it would suggest future directions so that it can be used as a basis for establishing data governance for KOICA development cooperation projects.

Due to the massive increase in an amount of data and its value, many organizations and institutions emphasize the need for data governance, yet there seems to be no established consensus on the definition of data governance. Reviewing at a number of previous studies, data governance is defined differently, but some similarities are found that it aims to collect and manage high-quality data which can be continuously used to create various values within the organization. In this perspective, this study intends to define data governance as a framework consisting strategies, policies, standards, roles, and responsibilities that are required in the entire management process from data creation to utilization.

Data as an asset is rapidly increasing in value, yet it is believed that data is still

undervalued in organizations or there is a significant difference of opinion within the organization over making an investment for systematic management. The data governance is necessary for better communication between participants in the entire data supply chain from data creation to utilization, to make good use of data while reducing risks and costs. Data is generated and collected through various channels and there are numerous constraints until it makes a good usage. Also, while data is everywhere, it is often difficult to find the exact data that is needed, and many times it requires a considerable time to process. This can be attributed to the lack of connection between the supply and demand of data. After all, data governance is a framework that serves as a link between each role in the data supply chain. Data governance manages data as a whole, thus it must be structured to conform with strategies and goals according to the nature of the organizations. And it is necessary to examine at what data is generated, who has authority, and how it is used within the organization. There is no fixed order in building data governance, but it is suggested to define suitable data governance under the given circumstances of the organization, from a small part, then gradually expanded to the whole.

The key element of data governance include standardization, quality management, and data life cycle. Data standardization is a process of organizing diversely existing data according to specific criteria, and setting common language and rules between participants, for better data utilization. Establishing standards is an essential procedure for consistent data quality management, and in the process of reorganizing existing data as well. Also, as the scope of data sharing expands, it is important to manage data properly to meet the quality and prevent misuse of them. Data quality management is a process of reviewing its accurate input according to set standards and managing to convey the meaning of the input data value to the end users. Moreover, detailed quality standards need to be established and a quality dedicated team should be organized, to handle management of considerations in data storage, quality diagnosis and improvement, and performance evaluation. Meanwhile, it is almost impossible to collect and store the generated data indefinitely. And sensitive data such as personal information needs to be discarded after a certain period of time,

and managed to prevent long-term storage. Therefore, organization's data should be managed and stored in accordance with the regulations, and discarded when its utilization is complete or transferred to a separate archive for a long-term storage. For this, it is necessary to examine data's life cycle and establish standards for each stage and a rating system according to its nature.

The data mainly generated by KOICA can be seen as various information generated throughout the implementing process of development cooperation projects, such as recordings earned from the projects including performance and evaluation results. Currently, among these data, unstructured data accounts for a large portion due to KOICA's document-oriented work process, and above all, digitization of these data is still insufficient. For ideal data management, it is required to establish an integrated work and data management system by analyzing the life cycle and step-by-step work process of each data. However, as with most institutions, management systems for each business field are individually established at different times, consequently data are distributed and managed in multiple systems.

Currently, a large part of KOICA's work area is managed through the system, and each work is carried out in connection with a number of systems. Yet, the linkage between the systems are limited, and digitization and standardization of data are insufficient, resulting in restrictions on data utilization. As a beginning point for establishing data governance, it is highly required to discuss standardization focusing on development cooperation project data, which is the KOICA's core data. Examining various documents and data at each project stage within the Integrated Project Management(IPM) system, the quantity was considerable yet revealing the need for efficient management in terms of quality control. In particular, clear guidelines on which data to be collected and managed are insufficient, generating numerous data management standards of each person in charge. When utilizing the data, this leaves only little valuable data. Thus, it is necessary to recognize the importance of data management across the organization, strive to digitize data, and build a digital environment, from establishing rules for the definition and input method of data at

each step and task, to establishing quality standards.

Especially in developing countries, overcoming language barriers and limitations of system infrastructures and managing them to quality data will require much more efforts. As the development cooperation data obtained through KOICA's projects includes unique values, systematic management of the data would enhance its great significance. KOICA's data can be used internally to establish business directions and strategies, and externally to provide an opportunity to expand accessibility to developing countries and share information in the field of international development cooperation. It is close to impossible to establish institutional data governance in a short period of time. Establishing data governance is not a one-time project, but a framework that needs to be continuously managed and improved with flexibility. In addition, sufficient financial resources and manpower must be put in place for the institution-wide. Therefore, in short term, it is necessary to establish standards for KOICA's projects, then internalize the quality and life cycle management. And in the long run, KOICA should consider even the establishment of digital infrastructure in developing countries.

목 차

국문 요약	i
Executive Summary	iv
I. 서론	1
1. 연구 배경 및 목적	1
2. 연구범위 및 방법	4
II. 데이터 거버넌스	5
1. 데이터 거버넌스란	5
1) 데이터 거버넌스의 정의	5
2) 데이터 거버넌스의 필요성	8
3) 데이터 거버넌스의 범위 및 구성요소	10
4) 데이터 거버넌스 고려사항	13
2. 데이터 거버넌스와 데이터 관리	15
1) 표준화	16
2) 품질	23
3) 데이터 생애주기 관리	30
III. KOICA의 데이터 거버넌스 진단: 전략과 관리체계, 프로젝트 데이터 중심으로	33
1. KOICA 데이터 관리 현황 및 디지털 전환	33
1) KOICA의 주요 데이터 현황	33
2) KOICA 디지털 전환 전략	35
3) 디지털 전환에 따른 데이터 관리 방향	37

2. 개발협력 프로젝트 데이터 및 수집현황	39
1) 개발협력 프로젝트 데이터	39
2) KOICA 프로젝트 데이터 수집 현황	45
3) KOICA 프로젝트 데이터 분석	49
 IV. KOICA 데이터 거버넌스 구축을 위한 방향성 제안 및 결론	50
1. KOICA 프로젝트 단계별 핵심 문서 및 데이터 제안	50
2. 향후 과제	53
 참고문헌	56
 부록	
부록 1. IATI 공개항목 및 DAC 보고항목	60
부록 2. 데이터 관리체계 구축방안 연구: KOICA 개발협력 프로젝트 데이터 중심으로 보이스아이코드	63

표 목차

〈표 II-1〉 데이터 거버넌스 정의	6
〈표 II-2〉 데이터 거버넌스 구조화	11
〈표 II-3〉 데이터 표준화 절차	18
〈표 II-4〉 데이터 표준화 수준	20
〈표 II-5〉 데이터 품질 진단 수행 조직구성 및 역할과 책임	26
〈표 II-6〉 세부 품질기준 설명 및 활용 예시	27
〈표 II-7〉 데이터 품질지표	29
〈표 II-8〉 KOICA 단위과제 보존기간표(2021.2.기준)	31
〈표 III-1〉 부서 간 생성 및 관리 데이터	35
〈표 III-2〉 KOICA 프로젝트 추진 절차	39
〈표 III-3〉 프로젝트 데이터 종류	41
〈표 III-4〉 PDM의 구조와 내용	43
〈표 III-5〉 간이 M&E 매트릭스(예)	44
〈표 III-6〉 USAID 성과지표 워크시트	44
〈표 부록-1〉 IATI 공개항목	60
〈표 부록-2〉 DAC 보고항목	61

그림 목차

〈그림 Ⅱ-1〉 데이터 거버넌스 정의	7
〈그림 Ⅱ-2〉 데이터 거버넌스 프레임워크	11
〈그림 Ⅱ-3〉 데이터 관리 프레임워크	11
〈그림 Ⅱ-4〉 데이터 거버넌스 프레임워크 예시	13
〈그림 Ⅱ-5〉 데이터 표준화 의미	16
〈그림 Ⅱ-6〉 표준 단어의 도출	19
〈그림 Ⅱ-7〉 표준용어 및 표준단어 정의서 예시	21
〈그림 Ⅱ-8〉 기관의 데이터 품질관리	24
〈그림 Ⅲ-1〉 디지털 전환 전략 체계도	37

I. 서론

1. 연구 배경 및 목적

4차 산업혁명 시대에 데이터는 그 가치가 빠르게 증대되면서, 이를 제대로 활용하기 위한 다양한 노력이 이어지고 있다. 특히, 전 세계를 강타한 코로나바이러스 감염증은 인간의 물리적인 이동을 제한하며, 역설적이게도 디지털화를 가속했다. 이러한 사회적 변화에 발맞춰 정부에서도 디지털 전환을 강조하며 디지털 뉴딜을 선포하였다. 민간에서는 급변하는 현대사회에 대응하기 위해 업무 효율성을 강조하며, 데이터가 혁신성장을 견인하는 필수 자본으로 부상하고 있다. 이에 데이터 분석 및 활용의 중요성이 증대되며 데이터 활용 역량은 개인 및 기관의 핵심 경쟁력으로 대두되고 있다.

민간에서의 데이터 활용은 곧 기업의 이익 창출과도 연관된 만큼, 이미 많은 투자와 관리 노력을 투입하고 있다. 이에 정부는 공공부문에서 창출되는 데이터를 활용하여 민간에서 활성화될 수 있도록 공공데이터 포털¹⁾을 개설하고, 현재 25개 분야 공공데이터 개방을 본격화하며 데이터 산업의 성장을 이끌어가고자 한다. 한편, KOICA는 공공기관으로 데이터를 제공할 의무가 부여되고 있으며, 이러한 정부의 주요 정책 방향성에 부응하고 향후 개발협력 사업의 효과성 확대를 위해 2021년 디지털 전환 전략 ‘21-’25를 수립하고 디지털 중심의 경영기반 혁신을 이끌어가고자 한다. 즉, 데이터 수집·관리·활용이 가능하도록 프로세스 및 시스템 개선, 인프라 최신화로 디지털 KOICA로 도약하고자 한다.

KOICA는 공공데이터 개방 외에도 데이터를 활용하여 기관의 가치를 확대하고 보다 효율적인 기관 및 사업 운영을 위해 협력단 내외의 디지털 전환 및 체계적인 데이터 관리 요구가 지속적으로 이어질 것으로 예상된다. 디지털 전환의 핵심은 데이터의 효과적 활용에 있을 것이다. 따라서 현재 KOICA의 상황을 고려할 때 데이터 표준화 및 거버넌스 정립이 우선되어야 할 것이다. 또한 데이터를 안전하게 관리하는 데 있어 국내 데이터 3법²⁾을 참작하여 보호되어야 하는 데이터는 명확히 하고, 그 외 합리적인 수준에서 개방하여 혁신에 활용함으로써 공공기관에 요구되는

1) 2019년 빅데이터 플랫폼 구축하여 1,400여종 데이터 전면 개방

2) 개인정보보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률(정보통신망법), 신용정보의 이용 및 보호에 관한 법률(신용정보법) 등 3가지 법률을 통칭

정부의 디지털 뉴딜 정책 방향³⁾에도 부합해야 할 것이다.

체계적인 데이터 관리의 필요성은 꾸준히 제기되고 있으나, 데이터 관리를 위한 인력 및 예산 배정에 있어서는 상대적으로 소극적이며 데이터 투자 타당성을 인정받기 쉽지 않은 경우가 많다. 데이터는 기관의 중요 자산으로 이를 활용하여 경쟁력 향상 및 추가적인 가치 창출을 한다고 강조하고는 있으나, 실질적으로는 데이터 자산에 대한 보유 현황 및 자산의 가치를 정확하게 파악하지 못하고 있는 셈이다. 정보 및 데이터의 중요성이 증대되고 있으며, 이러한 자원의 올바른 관리가 개인, 기관 및 기업의 가치를 결정하는 주요한 요인이 되고 있다. 실질적으로 양질의 데이터에 기반하여 추가적인 가치를 창출하기까지는 데이터 품질 관리, 데이터 표준화 및 통폐합, 메타데이터 및 마스터데이터 관리 등 상호 보완적이고 복합적인 과정들이 요구되며, 이에 상황에 적합한 의사결정을 내리고 효율성을 창출하기 위한 기반으로써 체계적인 데이터 거버넌스 구축의 필요성이 강조되고 있다.

기관에서 생성되는 데이터를 파악하여 하나의 시스템을 통해 관리하기 위해서는 관련 규정, 특히 표준화 지침 수립에 많은 노력이 요구된다. 전사 동일한 기준 및 규정으로 관리되는 신뢰성 있는 데이터는 결과적으로 사업 현황, 사업 성과 및 효과성을 분석하고 향후 방향성 제시 등의 데이터 기반 의사결정이 가능해지도록 한다. 데이터 거버넌스는 데이터 표준화 및 품질관리 등의 관리체계를 아우르는 전체 프로세스로, 데이터 관리의 전사적인 목표와 방향 정립을 가능하게 한다. 데이터의 가치를 최대한으로 활용하기 위한 전사 차원의 경영 체계라고도 할 수 있을 것이다. IT 및 데이터 중요도가 상승하면서 데이터 저장 및 호환 등 관련 관리 기술의 우수성이 강조되어 왔으나, 데이터를 활용하기 위해서는 기술 외에도 표준화에 기반한 양질의 데이터 품질 유지도 못지않게 중요하다. 이를 위해서는 데이터 거버넌스를 인지하고 관련 지침 및 표준을 만드는 것은 물론, 책임 있는 조직 구성까지도 함께 이루어져야 한다.

현재 KOICA에서는 다양한 데이터가 생산되고 있으며 크게 경영관리와 사업관리 데이터로 구분이 가능하다. 다만, 그동안 수집되어 온 데이터는 각각의 시스템 하에 중복 관리되고 있으며 이마저도 노후화되어 실질적으로 활용하는데 상당한 제약이 있는 상황이다. 효율적인 데이터 관리 체계 부재로 활용도 측면에서 효율성이 다소 떨어진다고 할 수 있다. 또한, 사업 이행에 따른 개발도상국 내 개발효과성이 어떻게 발현되고 있는지 판단할 필요가 있으나, 이를 측정하는데 필요한 양질의 데이터는 충분하지 못한 경우가 많다. 사업의 특성에 따라 개발효과성 분석에

3) 데이터를 거래하는 산업 확대될 예정인 가운데, 공공기관의 데이터를 적극 개방함으로써 공공성 확보 및 중소기업의 데이터를 원천으로 하는 다양한 산업 활성화를 도모하고자 함

필요한 데이터 및 수집 방법이 다를 것이기에, 적용 가능한 방법론 및 단계별 데이터 수집 프로세스에 대한 기관 내 인지가 필요한 상황이다. KOICA는 개발협력사업 집행기관으로 사업의 원활한 수행도 중요하지만, 데이터에 기반하여 개발효과성을 분석하고 이를 향후 사업 방향성 및 전략 수립 등에 적극 활용할 필요가 있다. 사업 효과성 분석 경험의 축적은 사업기획 및 수행에 있어 KOICA의 강점을 발견하고 이를 극대화할 수 있는 중요한 발판이 될 것이다.

데이터 중요도가 급상승하는 가운데, 데이터는 기관의 자산이며 조직 운영의 기반이 되어야 한다는 인식의 확산이 우선 이루어져야 한다. 기관에서의 적극적인 데이터 활용을 위해서는 데이터 관리 전략 및 프로세스, 즉 데이터 거버넌스가 수립되어야 한다. 데이터의 생산자, 처리자, 사용자는 모두 다르다. 많은 경우에서 데이터 사용자는 어떤 데이터가 어떻게 생산되고 어떤 기준으로 처리되었는지 알기 어려우며, 값과 명칭만 있는 데이터를 보고 컨텍스트까지 파악하기엔 한계가 있을 것이다. 체계적인 데이터 축적과 활용, 나아가 가치를 창출하는 데는 데이터 거버넌스 수립이 우선적으로 필요하다. 데이터 거버넌스를 통해 데이터 사용 목적에 따른 데이터의 개념과 분류 기준 및 가공 방법 등의 기준을 마련함으로써, 데이터의 원래 의미와 다르게 활용되는 것을 방지할 필요가 있다. 또한 데이터의 흐름에 따른 데이터 관리 인원 및 책임권한을 명확히 함으로써 양질의 데이터가 체계적으로 관리될 수 있도록 할 필요가 있다. 이는 결과적으로 성과 창출과도 직접적으로 연결될 것이다.

이에, 본 연구에서는 데이터 거버넌스의 정의 및 핵심 요소를 살펴보고, 나아가 KOICA 개발협력 사업의 데이터 거버넌스 수립을 위한 근거자료로 활용될 수 있도록 데이터 현황을 분석하고 향후 방향성 제시하는 것을 주요 목적으로 하고자 한다.

2. 연구범위 및 방법

먼저 본 연구에서는 데이터 거버넌스란 무엇인지 살펴보고, 데이터 표준화, 품질 및 생애주기 관리를 위한 이론적 배경을 조사하고자 한다. 또한 현재 KOICA의 현황을 분석하고, 디지털 전환 전략 수립에 따른 방향성을 논의해보고자 한다. 그 다음 KOICA의 개발협력사업 사례를 중심으로 사업 데이터 종류 및 관리 현황에 대해 살펴보고, 나아가 데이터 거버넌스 수립 방향을 고민해보고자 한다. 기관에서 생산되고 있는 다양한 데이터를 관리하기 위한 거버넌스는 무엇인지, 그리고 관련 기준 마련 등에 어떠한 노력이 필요한지 살펴보고, 나아가 KOICA의 중요 자산인 데이터 관리를 위한 적합한 거버넌스를 구축하는데 활용할 수 있는 기초 자료로써 시사점을 제시하고자 한다.

이를 위해, 국내외 관련 문헌자료, KOICA 및 타 원조 기관의 공식문서를 통한 문헌조사, 그리고 협력단 내 관계자 인터뷰를 주요 연구 방법으로 활용하고자 한다.

Ⅱ. 데이터 거버넌스

1. 데이터 거버넌스란

급격하게 늘어나고 있는 데이터로 인해, 그리고 그 데이터를 제대로 활용하기 위해, 나아가 데이터를 통해 추가적인 가치 창출을 위해 많은 기관에서는 데이터 거버넌스를 수립하고, 이를 기반으로 디지털 전환을 모색하고 있다. 데이터 거버넌스의 필요성에 대해서는 강조되고 있으나, 데이터 거버넌스에 대한 합의된 공통의 정의는 없는 것으로 보인다. 거버넌스의 사전적 의미는 공동의 목표를 달성하기 위하여 주어진 자원의 제약 하에서 모든 이해 당사자들이 책임감을 가지고 투명하게 의사결정을 수행할 수 있게 하는 제반 장치이다⁴⁾. 여기에 데이터를 덧붙이면, 기관의 목표 달성을 위해 데이터 기반 의사결정을 할 수 있도록 하는 데이터 관리 체계라고 짐작해볼 수 있다. 다수의 선행연구에서 데이터 거버넌스에 대한 정의는 기관별·연구자별로 상이하게 나타났으나, 일반적으로 데이터 거버넌스는 양질의 데이터를 수집하고 관리하여 조직에서의 다양한 가치 창출에 지속적으로 활용되도록 하는 데 그 목적을 두고 있다는 점에서 유사하다.

1) 데이터 거버넌스의 정의

우선 데이터 솔루션 업체들을 중심으로 데이터 거버넌스의 의미를 살펴보면, IBM(2007)은 데이터의 관리, 개선 조직화를 통한 데이터 품질 관리로 정의하고 있다. SAS(2018)는 데이터 전반의 활동에 대한 관리 측면에서 기업의 데이터에 대한 전략, 목표, 정책 수립을 위한 조직 차원의 프레임워크로 정의하며, 이를 전략과 실행의 조합이라 하였다. 또한 비영리 글로벌 협회인 DAMA International(2017)은 데이터 자산에 대한 조직 차원의 의사결정과 책임으로 정의하며, 데이터 아키텍처, 데이터 품질, 메타데이터, 데이터 보안 등의 핵심으로 언급하고 있다. 즉, 데이터 거버넌스는 관련 정책과 우수사례를 기반으로 데이터를 적절하게 관리하는 것(Ladely 2012)을 의미한다.

여러 선행연구에서도 유사한 정의를 제시하고 있다. Abraham et al(2019)는 데이터 관리에

4) 국립국어원 우리말샘 사전

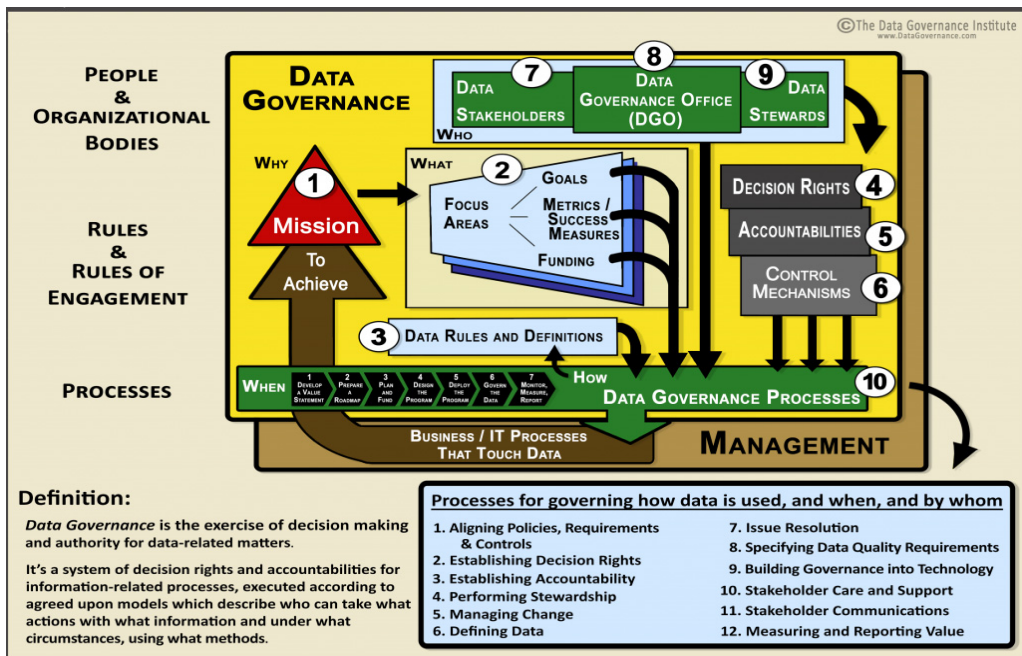
대한 권한과 통제권을 행사하는 것으로 정의한 바 있으며, 이후 연구에서는 데이터를 전략적 기업자산으로 관리하기 위한 프레임워크로 정의하였다. 또한 장경애 · 김우재(2016)는 조직이 데이터를 정확히 활용할 수 있도록 데이터를 관리하고 통제하고 조직화하며 각 구성원에게 역할과 책임을 부여하는 등의 활동을 통해서 데이터 품질을 향상시키는 힘으로 정의하였으며, 최영훈 외(2014)는 데이터를 활용하기 위한 조직으로 개편하고 구성원 간의 협조를 통해 데이터 활용 상의 문제를 해결해 나아가는 체계로 정의하였다. 이 외에도 Power(2011)은 전략적 로드맵을 구축하는 비즈니스 중심 프로그램으로, 사람, 처리 과정, 기술 및 정보의 4가지 측면 모두를 고려해야 한다고 정의한 바 있으며, 같은 맥락에서 Soares(2010)도 기업이 데이터를 전사적 자산으로 활용할 수 있도록 인력, 프로세스 및 기술을 조율하는 활동으로 정의하였다. 부창완(2020)은 데이터 표준 및 정책에 따라 비즈니스 데이터를 생성 · 변경하고 생성된 데이터의 가용성 · 유용성 · 무결성과 보안을 관리하는 프로세스로 정의하였다.

〈표 II-1〉 데이터 거버넌스 정의

저자	데이터 거버넌스 정의
Abraham et al. (2019)	데이터 관리에 대한 권한과 통제권 행사/ 데이터의 가치를 높이고 데이터 관련 비용 및 위험을 최소화하는 것이 목표임. 데이터를 전략적 기업자산으로 관리하기 위한 프레임워크. 데이터 관련, 조직의 의사결정에 대한 권한과 책임. 데이터 거버넌스는 데이터 정책, 표준 및 절차를 공식화하고 규정 준수를 모니터링함.
Weber et al. (2009)	데이터 거버넌스는 데이터 사용 시 바람직한 행동을 장려하는 의사결정권 및 책임
Khatri (2010)	데이터 거버넌스는 조직의 데이터 자산에 대한 의사결정에 책임을 지는 것
Merkus et el. (2019)	기업의 목적에 부합하는 데이터의 관리 및 결정권
김석수 (2011)	기업 데이터의 효과적 정의, 경영, 공유를 위해 정책, 전략, 표준화, 프로세스, 가이드라인, 프레임워크, 기술 등을 확립하여 운영하는 제반활동
장경애 김우재 (2016)	데이터의 전반적인 활동을 관리하고 통제하여 고품질 데이터 확보, 이를 통해 데이터 활용을 극대화하고 가치창출에 기여하는 전반적인 활동
조완섭 (2017)	데이터에 대해 관리 정책, 지침, 표준, 전략 및 방향을 수립하고, 데이터를 관리할 수 있는 조직 및 서비스를 구축하는 관리체계
최영훈 외 (2014)	데이터를 활용하기 위한 조직을 관리하고 구성원 간의 협조를 통해 데이터 활용상의 문제를 해결하는 체계
Otto (2011)	의사결정권의 할당과 전사적 측면의 데이터관리에 관한 업무
O'Neal (2015)	정보기술 자원과 조직을 기업의 경영 전략과 목표에 연계하기 위한 의사결정 및 책임에 관한 프레임워크
SAS (2018)	기업 데이터에 대한 전략, 목표, 정책 수립을 위한 조직차원의 프레임워크
Laudon and Laudon (2018)	데이터의 가용성, 유용성, 통합성, 보안성을 관리하기 위한 정책과 프로세스를 다루며 프라이버시, 보안성, 데이터품질, 관리규정 준수
DGI (웹사이트)	데이터 관련된 문제 해결을 위한 권한과 의사결정의 활동

위 다양한 정의들을 종합하여 보면, 데이터 거버넌스는 그동안 각각 발전해 온 사업운영과 정보통신기술을 조정 및 통합하기 위한 접근이라고 볼 수 있겠다. 데이터를 기관의 중요 자산으로 인식하고 관리 및 활용하기 위한 원칙으로, 인력 및 기술 등을 통해 데이터에서 최적의 가치를 창출하고자 함이다. 즉, 데이터 거버넌스는 데이터의 생성부터 활용까지의 전체 관리 과정에서 필요한 전략, 정책, 기준, 역할, 책임 등에 관한 체계로, 특히 데이터 표준화, 통제합, 보안 및 저장과 관련하여 역할 및 책임 소재를 명확히 하는 것이 핵심적이라 할 수 있다.

〈그림 II-1〉 데이터 거버넌스 정의



자료: Data Governance Institute 웹사이트

한편, 데이터 거버넌스와 유사하게 IT 거버넌스, 정보 관리(Information Management) 등의 개념도 있다. 이들 개념은 각 기관별, 문헌별로 조금씩 다르게 정의하고 있으며 일부 혼용되어 사용되고 있는 것으로 나타난다. 일반적으로 IT 거버넌스는 IT 자원과 정보, 조직을 기업의 경영 전략 및 목표와 연계해 경쟁 우위를 확보할 수 있도록 하는 의사결정 및 책임에 관한 프레임워크로 정의(한국정보통신기술협회 정보통신용어사전)되고 있으며, 국제정보시스템감사통제협회 (ISACA)⁵⁾에서는 IT자산의 효율적 관리를 위해 전략과의 연계되는 관리 프레임워크로 제시하고

5) ISACA는 IT 감사, 리스크, 거버넌스, 보안 분야 전문가들을 대상으로 한 글로벌 협회로 1996년 IT 환경의 성장을 원활하게 조정하기 위해 COBIT(Control Objectives for Information Technologies,

있다. 따라서 데이터 거버넌스는 데이터 관리 관점에서의 IT체계로 데이터 거버넌스가 IT 거버넌스의 일부로 보여질 수도 있다. DAMA International(2017)에 의하면 IT 거버넌스는 상대적으로 기술적인 부분에 초점을 둔 반면 데이터 거버넌스는 데이터의 실질적인 활용에 초점으로 두고 있다. 정보 관리는 기관에서 생성되고 있는 정보에 대한 관리를 하기 위한 시스템을 일컫는다. 데이터 거버넌스 하에 이러한 기술적인 접근이 이루어지는 것이 이론적으로 맞겠으나, 현실적으로는 기술발전에 따른 시스템이 먼저 구축되고, 이를 효과적으로 활용하기 위해 전사적으로 거버넌스 수립하는 경우가 많다고 한다. 본 연구에서는 데이터 거버넌스를 기술적 측면이 아닌 데이터 활용을 위한 관리 측면에서의 정의에 초점을 두고자 한다.

2) 데이터 거버넌스의 필요성

IT 산업 발전은 디지털 기술 장비의 고도화로 이어지며, 이를 통한 데이터 활용 능력의 중요성을 강조하고 있다. 다만, 기술은 상당히 고도화되고 있는 가운데, 실제 데이터의 활용 능력은 기관별로 상당한 편차를 보인다. 이는 데이터의 자산으로서의 가치에 대해 아직은 저평가하고 있거나, 혹은 가치를 인식하고 있더라도 실제로 이를 확대 관리하기 위한 투자로 이어지기까지 조직문화에 상당한 차이가 있기 때문인 것으로 사료된다. 특히 최근 발생하고 있는 데이터 유출에 따른 피해 사건은 많은 기관에 데이터 보안을 강화하고 주의할 것을 강조하고 있으나, 데이터를 활용하는 것에 대해서는 그 중요성 및 필요성에 대해 상대적으로 인지가 낮은 것으로 보인다. 기술 장비의 고도화 및 보안 강화도 중요하지만, 양질의 데이터를 관리할 수 있는 거버넌스를 수립하는 것 또한 매우 중요하다. 무한히 쏟아지고 있는 원자료 또는 미가공 데이터를 기관의 가치 창출 활동에 얼마나 효과적으로 활용할 수 있는지는 기관의 경쟁력과도 직결되기 때문이다. 데이터 관리에 필요한 정책, 기준, 역할, 책임 등에 관한 체계인 데이터 거버넌스를 수립하는 것은 기관의 데이터 활용에 대한 강력한 의지를 보여주는 것과 동시에 원칙을 통한 명확한 방향성을 제시하는 것과도 같다.

데이터 거버넌스 수립의 가장 근본적인 목적은 데이터를 더욱 수월하게 활용할 수 있게 하는 동시에 위험과 비용을 줄이기 위함이다. 메타데이터⁶⁾ 관리를 통해 기관 내의 데이터를 정의하고

정보기술 관리 목적) 발표. IT관리 프레임워크로 기업의 정보관리 및 거버넌스 관련된 전략 개발하고 조직하고 이행하는데 활용되고 있으며, 최신 버전은 COBIT 2019임.

6) 메타데이터란 “데이터를 위한 데이터”, 혹은 “데이터를 설명하기 위한 데이터”라고 할 수 있는데, 이는 메타데이터를 통해 원자료(raw data)의 여러 특징들을 기술함으로써 이후 그 자료를 다시 이용하는데 도움을 주고자 함이다(통계연구분석 제2권 2호 '97. 가을호).

규정을 준수하여 수집되는 데이터의 품질을 일정하게 관리하는 것은, 기관에서의 데이터 기반 의사결정을 위해 필요한 가장 기본적인 데이터 관리일 것이다. 이러한 작업은 데이터의 활용을 수월하게 함은 물론이려니와, 데이터의 잘못된 사용으로 인한 오류를 줄일 수 있다. 또한 최근 지속해서 쟁점이 되는 데이터 보안과 개인정보 보호 차원에서 사고나 고의적으로 자료가 유출되는 것을 방지하는 관련 규정과 절차가 필요하다. 데이터를 식별하여 저장하고, 나날이 늘어나고 있는 다양한 법과 규제를 즉각적으로 반영하는 데는 시간과 비용이 소요된다. 기관에서 수집되는 데이터를 파악하여 불필요한 데이터는 수집관리 대상에서 제외하고, 필요한 데이터에 대해서는 저장 및 활용에 관련 절차 및 규정을 마련하고 적절한 책임과 권한을 부여함으로써 개인정보의 유출 및 천문학적인 피해·비용 발생 가능성을 미연에 방지할 수 있다. 데이터 활용을 확대하는 것과 위험과 비용을 줄이는 것은 다소 상반되지만, 거버넌스 수립을 통해 전사 차원의 일관된 방법과 절차에 따라 데이터를 관리함으로써 위험 요소로부터 대비하고 피해 비용을 최소화하면서도 안전하고 효과적인 데이터 활용을 지원하고자 함이다.

한편, 수많은 데이터를 쉽게 접하고 있으면서도 정작 필요한 데이터는 찾기 어렵거나 가공하는데 상당한 시간이 소요되는 경우가 많다. 일례로 우버(Uber)는 지속적으로 생성되는 방대한 데이터 및 분산된 시스템으로 인해 사용가능한 정보에 대한 가시성 확보의 필요성을 느꼈으며, 데이터 활용에 드는 시간을 줄이기 위해 데이터북(Databook)을 만들었다고 한다⁷⁾. 이는 데이터의 공급과 수요(혹은 생성과 활용)가 원활하게 연결되어 있지 않기 때문이라고 볼 수 있다. 데이터의 생성부터 활용까지 다양한 역할자가 있으며, 일반적으로 데이터 생성자, 저장자 혹은 관리자, 그리고 활용자 등으로 나누어 볼 수 있다. 이들은 각각 다른 주체이며, 가지고 있는 권한이나 책임도 다르다. 대부분의 경우, 데이터 활용자는 언제 어디에서 어떠한 데이터가 어떻게 생성되고 있는지, 어디에 어떻게 저장되고 있는지, 어떤 방식으로 관리되고 있는지 모른다. 또한 데이터 관리자도 언제 어디에서 어떠한 데이터가 어떻게 생성되는지, 언제 어디에서 어떠한 데이터가 어떤 활용 목적으로 사용되는지 모른다. 이는 데이터 생성자도 마찬가지이다. 데이터가 어떻게 저장되고 관리되는지, 누가 어떤 목적으로 무슨 데이터를 활용하는지 알 수 없다. 데이터 생애주기를 통틀어 활용자가 직접 생성하고 저장하고 관리하여 활용하는 경우는 드물 것이다. 그리고 데이터 생성자의 목적과 다르게 분류되어 저장 및 관리되거나 활용될 가능성도 얼마든지 있다. 따라서 데이터의 활용 가능성을 높이기 위해서는 이들 간의 연결고리 역할을 하는 공통의 언어 또는 규칙을 설정하는 것이 필요하다. 즉, 데이터

7) 우버 웹사이트(<https://eng.uber.com/databook>)

거버넌스는 어디에서 어떤 데이터가 어떻게 생성되어 어디에서 활용되는지 전체를 아우르는 데이터 공급사슬 역할을 담당하는 것으로 볼 수 있다.

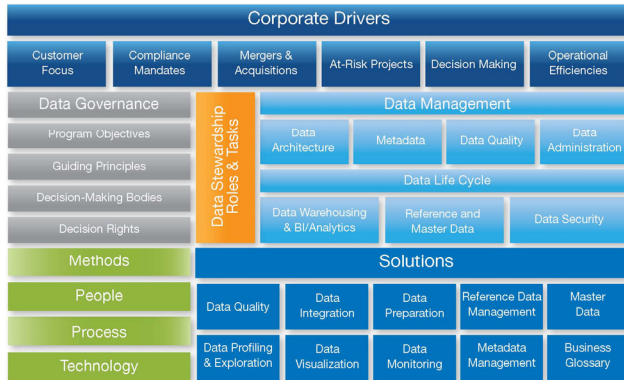
3) 데이터 거버넌스의 범위 및 구성요소

데이터 거버넌스 수립을 위해서는 데이터 자체에 대한 이해가 필요하며, 누가 접근 가능하였는지, 누구에게 접근 가능성을 제시해야 하는지 등, 기업이 보유하고 있는 데이터를 분류하고 접근권한에 대해 식별하는 작업이 필요하다. 결국, 이러한 기준 및 역할과 책임은 규정으로 명시되기 때문에 데이터 거버넌스는 통제 혹은 규정 준수(compliance)라고도 할 수 있다. 데이터는 품질 및 신뢰도에 따라 그 가치가 달라지기에 일정한 기준에 따른 신뢰성 있는 데이터를 활용 및 제공하기 위해서는 관리 규정 및 지침이 필요하고 정해진 일련의 과정들을 통해 체계적으로 수집 및 관리하는 것이 핵심이다.

데이터 거버넌스를 구성하고 있는 요소는 다수의 연구에서 조금씩 다르게 제시하고 있으며, 실제 거버넌스 구성 요소는 기관의 목적과 구조에 따라 다르게 설정된다. DAMA International(2017)은 전략, 정책, 기준과 품질, 관리 감독, 규정 준수, 이슈 관리, 데이터 관리 개선, 데이터 자원 가치 평가 등을, 김석수(2011)는 규정준수, 합병과 흡수, 비즈니스 프로세스 아웃소싱, 기업 정보 경영, BI(Business Intelligence)와 운영적인 분석, 데이터 보안 및 프라이버시를 데이터 거버넌스의 주요 요소로 언급하였다. SAS(2014)는 데이터 거버넌스를 위해 이용되는 기술 요소도 포함하기도 하였는데, 이는 기술 및 도구의 통합성 혹은 일관성이 데이터 거버넌스의 효율성을 증대시킬 수 있기 때문이다.

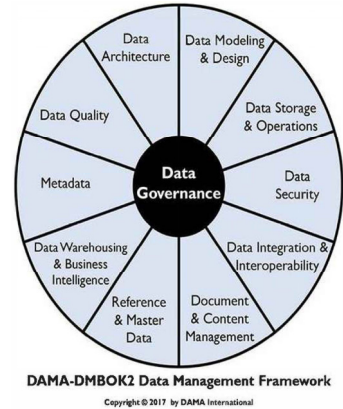
이들 중 공통적으로 언급되고 있는 요소로는 정책, 전략, 규정(표준, 가이드라인 등), 조직 구조 등이 있다. 정책 및 전략은 전사 차원에서 공통되게 방향성을 제시한다. 이는 원칙 및 규정과도 직결되기에 명확하지 않을 경우, 데이터 관리체계 및 최종적으로 제시되는 데이터의 품질에도 부정적인 영향을 미칠 수 있다. 규정, 표준, 가이드라인 등은 데이터 관리를 구조화하고 데이터 품질을 보장하기 위함이며, 데이터 처리 과정에서 데이터가 내포하고 있는 의미를 올바르게 전달하기 위해 반드시 지켜져야 하는 기준이다. 특히, 기관 내 투명성 증진 및 민감한 정보의 보안을 위해, 엄격한 프로세스 정책의 확립은 필수적이다. 조직 구조는 데이터의 생성부터 저장, 가공 및 활용까지의 전 과정에서의 역할과 책임, 그리고 각자의 역할과 책임에 따른 권한 부여도 함께 정의한다.

〈그림 II-2〉 데이터 거버넌스 프레임워크



자료 : SAS 2018

〈그림 II-3〉 데이터 관리 프레임워크



자료 : DAMA International 2017

장경애 · 김우제(2016)는 거버넌스 구성 요소를 분석하여 구조화하였다.

〈표 II-2〉 데이터 거버넌스 구조화

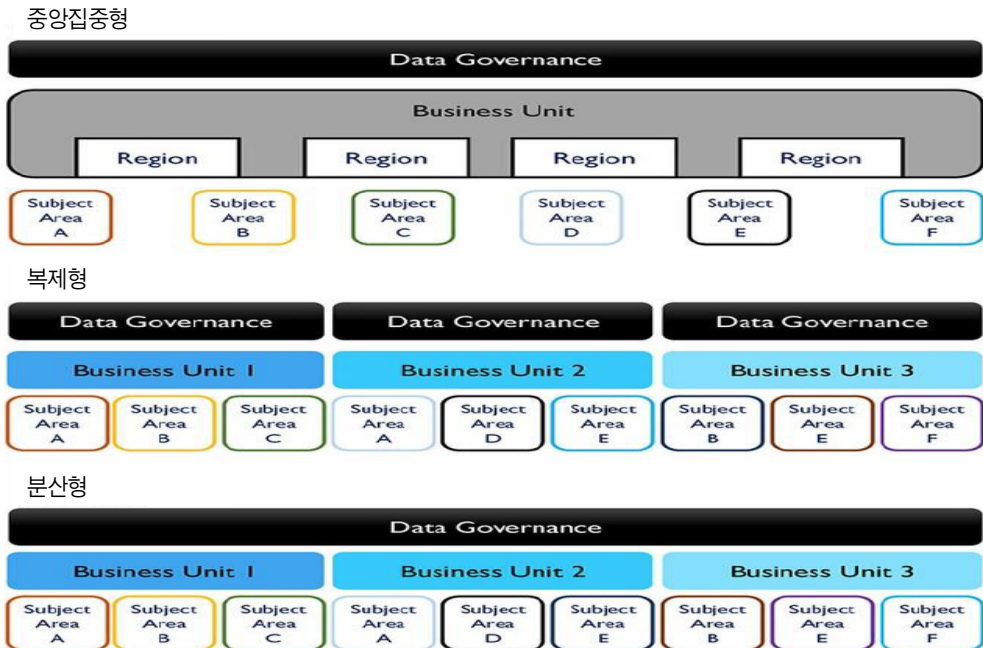
구성요소		가중치 (%)	정의
1레벨	2레벨		
Data Compliance (데이터 통제: 37.76)	Strategy & Mission	7.81	전사적 측면의 데이터 체계화를 위한 전략 및 비전 수립
	Data Quality (데이터 품질:48.56)	7.76	데이터 체계화를 위한 비전 달성을 위한 방침
	Data Organization (데이터 조직: 13.68)	7.31	전략, 비전, 정책 수립의 방식이나 수단
	Metrics	7.53	목표에 부합되는지 평가를 위한 지표
	System & Tools	7.36	시스템 및 데이터 아키텍처, 시스템, 솔루션/자동화 도구
Data Quality (데이터 품질: 48.56)	Data lifecycle	7.95	데이터 자원에 대한 수집/축적/가공/활용/폐기 등 전반적인 생명주기 관리
	Data Management	8.50	데이터 관리 업무 프로세스, 업무규칙 정의 등의 활동
	Monitoring	8.21	데이터 로그, 감사 및 보고 활동
	Quality Management	8.33	데이터 품질을 향상시키기 위한 일련의 관리 활동
	Quality Value	7.75	데이터 값의 품질을 향상시키기 위한 일련의 활동
	Data Security	7.82	데이터 보안을 유지하기 위한 일련의 활동
Data Organization (데이터 조직: 13.68)	R&R	6.99	데이터 체계화를 위한 인력의 역할과 책임 정의 및 활동
	Organizational Structures	6.69	데이터 체계화를 위한 조직구성 활동

자료: 장경애 · 김우제 2016

분석 결과 데이터 통제 관리, 데이터 품질 관리, 데이터 조직 관리의 3개의 최상위 구성 요소와 13개의 하부 구성 요소를 도출하였다. 구성 요소 간의 영향도 측면에서는 데이터 통제 관리 하의 5가지 요소는 원인 제공자 속성으로, 나머지 데이터 품질 및 조직 관리 하의 8가지 요소는 결과자의 속성으로 나타났다. 중심성 분석 결과를 살펴보면, 하부 구성 요소에서는 데이터 관리(data management), 품질관리(quality management), 모니터링(monitring), 최상위 구성 요소에서는 데이터 품질(data quality), 데이터 통제(data compliance), 데이터 조직(data organization) 순으로 높게 나타났다. 데이터 거버넌스는 데이터의 통제 관리를 위한 원인 제공자부터 데이터 품질이라는 결과자, 그리고 이를 위한 조직의 역할 및 책임 등으로 구성되어 있음을 제시하였다. 이러한 분석 결과는 데이터 거버넌스의 핵심은 데이터 품질관리에 있다는 점을 강조하고 있다.

한편, 데이터 거버넌스 운영 체계는 기관의 조직 운영 방식에 따라 다르게 나타날 수 있으며, 크게 중앙집중형, 복제형, 분산형 등, [그림 II-4]와 같이 구분 가능하다. 거버넌스 운영 체계는 기관 내 데이터 관리 책임 구조에 따라 달라진다. 단독 시스템 하에 데이터 관리가 이루어진다면, 거버넌스 역시 중앙집중형의 운영 체계를 갖게 될 것이며, 다수의 데이터 시스템이 운영되고 있다면 복제형 또는 분산형이 적합할 것이다. 가장 이상적인 모델은 중앙집중형이다. 중앙집중형은 기관 전체에 대한 통합적인 데이터 거버넌스를 수립하여 관리하는 것이다. 중앙집중형 운영 체계는 공식적인 경영진 지위를 확립함으로써, 데이터 거버넌스에 대한 기관 내 데이터에 대한 인식개선도 가능하다. 다만, 중앙집중형 운영 체계만이 최적은 아니다. 중앙집중형은 조직 구조의 많은 변화도 요구하기에 그만큼의 위험도 증가한다. 또한 기관의 구조적 특성과 다루는 데이터의 특성을 고려하여 복제형, 분산형, 하이브리드형 등 합리적인 체계 수립이 필요하다. 사업 분야가 다르거나 서로 간의 접점이 없는 데이터를 별도로 관리해야할 경우, 통합된 데이터 거버넌스 하에 사업 분야별 구분 또는 각각의 개별 거버넌스를 수립하는 등의 방안이 적합할 수도 있기 때문이다.

〈그림 Ⅱ-4〉 데이터 거버넌스 프레임워크 예시



자료: DAMA International 2017

4) 데이터 거버넌스 고려사항

EIU(2018) 설문조사에 의하면 데이터 거버넌스에 대한 전반적인 접근 방식은 주로 법률 및 규정 준수 고려사항(55%)에 의해 주도되거나, 개인정보 보호 문제(44%) 및 기술적인 관점에서 데이터를 관리해야 하는 필요성(44%)에 의해 주도되고 있다고 답하였으며, 실무부서에서 업무에의 활용(33%) 및 보안(22%)에는 다소 낮은 비율의 응답률을 보였다. 데이터 거버넌스 접근 방식의 성공을 측정하는 기준은 데이터의 정확성(accuracy of data 51%), 규정 준수(compliance 49%) 및 전략적 목표에 대한 기여(contribution to strategic goals 40%)라고 답하였다. 주로 규정 준수 및 통제 방식의 접근으로 데이터를 자산으로 인식하고 기관의 가치 창출에 활용하고자 하며, 데이터 거버넌스의 결과물은 결국 데이터의 정확성에 가장 무게를 두고 있는 것이다.

데이터 거버넌스 수립을 위해서는 다음의 사항을 고려해야 할 것이다. 데이터 거버넌스는 데이터의 생애주기를 전체적으로 관리하는 하나의 운영 체계인 바, 기관의 성격에 따라 사업전략과 목표에 부합하도록 구성되어야 한다. 일괄적으로 모든 기관에 동일하게 적용 가능한

체계가 아니다. 데이터 거버넌스는 데이터 기반 의사결정에 공동의 책임을 부여하는 것과도 같다. 데이터 생성부터 저장·관리, 활용까지 조직 전체 및 시스템을 통해 의사결정에 도달하게 되는 책임이 있기 때문이다. 데이터 거버넌스 수립은 단순히 데이터를 관리하는 차원의 체계 수립이라기보다는 기관 전체의 책임 구조를 데이터 기반 의사결정이 가능한 조직으로 재편하는 것이며, 여기에는 조직문화의 변화도 포함된다. 따라서 어디에서 어떤 데이터가 생성되고 어떻게 활용되는지 파악해야 하며, 의사결정에 필요한 데이터가 무엇인지도 함께 파악되어야 한다. 기관의 사업 분야 및 특성을 감안하여 어떠한 의사결정에 데이터를 활용하게 될 것인지, 그리고 어떤 데이터를 활용할 수 있는지 판단하여, 기관에 적합한 데이터 거버넌스 수립이 필요하다.

데이터 거버넌스 구축에 정해진 순서는 없다. 다만, 기관의 특성과 상황을 감안하여 데이터 거버넌스를 정의하고, 작은 부분부터 구축하여 확대해 나아갈 것을 제시하고 있다. KOICA는 공공기관으로 개발도상국 정부 및 관계자와 국내 사업 파트너들과 개발도상국에서의 다양한 사업을 진행하고 있다. 대략적으로 KOICA에서 생성 및 수집관리 가능한 데이터는 사업 정보, 사업 현장에서의 데이터, 사업 파트너 계약 정보, 전시·행사 및 참여자 정보, 예산 및 집행 내역, 기관의 미션·전략 및 조직구성, 직원 정보 등이 있으며, 각 정보들은 분산되어 관리⁸⁾되거나 혹은 누락되고 있다. 그리고 특이사항으로는 공공기관이기에 일부 데이터에 대해서는 공개해야 하는 의무가 있다. 데이터 거버넌스 구축을 위해서는 어떤 데이터가 어디에서 생성되어 누가 권한을 가졌는지, 그리고 누가 어떤 방식으로 활용하고 있는지를 살펴보아야 할 것이다. 그리고 기관의 중요 의사결정에 필요한 데이터, 업무에 지속적으로 요구되고 있는 데이터, 외부에서 요구하는 데이터 등에 대한 분석이 필요하다. 이를 토대로 데이터맵(data map) 혹은 데이터공급사슬(data supply chain)을 그려본다면, 보다 명확하게 데이터의 생성부터 활용까지 그 흐름을 파악할 수 있을 것이다. 대부분의 데이터 생성과 활용이 다르다는 점을 감안할 때, 이를 통해 실제 데이터가 생성되어 어디에서 활용되고 있는지, 혹은 생성되었으나 활용되지 못하고 있거나, 활용수요가 있으나 데이터가 전달되지 못하고 있는 점 등을 보다 쉽게 파악하여 이를 보완하는 거버넌스 수립이 가능할 것이다. 또한 데이터 거버넌스를 처음부터 전사 차원으로 확대할 수 없는 바, 기관에서의 데이터 중요도를 식별하고 수명주기(life cycle)를 고려하여 관련 정책 및 절차를 우선하여 설정하고, 이후 점진적으로 확대해가야 할 것이다.

한편, 변화하는 데이터의 가치에 따라 데이터 거버넌스 수립의 중요성이 증대하고 있으나,

8) [예] 지역실 : 사업 관련 데이터, 조달실 : KOICA 사업 참여 업체 데이터, 재무실 : 지출결의 관련 사업 및 업체 데이터

치밀하게 준비하였어도 여러 가지 사유로 인해 거버넌스가 제대로 작동하지 않을 가능성도 존재한다. SAS(2014)는 데이터 거버넌스의 실패 이유를 다음과 같이 제시하였다.

-
- ① 기관의 데이터 거버넌스 정의가 명확하지 않음
 - ② 데이터를 IT 문제로 간주하며, 사업부서와 IT부서가 소통하지 않음
 - ③ 조직 고유한 문화를 고려하지 않음
 - ④ 조직 구조의 과도한 분산화, 중앙집중식 의사 결정을 지원하지 않거나 의사 결정 과정이 불명확한 조직 구조
 - ⑤ 데이터 거버넌스를 학술적 이론적인 활동으로 간주 혹은 개별 프로젝트처럼 취급
 - ⑥ 지식 및 영향력 있는 기존의 운영위원회를 간과함
 - ⑦ 실행력 부족: 구조화되고 전문적이며 반복 가능한 방식으로 데이터 관리가 이루어지지 않음
 - ⑧ 투자수익률(ROI)이 명확하지 않고 데이터 거버넌스 활동을 비즈니스 가치와 연결하기 어려움
 - ⑨ 핵심 자원이 이미 과부하되어 거버넌스 활동을 수행할 수 없음
-

결과적으로 데이터 거버넌스는 어떤 데이터를 갖고, 어떠한 방향성을 제시할 것인지, 누가 어떤 역할과 책임을 갖고 갈 것인지, 어떤 기술과 솔루션을 통해 재현해낼 것인지에 대한 답을 하는 과정이라고 볼 수 있다. 데이터의 가치는 누가 어떠한 관점에서 살펴보느냐에 따라 달라진다. 데이터는 IT 관련 부서만의 관리 업무 대상이 아니다. 전사의 모든 직원들이 직접 필요한 데이터를 수월하게 활용할 수 있는 체계를 갖춰야 한다. 많은 이들이 과거의 사업 운영 결과를 수치로 살펴보기 위해 데이터를 언급하고 있으나, 오늘날 데이터는 주어진 문제를 해결하고 새로운 기회를 파악하는 데 활용할 수 있다. 데이터를 가치 창출의 원동력(driver of value)으로 인식하고, 업무의 결과물이 아닌 자산으로써 체계적인 관리의 필요성이 다대하다.

2. 데이터 거버넌스와 데이터 관리

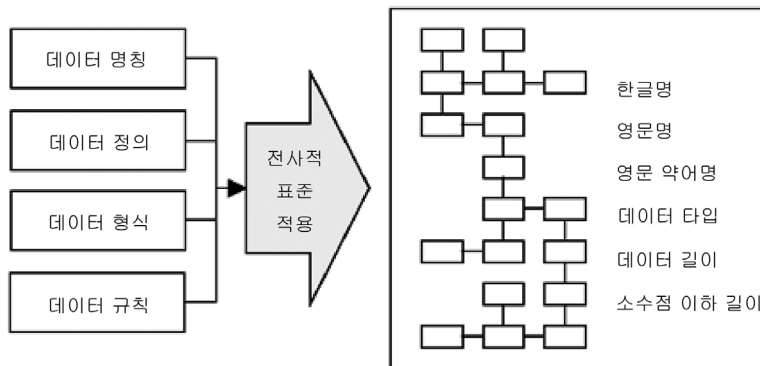
앞서 본 연구에서는 데이터 거버넌스를 데이터의 생성부터 활용까지의 전체 관리 과정에서 필요한 전략, 정책, 기준, 역할, 책임 등에 관한 체계라고 정의하였다. 많은 경우에서 데이터 거버넌스는 데이터 관리와 혼용해서 활용되고 있으나, 각각이 의미하는 바는 조금씩 다르다. 데이터 관리는 관리체계를 확립하는 구조적인 부분과 그 안에 어떠한 데이터를 포함하고 어떻게 활용할지 등의 기능적인 부분의 모든 과정을 의미한다. 또한 데이터가 제대로 있는지에 대한 상태를 살피는 것부터 데이터의 실제 활용 여부 및 가치가 있는지를 파악하는 것도 포함한다. 따라서 데이터 거버넌스라는 큰 체계 안에 데이터 관리 요소가 들어있다고 볼 수 있다. 보는 관점에 따라 반대로 주장하는 경우도 있으나, 본 연구에서는 데이터 관리를 거버넌스 하위 단의

실질적인 데이터의 효과적인 운영과정으로 보고자 한다. 이를 위해 데이터 관리의 주요 요소로 데이터 표준화, 품질관리, 생애주기관리 정도를 살펴보고자 한다. 데이터의 품질은 데이터의 활용을 좌우한다. 양질의 데이터로 관리하기 위해서는 합리적인 표준 마련이 필요하다. 합의된 기준에 따라 수집 및 중복데이터의 통합이 이루어져야 데이터의 신뢰성을 높일 수 있기 때문이다. 또한 데이터의 저장을 무기한으로 할 수 없는 바, 각 데이터의 생애주기에 따른 효율적 관리를 통해 양질의 데이터를 제공할 수 있어야 한다.

1) 표준화

무수히 생산되고 있는 데이터를 실질적으로 활용하기 위해서는 데이터의 체계적인 관리가 필요하며, 데이터 관리에 있어 기본은 데이터 표준화가 될 것이다. 데이터의 표준화는 다양하게 존재하는 데이터를 특정 기준에 따라 정리하는 것으로 볼 수 있다. 공공기관의 데이터베이스 표준화 지침⁹⁾ 제2조 2항에 따르면, 표준화란 코드, 용어, 도메인, 메타데이터, 데이터셋 등의 표준을 수립하여 공공데이터베이스에 일관되게 적용하는 일련의 활동으로 정의하고 있다. 한국데이터산업진흥원에서는 시스템별로 산재해 있는 데이터 정보 요소에 대한 명칭, 정의, 형식, 규칙에 대한 원칙을 수립하여 이를 전사적으로 적용하는 것으로 정의하고 있다.

〈그림 II-5〉 데이터 표준화 의미



자료: 한국데이터산업진흥원 데이터 온에어 플랫폼

9) 공공데이터베이스 구축 운영, 메타데이터 등록 관리 등의 표준화 업무에 적용하며(제3조), 공공기관이 생성 또는 취득하여 관리하는 데이터베이스의 표준화에 필요한 세부사항을 정하는 것을 목적으로 함(제1조)

데이터에 대한 표준 및 규칙 수립 적용한다는 것은 발생 가능한 데이터에 대해 사전에 입력 가능한 값을 정해둠으로 입력 오류 등의 위험 요소를 최소화하는 동시에 정확성 및 정합성을 향상시키는 역할을 한다.

앞서 언급한 바와 같이 데이터는 생산자, 처리자, 사용자가 모두 다른 경우가 많다. 따라서 값과 명칭만 있는 데이터를 올바르게 활용하기 위해서는 수집하는 데이터에 대해 공동의 합의된 기준이 요구된다. 그리고 합의된 기준에 따라 전사적으로 같은 값과 명칭을 사용하고, 공통의 언어로 수집 관리될 수 있도록 지속적인 노력이 필요하다. 이는 단순히 데이터로 저장되는 값과 명칭을 일괄적인 기준에 따라 동일하게 수집한다는 의미보다는 크다. 기관 내 모든 구성원들이 사용하는 용어와 데이터가 의미하는 내용, 나아가 데이터 목록까지도 표준화하여 관리하는 것을 포함한다. 또한 기관에서 생산되고 있는 데이터는 외부로도 개방되고 있는바, 명확한 기준 하에 일관되게 정보를 제공해야 데이터의 올바른 활용이 가능할 것이다. 즉, 데이터의 표준화는 기관의 지식 체계의 표준화라고도 할 수 있을 것이다.

데이터를 표준화하기 위해서는 일련의 절차가 요구되는데, 이춘열 · 황철현(2014)에 의하면 기관에서 먼저 필요한 데이터가 무엇인지, 실제 기관에서 보유하고 있는 데이터가 무엇인지 파악한 후, 이에 따라 데이터 표준화 방안 및 사용 방안을 설계해야 한다고 언급하였다. 즉, 데이터를 나타내는 개체(기관을 구성하는 물리적이거나 개념적인 실체, 예: 사원, 고객, 제품, 예산, 배송 등)¹⁰⁾가 무엇인지, 그리고 개체의 속성을 파악하여 필요한 데이터를 확인하고, 현재 운용되는 시스템을 중심으로 데이터 목록을 작성하여야 한다. 이를 통해 데이터의 중복 사용 및 저장의 실태를 파악할 수 있게 된다. 중복된 데이터를 통합하여 표준화하기 위해서는 기존의 데이터베이스들을 유지하면서 연합할 것인지, 새로이 통합 데이터베이스를 구축할 것인지에 대한 논의도 함께 이루어져야 한다. 간단히 KOICA의 예를 들면, 고객에 대한 데이터는 조달 시스템에도 저장되어 있고, 재무 시스템에도 저장되어 있으며, 사업 담당 부서에도 저장되어 있다. 그리고 이는 계약 체결, 지출결의, 사업진행, 증명서 발급, 고객만족도 조사, 감사 등에 활용되고 있다. 현재 각각의 데이터베이스는 생성, 저장, 활용이 모두 다르며, 중복되어 각각 관리되고 있다. 서로 연동되어있지 않기 때문에 데이터베이스의 갱신 시점도 각자 달라지면서 시간이 흐를수록 같아야 할 데이터 값이 달라지는 경우도 발생한다. 각 데이터가 일관된 기준으로 저장될 수 있도록 기관 내 표준화 지침을 마련하고 공유 방안에 대한 기관 내 합의가 이루어져야,

10) 개체들을 업무영역으로 구분 후, 영역별 상세 분석 가능. 업무영역은 2개 이상의 개체들을 포함함. 예: 고객 영역은 고객 개체와 지역 개체 포함, 제품 영역은 제품, 부품, 공급자, 공장 등을 포함 (이춘열 · 황철현 2014)

동일 개체의 데이터 공유 및 활용이 가능해질 것이다. 또한 데이터 흐름도 또는 배포 공정을 통해 어느 데이터가 어떻게 연동하여 갱신될 것인지에 대한 설계도 함께 고려되어야 할 것이다.

위와 유사하게 한국데이터산업진흥원에서도 [표 II-3]과 같이 데이터 표준화 요구사항을 수집하여 데이터 표준을 정의하고 확정하여 관리로 이루어지는 절차로 언급하고 있다. 이를 위한 세부 활동에는 데이터 표준 정립, 데이터 관리 책임을 부여받은 조직의 구성, 데이터 표준화 절차에 대한 합의가 전제된다. 데이터 표준 정립은 용어에 대한 표준을 정의하는 것이며, 동일한 개념에 대해 국문 및 영문명을 일관되게 부여하도록 한다. 데이터 관리 담당 부서는 부서간 데이터 구조를 조율하는 것으로, 데이터 표준을 정의하고 적용 및 준수 여부를 확인 등의 [표 II-3]의 절차를 수행하는 것이다.

〈표 II-3〉 데이터 표준화 절차

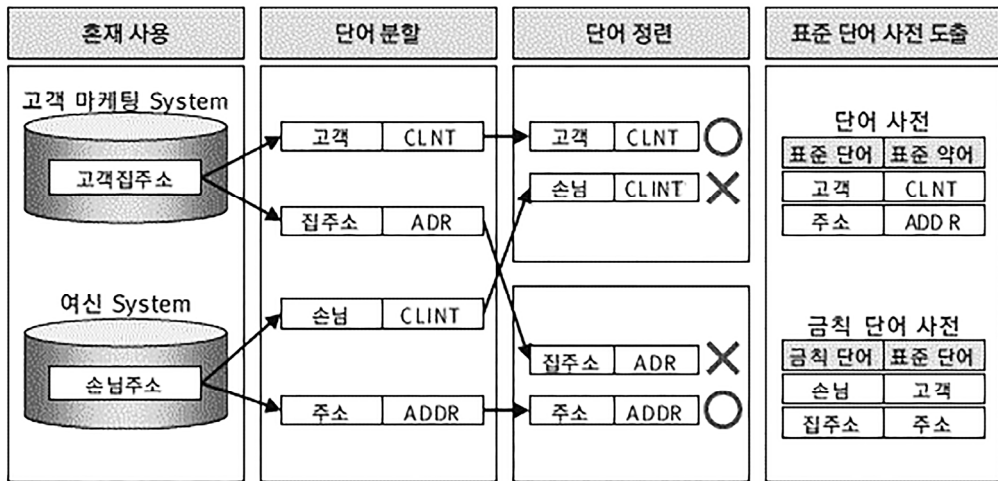
구분	주요활동
데이터 표준화 요구 사항 수집	- 개별 시스템 데이터 표준 수집 - 데이터 표준화 요구 사항 수집 - 표준화 현황 진단
데이터 표준 정의	- 표준화 원칙 - 데이터 표준 정의 : 표준 용어, 표준 단어, 표준 도메인, 표준코드, 기타 표준
데이터 표준 확정	- 데이터 표준 검토 및 확정 - 데이터 표준 공표
데이터 표준 관리	- 데이터 표준 이행 - 데이터 표준 관리 절차 수립 : 데이터 표준 적용, 변경, 준수 검사 절차

자료 : 한국데이터산업진흥원 데이터 온에어 플랫폼

앞서 언급한 문헌 외에도 다수의 문헌에서 유사한 절차를 언급하고 있다. 이들을 종합해보면, 표준화를 위해서는 기관 내 보유하고 있는 데이터와 필요한 데이터를 조사하여 정해진 규칙에 따라 정리한 후, 각 데이터의 표준화 정도를 분석하고 현황을 진단하여, 기관의 성격에 따라 표준화 원칙을 정립하는 것이 우선적으로 이루어져야 한다. 여기에는 각 데이터의 흐름도 함께 고려되어야 할 것이다. 각각의 데이터는 단독으로 활용되기보다는 다수의 데이터를 취합하여 의미 있는 정보를 만들어내기 때문이다. 그 다음, 원칙에 따라 각 데이터의 표준화 작업을 시행하고, 원칙이 데이터에 적용 가능한지 검토하여 이를 최종적으로 확정하면, 이후에는 수집되는 데이터에 대해 표준화가 제대로 적용되었는지 기준을 준수하였는지 검토 및 관리하는 것으로 정리해볼 수 있다.

데이터의 표준화가 문자 그대로 데이터의 표준화만을 의미하지는 않는다. 데이터 표준화는 기관 내 사용되고 있는 용어의 표준화라고도 할 수 있을 만큼의 영향력을 갖고 있다. 업무 효율화를 위해 데이터를 효과적으로 사용하기 위해서는 기관 내에서 사용되고 있는 용어들이 어떠한 의미로 사용되고 있으며, 암묵적 의미까지도 파악하여 이를 합의된 공용 언어로 만들어야 한다. 또한 동일한 기준을 적용하여 기관 내에서 생성되고 있는 데이터를 수집해야 혼선 및 중복 처리되는 것을 방지할 수 있을 것이다¹¹⁾.

〈그림 II-6〉 표준 단어의 도출



자료: 한국데이터산업진흥원 데이터 온에어 플랫폼

따라서 어떤 데이터가 어디에서 생성되는지 파악하고, 어떠한 용어로 작성되어야 할지에 대한 합의가 반드시 이루어져야 한다. 이는 결과적으로 기관 내 데이터만의 표준화 과제가 아닌 기관에서 활용되고 있는 표준화된 언어가 될 것이며, 나아가 공개되는 데이터를 통한 외부와의 소통 언어가 될 것이다.

절차에 따라 진행되는 데이터 표준화는 정도에 따라 진행 과정을 [표 II-4]와 같이 나누어 살펴볼 수 있다.

11) 국가명, 업체명, 사업명, 계좌번호, 원화/달러 단위 등 하나의 명칭만 허용하고 사용하도록 해야 함. 특히 외국어 표기 방식, 사업자번호, 주민번호, 사업번호, 전화번호 등의 ‘-’ 표시, 띄어쓰기 등에 대해 일관된 기준 필요

〈표 Ⅱ-4〉 데이터 표준화 수준

수준	내용
1. 데이터 연합단계	- 전사 차원의 데이터 표준화 및 공유 지침 미비 - 개별 데이터셋 관리 및 1:1 데이터 공유 - 개별 시스템 간 국지적으로 데이터를 연계하는 수준
2. 물리적 데이터 통합 단계	- 전사 보유 데이터 목록 구성 - 일부 데이터의 통합된 데이터베이스에 저장하나 저장데이터의 표준화는 아님 - 데이터의 물리적 공통의 장소에 저장 노력
3. 데이터 표준화 시작 단계	- 용어 사전 및 동의어 분류 등의 표준화를 위한 시도 - 데이터 표준화 전담 조직 설치 - 표준화 지침 및 절차에 대해 전사적 인식 확대
4. 데이터 표준화 관리 단계	- 표준화 절차 및 전담 조직 정착 - 표준화된 데이터 및 용어 사용에 적응 - 기관의 데이터를 주요 자산으로 인식하고 관리 - 전사적으로 데이터 표준화 수용 단계
5. 데이터 표준화 최적화 단계	- 표준화된 데이터 활용 - 데이터를 기관의 주요 자원으로 인지하고 적극 활용 - 데이터 전략과 기관 전략과 연계

자료 : 이춘열 · 황철현(2014) 참고하여 저자 정리

물리적 통합 수준까지는 각 데이터별 관리 담당자의 니즈 혹은 활용자의 요구에 의해 부분적으로 표준화가 이루어졌다면, 이후 수준에서는 전사 차원의 동일한 절차 및 지침에 따른 하향식(top-down)의 관리가 필요하다. 데이터 표준화 최적화 단계에서는 표준화된 데이터를 기반으로 필요에 의한 분산·셀프 서비스화된 분석이 가능해진다. 데이터 표준화는 중앙에서 통제하는 하향식 관리가 이루어지지만, 이를 활용하는 데는 각각의 니즈에 따른 다양한 분석이 가능해지도록 하는 것이 핵심이라고 볼 수 있다.

데이터 표준화 수준을 KOICA에 비추본다면 데이터 표준화 시작 단계에 해당한다고 볼 수 있다. 디지털 전환 전략을 발표함과 함께 체계적인 데이터 관리를 위해 전담 조직이 구성되었으며, 데이터 표준화를 위한 용어정리, 관련 지침 및 절차 마련 등을 위한 노력이 이루어지고 있다. 현재 핵심 데이터를 선정하고 이를 표준화하는 절차에 대한 논의가 진행되고 있다. 이러한 가운데 기관 내 사용되고 있는 용어에 대한 분류 및 정리가 이루어져야 하며, 전사 차원의 데이터 표준화에 대한 인지도 제고가 필요한 상황이다. 한편, KOICA는 공공기관으로 공공데이터 개방 의무가 있기에, 용어 분류 및 정리에 있어 아래와 같이 공공기관 데이터베이스 표준화 지침의 서식을 참고하는 것도 방안이 될 수 있다.

〈그림 II-7〉 표준용어 및 표준단어 정의서 예시

■ 공공기관의 데이터베이스 표준화 지침 [별지 제1호 서식]

표준용어정의서

표준 용어명	영문명	영문 약어명	용어 설명	표준 도메인명	허용값	관리 부서명	표준 코드명	업무 분야

□ 명세 지침

항 목	항목 정의 및 작성 지침
표준용어명	○ 엔터티, 속성 등 데이터 요소의 명명에 사용될 용어를 기재 - 동음이의어와 약어는 중복을 허용하나 사용 또는 작성을 최소화 ○ 한글을 원칙으로 하며, 특수 문자와 공백은 사용 불가 - 숫자의 사용은 가능하나, 표준용어의 직관적 의미 파악이 가능하도록 사용을 최소화 ○ 관용적으로 널리 알려진 영문약어에 한하여 표준용어로 사용 가능 (예: URL, IP 등)
영문명	○ 표준용어의 영문명으로, 대문자를 사용하여 Full Name으로 기재 - (예시) 표준용어명 "가공식품"의 경우 영문명은 "PROCESS FOOD"로 기재
영문약어명	○ 해당 표준용어에 대한 영문약어명을 기재 ○ 대문자로 작성하고, 불가피한 경우를 제외하고는 특수문자는 사용하지 않도록 유의 ○ "용어구분"이 동의어인 경우 해당 용어의 표준어와 동일한 영문약어를 기재 ○ 아래 예시를 참조하여 자체적인 영문약어 명명규칙을 작성하여 약어명을 정의할 것을 권고 - (예시) 모음을 제외한 자음의 결합으로 영문 약어 작성 (PRICE → PRC) - (예시) 같은 자음이 2개 이상 연속되면 1개만 선택하여 작성 (MAPPING → MPNG) - (예시) 용어 영문명의 앞부분만 사용하여 작성 (ADDRESS → ADDR) - (예시) 용어 영문명의 첫 글자와 마지막 글자를 조합하여 작성 (예: YARD → YD) - (예시) 용어 영문명이 긴 경우는 첫 자음과 마지막 자음 사이에 특징 있는 자음들을 조합하여 작성 (NETWORK → NW, SOFTWARE → SW, MESSAGE → MSG)
용어설명	○ 표준용어의 의미나 내용을 상세히 서술 ○ 표준어, 동의어 이외에 금칙어를 제공할 경우에는 금칙어 목록을 포함하여 기재할 수 있으며, 1개 용어에 대하여 다수의 금칙어가 존재할 수 있으므로 별도의 관리항목으로 정의하여 사용 가능 - (예시) '휴대폰'이 표준어로 정의된 경우 '핸드폰', '모바일폰' 등을 금칙어로 규정하여 표준용어로 사용하지 못하도록 할 수 있음
표준 도메인명	○ 도메인정의서에 정의된 용어의 경우 해당 도메인명을 기재
허용값	○ 해당 표준도메인이 가질 수 있는 최대/최소값이나 유효값을 기재 - (예시) 월 : 01~12, 여부 : 'Y', 'N' ※ 행정 표준코드는 행정 표준코드관리시스템(https://www.code.go.kr/)을 참조
관리부서명	○ 해당 용어를 생성 및 관리하는 주체인 부서명을 기재
표준코드명	○ 해당 용어가 표준코드에 해당하는 경우 해당 표준코드명을 기재
업무분야	○ 해당 테이블의 내용과 관련된 BRM 분류체계를 기재 ○ 해당 용어가 다수의 업무분야에서 일상적으로 사용되는 경우에는 "공통영역"으로 기재, 그 외에는 정부업무분류체계(BRM) 1단계 수준으로 분류하여 기재

210mm×297mm(백상지 80g/㎡)

※ 표준용어의 저장 및 표현형식, 데이터 타입 및 길이 등은 표준도메인 정의서 참조

표준단어정의서

표준 단어명	영문명	영문 약어명	단어 설명	형식단어 여부	도메인 분류명	이음동의어 목록	금칙어 목록

☐ 명세 지침

항 목	항목 정의 및 작성 지침
표준단어명	○ 표준단어에 대한 한글명을 기재
단어 영문명	○ 표준단어에 대한 영문전체명(full name)을 기재
단어 영문약어명	○ 표준단어에 대한 영문 약어명을 기재
단어 설명	○ 표준단어에 대한 정의 및 설명을 기재
형식단어 여부	○ 해당 표준단어가 도메인 특성을 지닌 형식단어(분류어)인지의 여부를 기재
도메인 분류명	○ 형식단어(분류어)인 경우 해당되는 표준도메인의 분류명을 기재
이음동의어 목록	○ 해당 표준단어에 대한 이음동의어 목록을 기재 - 이음동의어 : 소리는 다르나 의미가 동일한 단어 (예) 연도/년도, 비율/율 - 이음동의어는 공통표준단어와 한글명 이외에는 동일하여 한글명만 관리
금칙어 목록	○ 해당 표준단어에 대한 금칙어 목록을 기재 - 금칙어 : 사용이 허락되지 않거나 일정기간 사용하다가 특정시점 이후 사용이 중지된 단어 - (예시) 파일 (표준어) ↔ 화일 (금칙어, 사용×) - 금칙어는 표준단어 정의대상이 아니며 금칙어의 한글명만 관리

210mm×297mm(백상지 80g/㎡)

자료: 공공기관데이터베이스 표준화 지침

다만, 모든 데이터에 대한 표준을 한 번에 수립하여 변환하는 것은 불가능에 가까우며 많은 위험이 따른다. 표준화는 전사적으로 동일한 언어를 활용하는 것에 대한 약속이기에 초반에는 다양한 변수에 따른 여러 시행착오가 예상된다. 따라서 소규모 데이터부터 시작하여 기관 내 표준화에 대한 인식을 확산시킬 필요가 있다. 표준화를 시도할 소규모 데이터를 선정하는 것은 기관의 요구사항 및 중요도에 따라 우선순위가 달라진다. 기관의 데이터를 어떤 관점에서 분류할 것인지에 대한 고민도 필요하다. 예를 들면, 주제별로 재무 데이터, 고객 데이터, 인사 데이터,

지역 데이터 등으로 구분 가능하며, 또는 개발협력 사업 데이터, 기관 운영 데이터, 혹은 표준화가 진행된 정도에 따라서도 구분 가능하며, 데이터의 성격과 특성을 고려하여 분류해야 할 것이다. 데이터 관리와 관련된 다수의 문헌은 민간기업을 대상으로 하고 있는바, 대부분 고객관리를 중심으로 데이터 관리를 언급하고 있다. 즉, 상품을 구매하는 고객의 니즈를 빠르게 파악하여 향후 사업 방향성에 대한 의사결정을 내리기 위함이다. 이를 KOICA의 시각으로 본다면, 우선순위는 개발도상국의 현지 행정 및 시장 데이터와 진행되고 있는 사업의 성과 데이터 등에 있을 것이다. 따라서 이 중 일부 소규모 데이터를 선정하여 표준화를 시범적으로 진행하고 점진적으로 확대해 나아가는 것이 효과적일 수 있다.

표준화의 핵심은 기관 내 공통의 언어를 사용하기 위한 과정을 확립하는 것이라고 볼 수 있다. 이러한 기준은 기존의 중복 관리 및 데이터 관리체계 재정비 등 데이터를 통폐합하는 과정에서도, 향후 데이터의 품질관리를 위해서도 반드시 요구되는 절차이기도 하다. 표준이 명확해야만 데이터 신뢰도 유지 및 일관성 확보가 가능해지기 때문이다. 표준화 절차에 대한 가이드라인 수립 및 일련의 과정에서 누가 어떠한 권한을 갖고 결정을 할지에 대해서도 합의가 뒤따라야 한다. 중앙집중적으로 데이터를 관리하는데 있어 동일한 기준에 따라 표준화가 진행될 필요가 있으며, 이를 확대해 나아가는 과정에서 일관성을 유지할 수 있게 된다. 또한 데이터의 표준화는 일회성으로 완료되지 않는다. 꾸준히 새로운 데이터가 만들어지고 있으며, 때로는 기존의 데이터와 통폐합이 요구되기도 한다. 표준화 가이드라인도 지속적인 현행화가 필요하다.

2) 품질

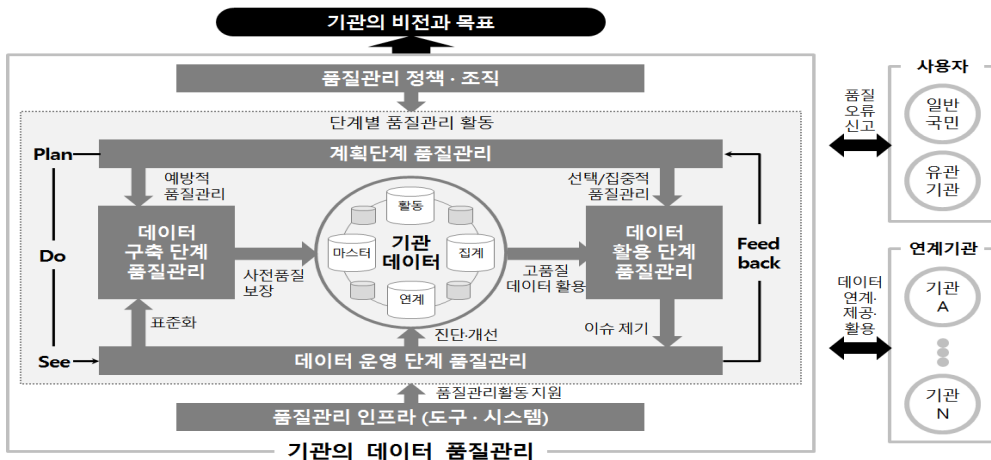
표준화된 데이터를 저장하고 관리하는 목적은 결국 양질의 데이터를 활용하기 위함이다. 품질의 사전적 의미는 물건의 성질과 바탕이다. 데이터의 관리 여부는 데이터 품질에 의해 판단해볼 수 있다. 관리가 효과적으로 이루어질수록 데이터의 품질은 양호할 것이며, 그렇지 못한 경우 관리가 제대로 이루어지지 않고 있다고 짐작할 수 있다. 데이터의 중요도가 증대될수록, 그리고 데이터에 기반하여 기관의 방향성 및 사업의 전략 수립 등 의사결정이 이루어짐에 따라 데이터 품질의 중요성도 함께 강조되고 있다. 과거에는 데이터의 공유는 상대적으로 제한적이었다. 따라서 데이터의 의미가 변질될 우려가 적었으며, 표준화 또는 중복에 대한 이슈가 많지 않았다. 당장의 비효율적인 데이터 관리가 딱히 문제 되지 않았다. 하지만 오늘날에는 무수히 많은 사람과 방대한 양의 데이터를 공유하고 있으며, 데이터의 내용이나 의미하는 바를 제대로 숙지하지 않은 채 활용하거나 유효기간이 만료된 잘못된 데이터를 활용할 가능성도 있다. 따라서 데이터의

잘못된 활용을 방지하고 양질의 데이터 활용 확산을 위해 명확한 기준에 따라 생성 및 관리하는 것이 중요하다.

데이터 품질이란 데이터의 최신성, 정확성, 상호연계성 등을 확보하여 사용자에게 유용한 가치를 줄 수 있는 수준을 의미하며, 데이터 품질관리란 사용자에게 유용한 가치를 제공하도록 데이터의 품질을 확보하기 위한 품질목표 설정, 품질 진단 및 개선 등 일련의 활동과 이를 지원하기 위한 관련 도구로 정의하고 있다(공공데이터포털). 한국데이터베이스진흥센터(2006)의 지침에 의하면 기관이나 조직 내외부의 정보시스템 및 DB 사용자의 기대를 만족시키기 위해 지속적으로 수행하는 데이터 관리 및 개선 활동을 데이터 품질관리라 정의하였다. 즉, 데이터를 활용 가능한 수준으로 유지 관리하여 사용자의 만족도를 극대화하기 위한 활동으로 이해할 수 있다.

일반적으로 데이터 품질관리를 데이터 생성·수집 이후 활용되기 전까지의 저장 단계에 해당하는 것으로 인식하는 경우가 많다. 하지만, 데이터 품질관리는 생성부터 활용까지 전 단계에 걸쳐 이루어지고 있다. 데이터 표준화도 품질관리의 가장 기본 중 하나다. 생성단계부터 표준화된 기준에 따라 수집 관리함으로써 양질의 데이터가 될 수 있도록 설계되어야 한다. 한국정보화진흥원(2018) 매뉴얼에 의하면 품질관리는 크게 계획단계, 구축단계, 운영단계, 활용단계 품질관리 등 4단계로 구분 가능하다.

〈그림 II-8〉 기관의 데이터 품질관리



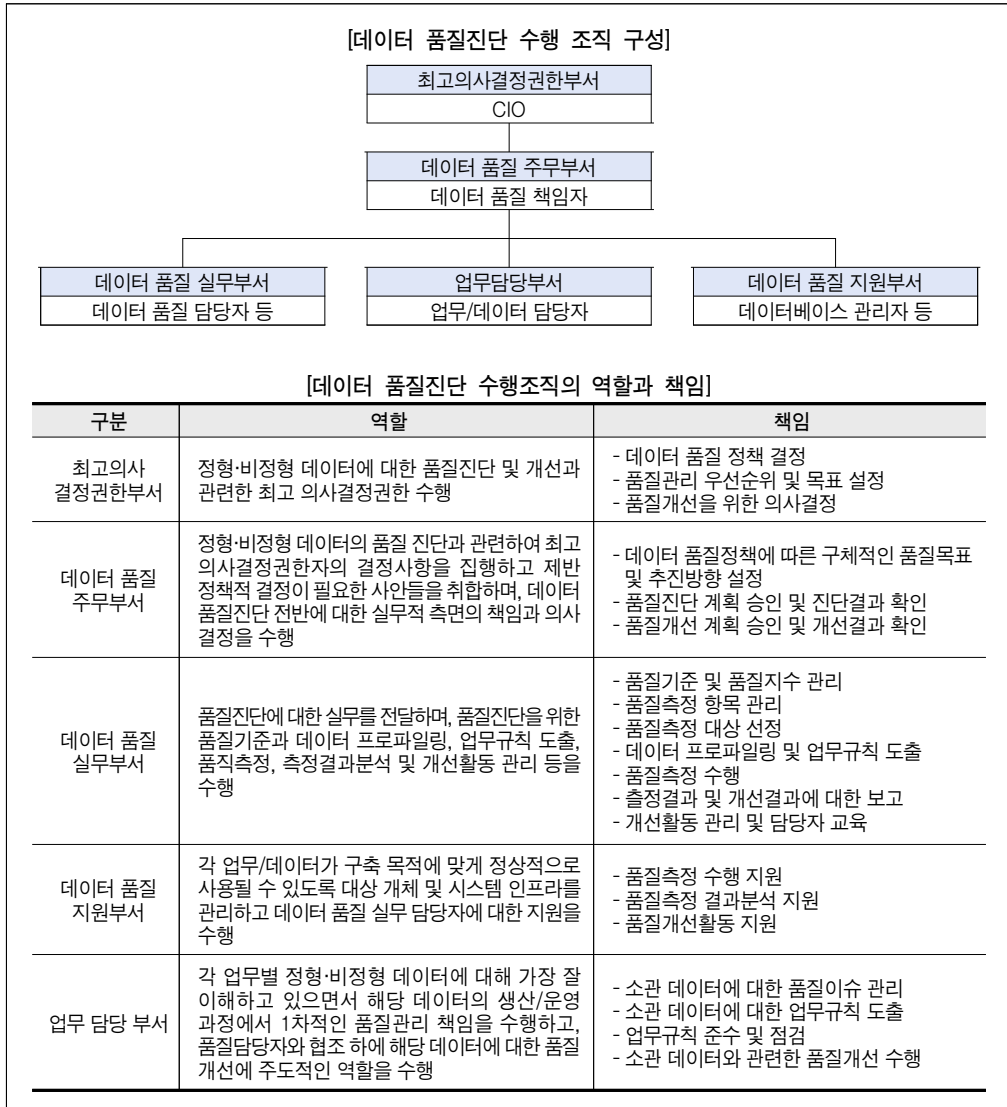
자료: 한국정보화진흥원 2018

계획단계에서는 데이터 수집 시 고려해야 할 데이터 품질관리 계획 수립을 의미하며, 구축단계에서는 표준화를 포함하여 데이터 저장 시 고려해야 할 데이터 품질관리 활동을 의미한다. 운영단계에서는 품질 진단 및 개선이 이루어지며, 활용단계에서는 사용자로부터의 피드백을 통한 오류 관리 및 활용에 대한 성과평가를 의미한다.

기관 내에서 생성되고 활용되는 데이터는 각각의 시스템에 혼재되어 있거나 오랜 기간 축적되면서 오류가 발생할 가능성이 높아진다. 특히 시스템의 변경 혹은 시스템 간의 연계는 데이터 오류 등의 다양한 이슈를 야기한다. 초기 데이터 값의 입력 오류, 시간의 흐름에 따른 데이터 손상 또는 유효기간 만료, 데이터 이동에 따른 누락 혹은 결함 발생, 메타데이터 부재에 따른 데이터 의미 전달 불가 등의 이슈로 데이터의 품질을 하락시킬 수 있다. 또는 데이터 오너십 부재 등의 관리체계 미흡, 데이터 표준화 지침서 미수립 등의 데이터 표준 미흡, 시스템 모델에 변경이 있었으나 관련 지침서 미비 등의 데이터 구조상의 이슈로 데이터 품질이 문제가 될 수도 있다.

품질관리를 위해서는 진단이 필요하다. 품질 진단을 위한 전담 조직과 그 역할에 대해 구체적으로 명시되어야 하며, 절차를 정의해야 한다. 그래야 정해진 기준과 방법에 따라 인력과 자원이 투입될 수 있으며, 절차에 따라 공정한 진단이 이루어질 수 있기 때문이다. 특히, 데이터의 품질관리에는 사용 편이성 등과 같은 주관적인 항목도 포함될 수 있으므로 명확한 기준과 절차가 요구된다. 예를 들면 [표 II -5]와 같은 조직구성과 역할이 주어질 수 있다.

〈표 Ⅱ-5〉 데이터 품질 진단 수행 조직구성 및 역할과 책임



자료: 한국데이터베이스진흥원 2009

데이터 품질과 데이터 활용 정도가 동일한 의미는 아니다. 데이터의 활용이 잘 이루어지기 위해서 필요한 여러 조건들 중 하나다. 데이터의 활용은 사용자가 필요한 양질의 데이터가 어디에 있는지 알아야 하며, 다른 데이터와의 상호 통합 및 공유를 할 수 있어야 한다. 따라서 데이터 품질은 일반적으로 정확성, 유용성, 완전성, 유일성, 유효성, 일관성 등의 원활한 활용을 위한 다양한 품질 기준에 의해 평가되며, 기관에 특성에 따라 구체화 및 변형하거나 기준을 추가하여 사용되고

있는 것으로 나타난다.

한국데이터베이스진흥원(2009)에서는 정형 텍스트 데이터에 대해 완전성, 유효성, 정확성, 유일성, 일관성을 기준으로 제시한 바 있다. 필수항목이 모두 입력되었는지(완전성), 데이터의 항목이 중복되지는 않는지(유일성), 정해진 데이터 유효범위 및 도메인을 충족하는지(유효성), 데이터의 구조, 값 등이 일관되게 정리되고 일치하는지(일관성), 객체의 표현 값이 정확히 반영되고 있는지(정확성)를 살펴보는 것이다. 아래 [표 II-6]은 품질 기준을 확장하여 상세 품질 기준을 제시한 사례이다.

〈표 II-6〉 세부 품질기준 설명 및 활용 예시

품질 기준	세부 품질기준	품질기준 설명	활용예시
완전성	개별 완전성	필수 컬럼에는 값의 누락이 없어야 한다	고객의 아이디는 NULL일 수 없다
	조건 완전성	조건에 따라 컬럼 값이 항상 존재해야 한다	기업 고객의 사업자 등록번호가 NULL일 수 없다
유효성	범위 유효성	컬럼 값이 주어진 범위내에 존재해야 한다	기준점 좌표각은 -360 초과 360 미만까지의 값을 가진다
	날짜 유효성	컬럼 값이 날짜유형일 경우에는 유효날짜 값을 가져야 한다	99991231, 20080231은 유효하지 않은 값이다
	형식 유효성	컬럼은 정해진 형식과 일치하는 값을 가져야 한다	주민번호형식은 '999999-9999999'의 형식이어야 한다
정확성	선후관계 정확성	복수의 컬럼값이 선후 관계에 있을 경우 이 규칙을 지켜야 한다	시작일은 종료일 이전 시점이어야 한다
	계산/집계 정확성	한 컬럼의 값은 다수 컬럼의 계산된 값일 경우 계산 값이 정확해야 한다	월 통계 테이블의 매출액은 기산일이 현월인 매출액의 총합과 일치해야 한다
	최신성	정보의 발생, 수집, 그리고 갱신 주기를 유지해야 한다	고객의 현재 값은 고객변경이력의 마지막 ROW와 일치해야 한다
	업무규칙 정확성	컬럼이 업무적으로 복잡하게 연관된 경우 관련 업무규칙에 일치해야 한다	지급원장의 지급여부가 'Y'이면 지급원장의 지급일자는 신청일보다 이전시점이어야 하고 NULL이 아니어야 한다
유일성	단독 유일성	컬럼은 유일한 값을 가져야 한다	고객의 이메일 주소는 유일해야 한다
	조건 유일성	업무 조건에 따라 컬럼 값은 유일해야 한다	교육과정의 강의시작일이 있으면, 강의실 코드, 임대일, 강사코드가 모두 동일한 레코드는 존재하지 않는다

품질 기준	세부 품질기준	품질기준 설명	활용예시
일관성	기준코드 일관성	컬럼이 통합코드를 기준코드로 사용할 경우 그 참조무결성을 유지해야 한다	고객의 직업코드는 통합코드테이블의 직업코드에 등록된 값이어야 한다
	참조 무결성	테이블간의 컬럼값이 참조관계에 있는 경우 그 무결성을 유지해야 한다	대출원장의 대출원장번호는 대출 상세내역에 존재해야 한다
	데이터 흐름 일관성	데이터를 생성하거나 가공하여 데이터가 이동되는 경우, 연관된 데이터는 모두 일치해야 한다	운영계의 현재 가입 고객 수와 DW의 고객 수는 일치해야 한다
	컬럼 일관성	관리 목적으로 중복 컬럼을 임의 생성하여 활용하는 경우 각각의 동의어 컬럼 값은 일치해야 한다	주문의 주문번호와 고객번호는 배송의 주문번호와 고객번호가 서로 일치해야 한다

자료: 한국데이터베이스진흥원 2009

이준열·황철현(2014)는 전통적 품질 평가 및 활용적 관점에서의 품질 평가 등의 프레임워크를 제시한 바 있다. 전통적 품질 평가 프레임워크에서는 데이터베이스의 구조, 데이터의 생성 프로세스, 데이터의 관리 정책과 절차, 시스템을 데이터 품질에 영향을 주는 원인으로 언급하며, 평가 기준으로 정확성, 무결성, 일관성, 표현 일관성, 해석 용이성, 접근성 성능, 보안, 유용성, 적시성, 완전성 등 11개의 기준을 제시하였다. 활용적 관점에서는 활용자의 시점에서 데이터의 유무여부를 판단할 수 있는지, 데이터 접근 및 데이터 간의 상호통합이 가능한지 등에 초점을 두고 품질을 분석하기 위해 가시성, 접근성, 이해가능성, 신뢰성, 상호운용성, 반응성 등 6개의 기준을 제시하고 있다. 활용적 관점의 데이터 품질 평가는 기존의 데이터 관리 관점과는 다소 다르다. 과거 데이터 활용이 제한적으로 이루어지던 것에 비해 현재 데이터는 누구나 보편적으로 사용 가능하다. 따라서 데이터 품질의 기준도 데이터 관리자 및 엔지니어적인 사고방식으로 생산 중심의 데이터 관리에서 사용자의 편의성에 도움이 되는 요소들을 포함하고 있다. 이는 데이터 품질관리를 통해 양질의 데이터가 체계적으로 수집되었는가에서 부터 체계적으로 관리되고 있는 양질의 데이터를 사용자가 그 소재를 파악하고 활용하는 것이 용이한지까지도 고려하고자 함이다.

이 외에도 다양한 품질을 관리하기 위한 프레임워크 및 지표들이 있다. 기관의 특성 및 품질 평가의 목적에 따라 필요한 항목을 선정하여 평가 절차를 진행하고 그 결과를 반영하는 관리가 이루어질 수 있다. 한편, 공공기관 데이터 품질 수준을 측정하기 위한 기준으로 제시된 데이터 품질 지표는 다음과 같다.

〈표 II-7〉 데이터 품질지표

지표	지표 정의	세부 지표
준비성	- 공공데이터의 데이터관리를 위해 기본적으로 관리해야 하는 요소들을 정의하고, 이에 대해 지속적인 활동을 통해 최신화하고 있는지를 측정하는 지표	- 관리지표 - 내용충실
완전성	- 공공데이터의 저장소인 데이터베이스를 구축함에 있어 논리적인 설계와 물리적인 구조를 업무요건에 맞게 데이터가 저장되도록 설계·구축되었는가를 측정하는 지표	- 논리모델 - 식별자 - 물리구조 - 속성의미
일관성	- 엔터티의 속성이 정의된 표준(표준 용어, 코드, 도메인)을 준수하고 있는지와 속성의 중복사용 또는 데이터 연계에 일관성이 유지되고 있는지를 측정하는 지표	- 속성 - 표준 - 중복값 - 연계값
정확성	- 사전 품질확보를 위해 데이터 입력단계에 오류를 방지하기 위한 방법이나 로직(Logic) 등의 조치 수준, 저장된 데이터가 정의된 기준에 맞게 유효한 정보의 범위와 형식으로 되어 있는 수준, 상호 참조관계에 대한 무결성 확보 수준, 실제 입력 값이 업무규칙에 맞게 저장되어 있는지 등을 측정하는 지표	- 입력값 - 업무규칙 - 범위·형식 - 참조관계 - 계산식
보안성	- 지속적인 품질확보를 위해 운영되는 데이터의 생성관리 주체가 누구인지에 대한 관리수준, 권한에 따른 데이터 접근관리 수준, 정보 서비스의 중단 최소화를 위한 DB 보완관리 등의 수준을 측정하는 지표	- 오픈쉽 - 접근제한 - DB보호
적시성	- 사용자가 만족하는 수준의 응답시간으로 프로그램 성능이 확보 되었는지, 정보 요구사항 접수로부터 수집·처리되어 제공되기까지의 작업시간이 최소화 관리되며 요구된 정보가 최신의 것인가에 대한 수준을 측정하는 지표	- 응답시간 - 데이터제공 - 최신값
유용성	- 사용자가 만족하는 수준의 충분한 정보가 제공되고 있는가, 정보에 접근이 사용자의 편의성이 확보되었는가, 그래서 사용자가 정보를 유용하게 활용하고 있는가에 대한 수준을 측정하는 지표	- 충분 - 접근 - 활용

* 세부적인 체크리스트는 ‘공공데이터 품질관리 매뉴얼 : 데이터 품질 지표별 체크리스트’ 참조

자료: 한국정보화진흥원 2018

품질관리와 관련하여 국가에서 제시하고 있는 주요 지침은 공공데이터 관리지침, 공공데이터 개방표준, 공공기관의 데이터베이스 표준화 지침, 행정기관 코드표준화 추진 지침, 행정기관 및 공공기관 정보시스템의 구축·운영 지침, 정보기술아키텍처 도입·운영 지침 등 다양하다. 또한 KOICA는 공공기관으로 공공데이터 제공 의무가 있으며, 품질 평가 대상이기도 하다. 공공데이터법 제12조 및 공공데이터법 시행령 제13조에 의하면 품질 유지를 위한 관련 조직을 만들고, 담당자를 지정해야 할 의무가 있다. 관련 계획 수립 및 절차 이행에 있어 중복되거나 누락되는 사항이 없도록 상위 법령 검토 및 협력단 내 관련 절차 수립에 참고할 필요가 있다.

3) 데이터 생애주기 관리

데이터가 생성되면 이를 확보하여 저장하고, 분석 및 가공 처리하여 데이터를 이용하게 된다. 하지만, 생성되고 있는 데이터를 무한히 저장하고 보관할 수는 없다. 생성되고 있는 데이터는 언젠가 그 활용도가 종료될 것이다. 또한 개인정보 등의 민감한 데이터에 대해서는 일정 기한이 지나면 폐기하고 장기간 저장을 방지하는 별도의 관리체계가 있다. 즉, 기관에서 생성되고 있는 데이터 또한 일련의 기준에 따라 저장 및 관리되다가, 그 활용도가 없어지면 폐기하거나 특별히 보관하는 것이 가치가 있다고 판단되는 경우 별도의 장기간 보관이 가능한 저장소에 이관하여 관리하는 것이 합리적인 절차일 것이다.

기관에서 생성되고 있는 데이터에 대해 생성단계부터 활용이 종료될 수 있는 시점까지 판단하여 각 단계별 관리가 요구된다. 생성된 데이터는 활용 가능하도록 기본적인 가공이 필요하며, 사용자의 필요에 따라 추가적인 가공을 통해 활용될 것이다. 그리고 데이터의 종류에 따라 접근 가능 범위가 달라질 수 있으며, 이에 따라 기관의 내부에서의 접근성 혹은 외부로 공개 가능 여부 등이 나누어질 수 있다. 또한 사용 빈도가 현저히 감소하거나 활용이 거의 이루어지지 않는 데이터에 대해서는 별도의 저장공간 등에 이관함으로써 필요한 데이터 활용의 편의성을 확보하는 것도 중요하다.

결과적으로 데이터의 생애주기는 데이터가 어떻게 생성되어 어떻게 활용되고 더 이상 활용되지 않는 시점이 언제인지 예측해 데이터의 생성부터 폐기까지의 기준을 마련하는 것으로 생각해볼 수 있다. 데이터의 성격에 따른 등급이 마련되어야 하며, 이에 따른 보존 기간도 설정할 수 있다. 또한 등급에 따라 내부에서 열람 가능한 범위를 설정할 수 있으며 외부로의 공개 여부도 결정될 수 있다.

이와 유사한 관점에서 현재 문서 중심의 기록관리 체계가 있다. 물론 데이터 관리체계와 동일하지 않으며, 데이터와 기록물 관리의 최종적인 목적은 다르다. 많은 경우 기록관리는 보존 관점에서 기록물을 분류 및 저장한다고 인식되고 있으며, 데이터 관리는 디지털화된 값을 활용 관점에서 저장한다고 인식하고 있다. 하지만, 기록 또한 데이터를 기반으로 작성되는 다양한 형태의 데이터 집합으로 구분할 수 있다. 즉 기록과 데이터는 각 영역의 특성이 조금 다르게 나타나는 것이지 완전히 다른 개념은 아니라고 볼 수 있다. 따라서 기관 차원에서 효율적으로 방대한 양의 데이터를 분류하여 관리하는 측면에서, 그리고 기존의 기록물로 관리되고 있는 문서 또한 비정형 데이터라는 점에서, 데이터 생애주기 관리를 위해 기록관리의 분류 체계를 참고하는 것은 상당히

합리적인 선택이 될 수 있다. 현재 KOICA 기록관리분류 기준표는 정부기능분류체계(BRM; Business Reference Model)¹²⁾의 일부이며, 업무기능에 따라 기록물을 분류하여 관리하는 체계이다. 업무조직에 따라 분류되어 있어 데이터 거버넌스에서 데이터 흐름 분석에 따른 조직구조가 구축되어야 한다는 점과 유사한 맥락이다.

〈표 II -8〉 KOICA 단위과제 보존기간표(2021.2.기준)

한국국제협력단(KOICA)단위과제보존기간표											
2021. 2. 기준											
번호	기능분류							기능 유형	기록관리항목		
	처리과 (부서)명	(1레벨) 정책분야	(2레벨) 정책영역	(3레벨) 대기능	(4레벨) 중기능	(5레벨) 소기능	(6레벨) 단위과제(업무)		단위과제(업무) 설명	보존 기간	보존기간 책임사유
12	ODA 연구·정보 센터	통일· 외교	외교	개발협력	공적개발 원조	연구조사	KOICA 연구 기획 및 발굴	고유	한국국제협력단 의 연구과제를 기획하고 새로운 연구과제 를 발굴하는 업무	준 영 구	연구과제를 기획하 고 개발·발굴하는 업무로 업무참고 및 중요성을 고려하여 준영구로 책정
13	ODA 연구·정보 센터	통일· 외교	외교	개발협력	공적개발 원조	연구조사	개발협력 관련 조사 분석 및 연구	고유	개발협력 관련 조 사를 분석하고 연 구하는 업무	준 영 구	개발협력 관련 조 사 분석 및 연구와 관련된 업무로 업무 참고 및 중요성을 고려하여 준영구로 책정
14	ODA 연구·정보 센터	통일· 외교	외교	개발협력	공적개발 원조	연구지원	ODA 지식 공유 및 확산	고유	국제회의의 포럼, 컨퍼런스, 세미나 등 주요 연구결과 를 대내외로 공유 하는 업무	준 영 구	주요연구결과를 대내외로 공유하는 업무로 업무참고 및 중요성을 고려 하여 준영구로 책 정
15	ODA 연구·정보 센터	통일· 외교	외교	개발협력	공적개발 원조	연구지원	국내외 ODA 연구역량 강화 지원	고유	국내외 ODA 연구 역량이 강화될 수 있도록 지원하는 업무	5	ODA연구역량 강 화를 지원업무로 업무참고를 위해 5년으로 책정
16	ODA 연구·정보 센터	통일· 외교	외교	개발협력	공적개발 원조	공적개발 원조 통계	KOICA ODA 통계 정책수립 및 주요 보고	고유	KOICA ODA 통계 관련 이슈에 대응, 정책을 수립하고 OECD, 국조실 등 에 공식보고하는 업무	영 구	KOICA ODA 통 계를 위한 국제통 계기준과 공식보 고사항의 중요성 을 고려하여 영 구로 책정
17	ODA 연구·정보 센터	통일· 외교	외교	행정지원	기획	연보및통 계집발간	연보 및 통계 집발간	기관 공동	한국국제협력단 연보발간 및 통 계집발간과 연 관한 업무	영 구	기록물관리지침 의 보존기간 준 용
18	ODA 연구·정보 센터	통일· 외교	외교	개발 협력	공적개 발원조	ODA 도서관	ODA도서관 운영	고유	ODA 도서관 운 영에 관한 업무	준 영 구	ODA도서관 운 영 및 관리와 연 관한 업무로 준 영구로 책정

자료: 내부자료

12) 행정기관이 수행하는 모든 업무를 기능별, 목적별로 분류하여 단위과제별 업무의 기본정보(수행기관, 수행부서, 기능휴형 등), 속성정보(수행주체, 수행절차 등), 유관정보(법령, 지침, 규제, 예산 등), 업무편람(처리절차, 과정정보, 결과정보 등) 다양한 연계지식, 정보를 공유할 수 있도록 체계화된 정부기능분류 체계(서울시 홈페이지: opengov.seoul.go.kr/brm)

물론 데이터의 관리 범위는 기록물 관리 범위를 넘어서며, 기록물관리에서 다루지 않는 데이터가 훨씬 많다. 다만, KOICA는 공공기관으로 기록물관리가 법적으로 규정되어 있으며, 공공행정 오픈데이터 등 관련 국가 차원의 규정이 수립되고 있는 바, 별도의 데이터 생애주기 관련 지침을 새롭게 마련하는 것보다는 국가 관리 기준을 포용하는 관리체계 마련하는 것이 보다 효과적인 것으로 사료된다. 특히 분류체계 및 (보안)등급은 기록물을 포함하고 있는 데이터 관리 측면에서 혼선을 방지하기 위해 유기적으로 연동되어야 할 필요가 있다. 기관의 기존 기록물관리와 상충될 경우, 결국 데이터의 품질과 신뢰성 문제로도 연관되기 때문이다.

데이터 생애주기 관리는 결국 데이터 분류체계 및 등급을 마련하고 이에 따른 권한을 부여함으로써 등급별로 데이터 공개하고 활용 빈도 및 중요도에 따라 일정 기간 이후 보관하는 것을 체계화하는 과정이라고 볼 수 있다. 우선 사업 데이터의 경우 기관의 대다수 직원에 공개되어야 하므로 등급에 대한 고민이 당장은 불필요할 수도 있겠으나, 향후 경영관리 데이터 혹은 대외비 전자문서 등 기관의 전체 데이터를 고려한다면, 등급 기준 및 분류 과정에 대한 합의가 먼저 이루어져야 할 것이다.

Ⅲ. KOICA의 데이터 거버넌스 진단: 전략과 관리체계, 프로젝트 데이터 중심으로

1. KOICA 데이터 관리 현황 및 디지털 전환

1) KOICA의 주요 데이터 현황

데이터는 크게 정형 데이터와 비정형 데이터로 구분된다. KOICA에서 생성되고 있는 데이터는 개발협력사업 수행을 통해 얻게 되는 자료, 사업수행과정의 기록, 사업결과 및 평가결과 등으로 중심으로 기관 운영과정에서 발생하는 다양한 자료 및 정보 등이라고 볼 수 있다. 본 연구에서는 비정형 데이터를 문서, 흔히 값으로 언급하는 정형 데이터를 데이터로 언급하고자 한다.

일반적으로 데이터 중심 경영 또는 데이터 기반 정책결정은 구조화된 정형 데이터의 분석 결과에 기초한다. 따라서, 디지털 전환은 구조화된 정형데이터를 중심으로 논의되는 경우가 많다. 하지만, 다른 공공기관들과 마찬가지로 KOICA의 경우 문서 중심의 업무 프로세스를 감안하면 비정형 데이터를 제외하고 논의하기에는 한계가 많을 것이다. 정형 및 비정형의 구분도 중요하지만, KOICA 내부에서 관리하는 데이터가 디지털화가 되어 있는지도 중요하다. 실제로 데이터가 있다 하더라도 디지털화가 되어 있지 않다면, 결국 활용을 위해서는 별도의 가공 절차가 필요하기 때문이다.

이상적인 상황에서 데이터 관리를 위해서는 비정형 및 정형데이터의 생애주기를 면밀히 분석하고, 생산-저장-활용의 단계별 업무프로세스를 세밀히 정리하여, 그 논리체계에 따라 업무 관리 및 데이터 관리 시스템이 구축되는 것이다. 그러나 KOICA의 경우, 예산 등의 제약 상황 등으로 인해 각 사업별, 각 업무 분야별로 관리시스템이 각기 다른 시기에 개별적으로 구축되었다.

그 결과, 현재 KOICA의 데이터는 여러 시스템에 분산되어 관리되고 있다. 주요 업무영역을 살펴보면, 경영관리 측면과 사업관리 측면으로 나누어 볼 수 있다. 모든 업무를 나열할 수는 없겠으나, 대표적으로 몇 가지만 살펴보면 경영관리 측면에서의 업무영역으로는 인사관리, 경영기획 및 전략, 경영평가, 예산·재무관리, 규정관리, 정보관리, 계약관리, 기록관리, 국회대응관리,

연구, 홍보, 국내외 안전관리, 고객관리, 시설관리 등이 있을 것이다. 사업관리 측면에서의 업무영역으로는 사업별 통계, 국별협력사업, 민관협력사업, 연수사업, 다자협력사업, 인도적지원사업, 국제질병퇴치기금사업, 해외봉사단 및 국제개발협력인재양성, 교육사업, 사업평가 등이 있다.

업무영역의 상당 부분이 시스템을 통해 관리되고 있으며, 각 업무는 다수의 시스템과 연계되어 진행되고 있다. 전자결재의 경우 전 업무영역에서 활용하고 있으며, 예산재무 및 전자조달 또한 많은 비중을 차지하고 있는 주요 시스템이라 볼 수 있다. 이는 부서 간의 관계가 매우 긴밀함을 보여주는 부분이기도 하다. 또한 사업부서의 경우 통합사업관리(IPM, 봉사단, 사업평가)를 주요 시스템으로 볼 수 있다. 다만, 시스템 간의 연계 및 부서 간 문서 열람 등의 정보 공유 등 다소 제한이 있다. 따라서 데이터 관리 측면에서 정확성, 접근성, 완전성, 일관성 등의 문제를 갖고 있다고 볼 수 있다. 이 외에도 시스템을 갖추고 있으나 활용되지 않고 부서 내 별도로 관리되고 있어 담당자를 제외하고는 활용하기 어려운 데이터도 있다. 더욱이 각 부서별 파일로 관리하는 데이터의 경우, 최신 또는 최종 데이터가 무엇인지도 확인하기 어려워진다.

한편, 국별협력 프로젝트 사업을 중심으로 문서 및 데이터의 흐름을 살펴보면 다음과 같다. 우선 주로 생성되는 데이터는 비정형 데이터인 전자문서 형태이다. 모든 절차를 포함하지는 못하지만, 간략히 살펴보면 주로 생성되는 데이터는 비정형 데이터인 전자문서 형태이다. 또한 개발도상국의 사정을 감안하여 종이 문서 형태도 함께 존재하고 있다. 정형데이터인 디지털화된 값으로 존재하는 데이터는 대체적으로 사업조사 및 사업수행 결과 확인에서 생성되고 있다. 전자문서는 전자결재시스템에 자동적으로 저장되고 있으나, 종이문서의 경우 수집 및 저장이 온전히 이루어지지 않는 경우도 있다. 시스템 입력값으로는 현재 사업번호, 사업명 등의 기본 정보가 사업발굴 단계부터 사업관리(IPM) 시스템에 생성 및 입력되고 있다. 디지털화된 데이터값은 모든 단계에서 입력되고 있으나, 대체적으로 사업의 기본 정보이다.

현재 KOICA의 많은 데이터는 전자결재 시스템에 문서화되어 저장 관리되고 있다. 사업수행을 위해 통합사업관리 시스템, 조달시스템, 예산재무시스템을 활용하고 있으며, 일부는 문서로 업로드 되며 일부는 디지털화되어 값으로 입력되어 있다. 다만, 각각의 시스템은 연동이 제한적이며, 대부분 각 시스템에서 재작성되는 경우가 많다. 또한 각 사업수행으로 나타나는 결과 값은 모두 문서에 서술되어 있으며 원 데이터에 대한 관리는 별도로 이루어지지 않고 있다.

〈표 Ⅲ-1〉 부서 간 생성 및 관리 데이터

부서	생성 및 관리 데이터
인사	사업담당자 데이터
전략경영	사업예산 배분 데이터
사업전략	사업심의 결과 데이터
조달	사업 파트너 계약: 계약명, 사업자 정보(기업명, 사업자등록번호, 사업참여자 이력, 기관신용등급, 유사사업실적내역 등) 계약서, 계약이행증권, 각종 서약서 및 동의서
재무	사업비 지출: 계약명, 사업자 정보(기업명, 사업자등록번호, 은행계좌 등), 계약서, 계약이행증권, 사업 수행 결과물, 지출내역서 및 증빙서류)
사업	사업지역, 수원기관, 수혜자 정보, 사업분야, 원조목적, 정책마커, SDG, 사업유형, 지원규모, 수행기관 및 참여인력 정보, 조사 결과, 마일스톤 달성현황, 사업변경 횟수, 계약/약정 현황, 예산집행률, 성과지표, 사업평가 결과 등

자료: 내부자료 및 저자 작성

사업수행에 있어 기본적인 데이터를 살펴보면, 사업명/기간/예산/계약업체/집행결과 등의 데이터는 각각 생성된다. 계약과 관련된 데이터는 조달부서에서, 사업수행과 관련된 데이터는 사업부서에서, 사업 예산 배정 및 지출 관련 데이터는 예산·재무부서에서 각각 관리한다. 사용에 있어서는 각각의 시스템에서 데이터를 확인하여 활용도에 따라 가공하여야 한다.

다만, 현재 이들 데이터는 통합적으로 관리되고 있지 않으며, 시스템에 입력이 되지 않거나, 타 부서에서는 검색되지 않는 경우도 있다. 또한 시스템 외에 별도의 개별 파일로 관리되고 있는 데이터도 있다. 효과적인 관리 및 활용을 위해서는 공통의 표준을 마련하여 분산 관리되고 있는 데이터의 통합이 필요하다. 이 외에도 생성되고 있으나 문서 내에 존재하거나, 시스템에 입력되지 않거나 누락된 데이터를 수집하여 디지털화가 필요한 상황이다. 사용자 입장에서 데이터를 통해 추가적인 가치 창출이 가능하더라도, 수집되지 않거나 제대로 관리되고 있지 않다면 활용 가능성이 낮거나 없다고 볼 수 있다. 의사결정을 위해 분산되어 있던 정보 및 데이터를 가공을 하여 새로운 지식 및 관련 근거를 생산하였더라도 이는 다시금 재사용되는데 한계가 있기 때문이다. 따라서 데이터 흐름에 따른 업무의 흐름도 및 부서 간의 관계를 파악하고 이에 따라 데이터 관리 시스템 구축 및 통합이 필요하다.

2) KOICA 디지털 전환 전략

앞서 살펴본 바와 같이, KOICA는 디지털 전환 초기 단계로 데이터 자산에 대한 조직 차원의 인식 및 시스템 구축은 미흡한 상황이다. 데이터 생산, 수집 및 관리 지침 등 프로세스 부재로

데이터 공유 또는 활용이 어렵거나 많은 시간이 소요된다. 하지만, 내·외부적 데이터 활용의 필요성은 나날이 증가하고 있으며, 다양한 유형의 데이터를 원하는 방식으로 분석할 수 있는 환경이 요구되고 있다. 이를 위해서 전사적 차원의 동일한 기준에 의한 데이터 관리체계가 수립되어야 하며, 무엇보다도 IT 관련 부서 및 현업 부서 모두 참여하는 역할 및 책임 정립이 이루어져야 한다. 단계적으로 기존 업무방식과 시스템의 비효율성을 제거하고, 디지털 서비스와 신기술을 수용할 수 있는 인프라를 구축해야 하며, 업무 프로세스에 혁신이 필요하다.

더욱이 개발협력 현장에서도 데이터기반의 협력, 디지털 전환은 거스를 수 없는 요구가 되었다. 데이터가 새로운 핵심 자원으로 부상하며 데이터를 통해 협력국 내 개발수요를 파악하고, 정책을 수립하며, 사업을 수행한다. 나아가 데이터는 사업효과성의 과학적이고 객관적인 증거로서, 이를 활용하여 개발성과를 극대화할 수 있다. 또한 디지털 신기술을 활용하여 효율적이고 체계적인 업무가 가능해진다. 주요 국가 및 타 기관들 역시 데이터 경쟁력을 확보하며 디지털 전환을 추진 중이며, 우리 정부도 디지털 뉴딜을 선포했다. 이러한 상황에서 개발협력 대표기관으로서 KOICA 역시 전문성을 높이고, 효과적인 사업을 수행하기 위해서는 사업기획 및 성과관리에서 데이터 활용도를 높이는 디지털 전환의 노력이 시급하다.

이에 따라 KOICA는 21년 디지털 전환 전략을 수립하고 사업, 고객, 프로세스, 조직 역량 관점에서 4대 전략목표, 12대 전략과제 및 세부과제 33개를 도출하였다. 중장기 주요 목표로는 사업방식의 디지털 전환 및 디지털 중심의 경영기반 혁신으로, 2023년까지 프로세스, 시스템, 조직의 디지털 전환을 시행하며 데이터 거버넌스와 표준을 정립할 계획이다. 더불어 사업관리 시스템 고도화 및 최신화, 업무 절차 간소화와 효율화를 추진해가며 2025년까지는 전사 업무 및 분야에 디지털 전환이 내재화되는 디지털 KOICA로 나아갈 것을 목표로 한다. 이를 통해 디지털 시대에 맞는 개발협력 사업을 수행하고 그 성과를 높이고자 한다. KOICA의 디지털 전환 전략은 크게 사업, 고객, 프로세스, 조직역량으로 구분하고 있다. 하지만, 결국 데이터의 체계적인 관리 및 데이터 기반 의사결정을 위해서는 사업을 통해 생성되는 데이터를 어떻게 관리할 것인지, 그리고 이걸 내부적으로 어떻게 활용할 것인지, 외부활용을 위해서 어떻게 개방할 것인지를 고민하고 이를 한데 아우르는 거버넌스 수립이 먼저 이행되어야 할 것이다. 따라서 본 연구에서는 KOICA의 데이터 거버넌스 구축 및 실현을 위한 경영관리 및 시스템적인 측면에서 프로세스 전략을 중심으로 살펴볼 예정이다.

〈그림 Ⅲ-1〉 디지털 전환 전략 체계도

조직 비전	글로벌 사회적 가치를 실천하는 대한민국 개발협력 대표기관		
DT 비전	디지털 시대 개발협력 성과 제고를 위한 사업·경영의 디지털 전환		
전략 목표	A. 사업 디지털 기반 ODA 사업 확산	B. 고객 개발협력 분야 고객 맞춤형 서비스 제공	C. 프로세스 사업방식의 디지털 전환
전략 과제	1. 디지털 핵심사업 및 주류화 확대* 2. 개발협력 데이터 구축 및 개방 활성화	1. 사업 참여 고객 맞춤형 서비스 제공 2. 고객 친화적 디지털 참여 환경 조성	1. 프로세스 기반 데이터 관리체계 수립 2. 사업관리 시스템 고도화 3. 현장 중심 업무 활성화 기반 조성
전략 목표	D. 조직 역량 디지털 중심의 경영기반 혁신		
전략 과제	1. 효율·연계중심 경영관리시스템 구축 2. 디지털 신기술 활용 업무 효율화 3. 디지털 전환 추진기반 인프라 강화	4. 디지털 전환 성과 창출을 위한 조직 역량 강화 5. 디지털 체질전환을 위한 변화관리	

자료: 내부자료

3) 디지털 전환에 따른 데이터 관리 방향

2021년 디지털 전환 전략은 내부적으로 데이터 거버넌스를 정립하여 업무 프로세스와 시스템의 연계를 강화하고자 하며 나아가 디지털 환경을 재구축하는 것이다. 디지털 전환을 통해 업무 효율성 개선 및 고객 편의 증대, 나아가 ODA 사업에서의 디지털 요소를 강화할 것을 목표로하고 있다.

데이터 거버넌스는 기관 운영에 밀접하게 구축된다. 운영 전면에 걸쳐 연관되기 때문이다. 다만, 처음부터 모든 것을 아우르는 거버넌스 수립은 현실적으로 불가하기에 핵심적이고 작은 영역부터 수립하여 점차적으로 확대할 필요가 있다. KOICA는 개발협력사업을 통해 수집 가능한 데이터이자 주요 활용 가치가 있는 데이터와 밀접히 연관지어 이를 중심으로 데이터 거버넌스 구축이 필요하다. 이미 다양한 데이터가 지속적으로 생성되고 있으며 어딘가로 수집되고 있다. 따라서 이중 무엇을 수집하고 관리할 것인지에 대한 합의 및 이에 대한 표준 수립이 필요하다.

이에 그 첫 번째 단계로 KOICA 디지털전환탐을 중심으로 데이터 거버넌스 수립을 위해 핵심 데이터를 선정하고 IPM 개선 방향 등에 대한 여러 논의를 거듭하고 있다. 현재 KOICA 내 수집되고 있는 주요 데이터는 개발협력 사업을 수행하면서 생성되는 정보다. KOICA 8대 사업 유형인 국별협력, 국제기구협력, 시민사회협력, 기업협력, 봉사단, 연수, 인도적지원 민관협력, 질병퇴치기금 사업을 대상으로 재원 배분 현황 데이터, 사업 수행현황 데이터, 사업 성과 데이터,

사업 수혜자 및 사업 참여자 데이터가 해당된다.

수집 및 저장의 관점에서 본다면, 자원 배분 현황 데이터, 사업 수원기관 및 사업 참여자의 경우 대부분 IPM, 조달시스템 및 예산·재무 시스템을 통해 관리되고 있다. 사업 수행현황 데이터 및 사업성과 데이터의 경우 IPM에 보고서 형태로 수집 및 저장되고 있으며, 사업을 통해 얻을 수 있는 원자료는 별도로 관리되고 있지 않다. 하지만, 활용의 관점에서는 대부분의 데이터는 활용이 어려운 상태라고 볼 수 있다. 필요한 데이터에 접근부터 가공 후 활용까지 많은 노력을 요구하는 경우가 많다.

이중에서도 특히 IPM 활용도가 높은 국별협력사업, 즉 프로젝트에 대해서는 사업 주요 단계별 핵심데이터 입력 및 변경을 용이하게 하여 전자결재 등 시스템 간 연동성을 강화하고 사용자의 편의성을 제고하고자 한다.

KOICA의 데이터는 개발협력 사업 데이터가 핵심이나, 개발협력 사업 데이터는 수집, 관리하기 가장 난해하고 어려운 데이터이기도 하다. 따라서 KOICA의 개발협력 사업 데이터는 그 자체로도 충분한 활용가치가 있으면서 동시에 SDG17에서 강조하고 있는 통계역량강화 측면에서도 나름의 의미가 있을 것이다. KOICA 내부적 활용도 측면에서 사업현황 파악 및 사업 관리부터 개발협력사업의 영향력 평가 및 사업방향성 수립 등에 보다 수월하고 합리적인 데이터 기반 의사결정이 가능해지며, 이에 대한 요구는 나날이 증가하고 있다. 특히 KOICA는 공공기관으로 외부적으로도 기관이 보유하고 있는 정형 및 비정형 데이터에 대한 공개 요구와 기관의 성과에 대한 가시적인 증명 필요성이 높아지고 있는 가운데, 공공데이터 개방을 통해 데이터경제 및 국제개발협력 분야 저변확대에 크게 기여할 수 있을 것이다. KOICA는 디지털 중심 경영혁신의 일환으로 데이터 거버넌스 확립을 통해 관련 정책 및 규정, 기준을 마련하고 그 절차를 내재화할 필요성이 그 어느 때보다도 강조되고 있다.

따라서 다음에서는 KOICA의 개발협력 프로젝트 데이터를 중심으로 구체적으로 어떤 데이터가 존재하며 어떻게 관리 중인지 그 현황을 상세히 살펴보고, 데이터 거버넌스 구축을 위해 고려해야할 사항들을 함께 논의해보고자 한다.

2. 개발협력 프로젝트 데이터 및 수집현황

1) 개발협력 프로젝트 데이터

개발협력 프로젝트 데이터를 정의하고 데이터 유형을 분류하기 위해서는, 먼저 개발협력사업 프로젝트에 대한 정의를 내릴 필요가 있다. OECD/DAC(2013)는 개발협력 프로젝트에 대해 ‘정해진 기간, 예산, 지역 내에서 특정 목표 및 성과를 달성하기 위해 협력국과 협의된 투입물, 활동, 산출물을 종합한 사업 유형’으로 정의하였다. 또한 세계은행은 ‘일정 기간 이내에 특정한 개발 목적을 달성하기 위해 계획된 투자와 정책수단, 조직 및 그 외의 조치가 하나로 결합된 패키지’라고 정의한 바 있다(Baum and Tolbert 1985).

KOICA는 국별협력사업, 글로벌연수사업, 민관협력사업, 국제기구협력사업, 봉사단 파견사업 등 다양한 유형의 사업을 실시하고 있으나, 프로젝트 기반 사업의 경우 국별협력사업의 프로젝트형 사업¹³⁾과 개발컨설팅 사업¹⁴⁾에 국한된다(고요한 외 3인 2020, 8). KOICA 프로젝트는 [표 Ⅲ-2]과 같은 절차로 추진된다.

〈표 Ⅲ-2〉 KOICA 프로젝트 추진 절차

일정	단계	내용
n-2	사업 발굴	PCP 제출 (KOICA 사무소)
		(상시) 수원국 사업제안서(Project Concept Paper, PCP) 발굴 및 본부 송부(협조문)
	사업 형성	PCP 심사 (KOICA 본부)
		(매월 1회) PCP 심사 - 적격여부 판단, 보완 필요사항 및 부적격 사유 정리
		예비조사 (사무소 주도/ 본부 지역실 지원)
		(연중) 통과된 PCP 대상 예비조사 실시 및 사업집행계획 (Project Document, PD) 생산
		사업형성 결과보고 (공관/사무소)
		(11월) PD 중 우선순위 정하여 본부 송부(전문) - 공관 협의 및 수송기관 요청서 첨부 필수

13) 협력대상국의 경제 사회 발전 및 복지 향상의 기여 등 특정 개발목적의 달성을 위하여 여러 사업유형을 연계하여 종합적으로 지원하는 사업

14) 기술협력 중심의 국별협력사업으로 소프트웨어 중심, 제도 구축 지원에 특화된 사업

	사업 심사	사업 심사 (본부)	(12월~1월) PD로 사업심사위원회 개최 - 미선정 PD는 차년도 '후보사업 POOL'로 관리
n-1	사업 기획	시행계획 제출 (본부)	(1분기) 시행계획 외교부 제출 및 예산 요구
		집행계획 업데이트 (사무소)	(하반기) 심층기획조사 등 후속 조사 통해 PD 지속 업데이트 활용 - PD ver.1(예비조사보고서) → ver.2(기획조사보고서) → 사업집행계획
사업 기간	사업 수행	사업시행자 선정 및 사업 수행 (사무소/본부)	입찰공고(KOICA), 입찰제안서 제출(응찰업체), 기술평가 및 가격평가(KOICA), 계약 체결
		본 사업 실시 (사무소/본부)	사업 실시
종료 평가 (종료 후 1년 내)		모니터링 및 평가 (사무소/본부)	사업 모니터링(상시), 연차점검(매년 초) 및 종료보고/ 종료평가
종료 후		사후관리 (사무소)	(필요시) 종료사업에 대한 사후관리 실시

자료: KOICA 2018, 38-40, 고요한 외 3인 2020, 9 종합하여 작성

KOICA에서 개발협력 프로젝트 데이터는 위의 프로젝트 추진 절차를 거치는 동안 수집, 생산, 관리, 모니터링하는 모든 데이터라고 통칭할 수 있다. KOICA 프로젝트는 변화이론을 바탕으로 프로젝트 디자인 매트릭스(Project Design Matrix, PDM)¹⁵⁾에 근거하여 기획 및 관리된다. PDM에서 제시하는 투입물(inputs), 활동(activities), 산출물(outputs), 성과(outcomes), 영향(impacts)을 측정하는 객관적 검증지표 값이 대표적인 프로젝트 데이터이다.

15) 고요한 외 3인 2020, 10 <표 3> PDM 예시 참조

〈표 Ⅲ-3〉 프로젝트 데이터 종류

프로젝트 데이터	내용
투입물 데이터	사업시행을 위해 투입하는 예산, 인력 자원
활동 데이터	사업을 통해 실제 수행하는 활동
산출물 데이터	사업 활동을 통해 직접적으로 도출할 수 있는 재화 또는 서비스를 측정하는 지표값
성과지표 데이터	활동목표, 또는 산출물로 발생하는 행동변화 상태를 측정하는 지표값
영향지표 데이터	사업이 기여할 수 있는 전략적 목표 달성 여부를 측정하는 지표값

자료: 고요한 외 3인 2020, 11 바탕으로 저자 작성

또한 [표 Ⅲ-3]의 프로젝트 데이터 외에 각 지표 값의 출처를 나타내는 검증 수단, 즉 해당 데이터가 언제 어떠한 방법으로 수집되었으며 누가 제공하는지, 책임자가 누구인지에 대한 정보도 해당 데이터의 품질, 신뢰성 및 적절성을 보장하고 추적관리에 필요하다는 점에서 중요하게 관리되어야 한다(고요한 외 3인 2020).

프로젝트는 여러 데이터를 생산한다. 본 연구는 체계적인 분석을 위해 프로젝트에서 생산되는 데이터를 일반관리 데이터와 통계 데이터, 그리고 성과 데이터로 구분한다.

(1) 일반관리 데이터

프로젝트의 원활한 관리를 위해서 기본적으로 10개의 지식영역에 대한 관리가 필요하다(Project Management Institute 2021). 프로젝트 생애주기별로 통합관리, 범위관리, 일정관리, 예산관리, 품질관리, 인적자원관리, 커뮤니케이션관리, 위험관리, 조달관리, 이해관계자 관리범위, 일정, 예산, 조달(계약관리), 인사, 변경관리, 위험관리, 이해관계자 관리의 10개 영역별 관리가 필요하다. 이렇게 프로젝트의 일반 관리과정에서 생산되는 데이터를 프로젝트의 일반관리 데이터라고 할 수 있다. 이러한 행정적인 관리에 필요한 데이터를 일반관리 데이터라고 할 수 있다. [표 Ⅲ-3]에서 제시된 프로젝트 데이터 중 프로젝트 관리통제 내에 있는 투입물, 활동 데이터 및 산출물 데이터의 일부가 이에 해당한다(KOICA 2018, 275). KOICA는 공공기관으로서 국내에서 상위기관 및 감사기관과 국회 등의 감사와 평가를 받는다. 대부분의 경우, 외부 책무성을 위해 관리되는 데이터는 위의 10개 영역별 관리 데이터와 중복된다. 따라서 동 연구는 이러한 국내 책무성을 위한 관리 데이터도 일반관리 데이터로 포함하여 분석한다. 일반관리 데이터는 잘 관리된다면 프로젝트 관리뿐만 아니라, 국회, 감사원, 외교부 및 기재부

대응 시에도 용이하게 활용될 수 있으며 OECD, IATI 통계 보고 시에도 별도의 수고 없이 추출되어 제출될 수 있다.

(2) 국제 통계 데이터

KOICA는 매년 2차례 OECD 개발원조위원회(DAC)에 ODA 통계(잠정, 확정)를 보고하며, 익년도에 전전연도 확정통계 수치가 OECD 통계 홈페이지¹⁶⁾를 통해 공개된다. KOICA 자체적으로는 ODA 세부실적 잠정 수치는 매년 10월 이후 국·영문 통계집¹⁷⁾으로 발간되며, 확정 수치는 익년 초 KOICA 대표 홈페이지¹⁸⁾ 및 통계조회서비스 사이트¹⁹⁾를 통해 공개된다. 이렇게 대외적으로 보고 및 공개되는 데이터를 통계 데이터라고 한다. 통계 데이터에는 주로 투입물 데이터와 활동 데이터가 포함되며, 구체적으로 프로젝트 대상 국가 및 지역, 분야, 원조목적(CRS) 코드, 관련 SDG, 정책마커, 사업유형, 지원규모, 수원기관 등의 정보가 이에 해당한다. 통계 데이터의 경우 OECD DAC 통계 가이드라인에 맞추어 공개되기 때문에 상당 부분 표준화되어 있으며, 모든 사업부서가 주기적으로 업데이트하고 있다. 그럼에도 불구하고 같은 건의 사업에 대해서도 시스템 연동이 되지 않으며, 데이터 및 용어가 표준화되지 않아 종합적인 관리에 어려움이 있으며, 중복관리의 비효율성이 발생한다.

(3) 성과 데이터

성과 데이터는 개발협력 사업이 생산해낸 성과를 추적하고 관리하는 데이터를 의미하며, KOICA도 다른 공여기관과 마찬가지로 논리모형(logical framework)에 기초한 성과를 (results framework)로 관리한다. 사업추진 현황 및 성과 모니터링에 필요한 지표로서 [표 Ⅲ-3]의 산출물 데이터, 성과지표 데이터와 영향지표 데이터를 포괄한다. 프로젝트 기획 단계에서부터 목표 성과를 측정하기 위한 적정 지표를 선정하고 해당 지표에 대한 정의, 측정 단위, 측정 시기 및 빈도, 검증 수단(출처, 책임소재 등)을 명확히 해야 한다. 또한, 기초선 조사를 바탕으로 설정한 연도별, 혹은 프로젝트 단계별 목표치와 실제 측정치를 함께 제시하여 프로젝트의 체계적인 관리를 지원해야 한다. KOICA는 이미 2009년 프로젝트 성과관리를 위한 방법론을 제시한 ‘프로젝트 기획, 모니터링 및 평가 방법론’을 출간하고, 논리모형을 활용한 성과 기획 및 모니터링과 관리 방법을 체계적으로 아래와 같이 제시한 바 있다.

16) stats.oecd.org

17) (국문) KOICA 대외무상원조실적 통계, (영문) KOICA annual statistical report

18) koica.go.kr

19) stat.koica.go.kr

〈표 Ⅲ-4〉 PDM의 구조와 내용

Narrative Summary	Objectively Verifiable Indicators(OVI)	Means of Verification(MOV)	Assumptions
Goal(목표) 프로젝트가 기여하는 섹터수준 또는 국가수준의 장기적 개발목표 작성방법 : 장기에 걸쳐 이미 성취된 것처럼 과거시제 동사로 시작	상위수준 파급효과 즉 목표수준을 측정하기 위한 질문이나 지표 등의 방법 (수량, 질, 시간 포함)	필요정보 수집 방법 : 데이터 수집 주체, 빈도, 정보 수집원(Source) 기입	프로젝트의 장기적 지속 가능성을 위해 충족되어야 할 가정
Purpose(목적) 프로젝트의 직접적 목적, 프로젝트의 직접적 결과에 의한 (수혜자그룹의) 행동, 태도, 자원 등의 측면의 통합적이며 측정가능한 변화 작성방법 : 이미 성취된 것처럼 과거 또는 현재시제 동사로 시작	각각의 목적 달성도를 측정하기 위한 질문이나 지표 등의 방법 (수량, 질, 시간 포함)	필요정보 수집 방법 : 데이터 수집 주체, 빈도, 정보 수집원(Source) 기입	적에서 목표로 연계되기 위해 충족되어야 할 가정
Outputs(산출물) 목적달성을 위해 프로젝트가 산출해야하는 결과물, 서비스 또는 생산물 작성방법 : 이미 성취된 것처럼 과거 또는 현재시제 동사로 시작	산출물 생산 정도를 측정하기 위한 질문이나 지표 등의 방법 일반적으로 개발행위의 결과물임 (수량, 질, 시간 포함)	필요정보 수집 방법 : 데이터 수집 주체, 빈도, 정보 수집원(Source) 기입	산출물에서 목적으로 연계되기 위해 충족되어야 할 가정
Activities(사업구성요소/ 개발활동) 산출물 생산을 위해 프로젝트가 해야 하는 활동들 작성방법 : 생산해야하는 것을 명사형으로 표현	Inputs(투입/ 투입자원) 개발행위의 결과는 산출물 수준에서 측정되므로 이 칸에는 투입요소를 기입		개발행위에서 산출물로 연계되기 위해 충족되어야 할 가정

자료: 박수영 2009

대부분의 공여기관은 논리모형에 기초한 성과관리를 하고 있다. [표 14]와 [표 15]가 보여주듯 USAID의 성과지표 워크시트나 KOICA의 간이 M&E 매트릭스는 각 성과목표와 지표가 유사하게 관리되고 있다.

〈표 Ⅲ-5〉 간이 M&E 매트릭스(예)

1. 프로젝트성과부문				날짜 :		
프로젝트목표	Indicators	Baseline Information	MoV	모니터링 담당자	프로젝트 Target	모니터링 결과
Output						
Activities						
Input						
2. 프로젝트관리부문						
관리분야	1차년도 목표	2차년도 목표	X차년도 목표	Indicators /MoV	모니터링결과	대응계획
예산						
일정/범위						

자료: 박수영 2009

〈표 Ⅲ-6〉 USAID 성과지표 워크시트

Performance Indicator	Detailed Definition	Unit of Measure	Data Source(s)	Data Disaggregated By	Data Collection Method(s)	Data Analysis Method(s)	Frequency and Schedule	Party Responsible for Timely Data
Enter the results statement for this result here								
Enter Indicator								
Enter Indicator								
Enter Indicator								

자료: USAID 웹사이트²⁰⁾

성과 데이터는 정책 및 전략 수립, 프로젝트 기획 및 수행, 성과 모니터링 및 평가, 학습 및 개발성과 제고를 위한 적용에 이르는 프로젝트 전 단계에 걸쳐 증거에 기반한 합리적인 의사결정에 활용될 수 있다. 그러나 아직 KOICA의 성과 데이터를 수집하고 정리하는 절차 및 기준에 대한 합의가 미비한 상황이다. 또한, 성과 데이터가 디지털화된 데이터가 아닌 문서 형태의 연차보고서 또는 종료 보고서에 수록되어 있어 데이터 분석 및 이에 기반한 의사결정이 용이하지 않다는 문제가 있다.

USAID(2020)는 프로젝트 데이터에 대해 서면으로 작성된 보고서에 기술된 것 이상의 추가 분석을 가능케 함으로써 프로젝트 기획 및 수행에 대한 정보를 제공하며, 기초선과 상황별

20) [usaid.gov/project-starter/program-cycle/pmp/performance-indicator-elements/pmp-performance-indicator-worksheet](https://www.usaid.gov/project-starter/program-cycle/pmp/performance-indicator-elements/pmp-performance-indicator-worksheet)

데이터(contextual data)를 통합함으로써 프로젝트 계획, 성과 측정 및 평가, 적용 시 활용 가능하며, 데이터 시각화를 통해 프로젝트에 대한 인식 증진 및 교육 목적으로 추가적인 통찰력을 제시할 수 있다고 인식하였다. 이에, 데이터의 원활하고 효율적인 활용을 위해 USAID의 자금으로 집행되는 모든 사업의 수행과정에서 발생하는 보고서, 자료 및 데이터를 웹기반 저장소인 Development Experience Clearinghouse(DEC)²¹⁾와 Development Data Library(DDL)²²⁾에 보관하고 있다. DEC는 사업 수행과정 중 작성 및 취득되는 보고서, 교육자료, 정책자료, 평가자료, 계획문서 등 텍스트 형태의 정보를 관리하고 있으며, DDL은 USAID와 수행파트너로부터 생성되고 수집된 정형데이터, 데이터 세트 등 기계판독이 가능한 비독점형 포맷의 데이터를 보관한다. 즉, 앞서 일반관리 데이터 부분에서 지적한 바와 같이 별도의 데이터나 기록 저장시스템이 부재한 것이 사업성과관리에도 부정적인 영향을 미치고 있다.

노르웨이 개발협력청(Norwegian Agency for Development Cooperation, Norad)의 경우, KOICA와 마찬가지로 통계 데이터는 OECD 기준에 따라 일관성 있게 보고되는 반면, 성과 데이터는 개별 프로젝트에 따라 형식이나 내용에 큰 격차를 보인다는 점을 과제로 인식하였다. 개발성과에 대해 공공에 투명하게 개방한다는 목적 하에 개발원조 정보포털²³⁾을 개설하였고, 가이드선스 및 템플릿 등 표준화된 체계를 마련하여 그 안에서 프로젝트 성격에 따라 유연하게 정성적, 정량적 성과 데이터를 보고할 수 있도록 하였다. 동 포털에서는 국가 단위 기초정보에 해당하는 정형/ 비정형 데이터를 일괄 제공하며 성과 데이터를 보유하고 있는 사업을 조회할 수 있다. 재생 에너지와 기초 교육 분야 프로젝트부터 성과 데이터를 공개하며, 점진적으로 공개 범위 확장을 계획하고 있다. 그러나 여전히 정량/정성적 기대효과 및 성과 점검 현황이 모두 텍스트 형태로 기재되어 있으며, 데이터 입력이 자동화되어 있지 않다는 한계점을 가지고 있다(OECD 2021).

2) KOICA 프로젝트 데이터 수집 현황

KOICA는 프로젝트 단계에 따라 생산되는 문서 및 데이터를 통합사업관리시스템(IPM)에 업로드하여 관리하고 있다. 각 단계별로 수집 데이터 및 관리 현황을 살펴보면 다음과 같다.

21) <https://dec.usaid.gov>

22) <https://data.usaid.gov>

23) resultat.norad.no/en

(1) 사업발굴: 사업제안서(PCP)

PCP 등록 탭에서 게시된 정보를 보면 사업의 첫 단계인 PCP 단계에서 작성될 수 없는 사업 기본정보가 상당히 많이 기입되어 있다. 이는 해당 사업의 첫 화면에서 사업의 기본정보를 최대한 많이 확인할 수 있도록 하려는 의도에서 비롯된 것으로, PCP 단계에서 작성 불가능한 기본정보는 사업이 진행됨에 따라 담당자가 계속해서 업데이트한다.

PCP 단계에서 수집 및 작성되는 데이터는 첨부파일 문서인 수요요청서, PCP에 텍스트 형태로 수록되어 있다. 수요요청서 송부 시기, 구체적인 수요요청 사업내용 등이 이에 해당한다.

(2) 사업형성: 예비조사-사전조사 결과보고서

PCP 단계에서 조회되던 기본정보는 이후 모든 사업단계에서 사업 기본정보 항목으로 사업번호, 사업명, 사업목적, 사업내용을 확인할 수 있다. 사업명의 경우, 효과적인 데이터 추출 및 관리를 위해서 사업명, 사업기간, 사업예산이 분리되어 각각의 데이터로 입력될 필요성이 있다.

예비조사 단계에서는 조사와 관련한 정보가 입력되는데, 조사단 구성인력(구분/성명/소속/직위/업무/분야)을 파악할 수 있도록 디지털화된 데이터값으로 입력하게 되어 있다. 다만, 미입력되어 있거나 첨부문서 상 텍스트로만 확인 가능한 경우도 종종 찾아볼 수 있다. 효율적인 책임소재 파악 및 관리를 위해서는 동 정보가 데이터로 입력될 필요가 있다. 또한, 예산·재무 시스템과 연동하여 소요된 비용을 실시간 데이터로 확인할 수 있어야 한다. 핵심 문서인 예비조사 결과보고서의 경우, 개발도상국의 개황을 파악하기 위한 주요 지표 및 현황 등의 자료를 포함하고 있는데, 동 데이터의 경우 기초선 데이터 설정을 위한 초기 데이터로서 추출되어 관리될 필요가 있다. 이 외에도 기자재 목록, 예산안, 조달 계획 등의 협의 자료는 문서를 분리하여 각각 관리하는 것이 사업의 용이성을 높여줄 수 있을 것이다.

기획조사 단계에서 기획조사보고서를 작성하기도 하나, 사업에 따라 사전조사 결과보고서 혹은 기획조사보고서를 택일하여 작성하기도 한다.

(3) 사업기획: R/D 체결

R/D 체결 시 기본적으로 입력해야할 데이터는 R/D 서명 일자, KOICA 서명자(부서, 서명자명), 수원국 서명자(소속/직위, 서명자명), R/D 주요내용(전체, 대한민국, 수원국)이다. 주요

내용에는 KOICA와 수원국측 분담사항을 기재하도록 한다. R/D 상의 사업기간, 예산, 사업내용을 데이터로 입력하여, 사업 변경 시 기존 R/D와 상이한 부분을 직관적으로 확인하여 할 수 있도록 할 필요가 있으며, 향후 R/D 변경 체결까지도 확인할 수 있도록 개선이 필요하다.

(4) 사업기획: 집행계획

집행계획에는 사업 기본정보로 사업대상지역, 수원기관, 수원총괄기관, 사업유형, 원조구분, 지원분야, 건축·시스템 포함여부 등이 작성된다. 첨부파일인 사업집행계획 문서에 대부분의 주요 정보가 포함되어 있으며, 사업시행자 선정계획 내용 또한 포함되어 있다. 그러나 집행계획 단계에서 사업시행자 선정계획을 필수 업로드 문서로 지정한 만큼 이 부분을 따로 분리하여 첨부하는 것이 조달과 관련한 프로젝트 관리에 더 용이할 것으로 판단된다. 이에 더하여 프로젝트 관리 대상인 예산, 일정, 범위, 성과와 관련된 내용 또한 분리하여 관리하는 것이 필요하다. 예산 세부내역, PDM은 필수 첨부문서로 저장된 바, 사업 일정과 범위 문서 또한 필수 입력 문서로 지정이 필요하다.

집행계획에는 사업 대상 지역의 현황 파악 및 기초선 데이터 구성에 참고할만한 데이터를 포함하고 있기에 이를 추출하여 관리할 필요가 있다. 또한, 기본정보 상의 사업기간, 총 사업비와 첨부파일 내 동 정보의 숫자가 다른 경우가 종종 발견되는데, 이는 기본정보는 최신으로 업데이트되는 데이터임에 반해 집행계획문서는 작성 시점의 내용으로 업로드되기 때문이다.

(5) 사업수행: 점검 및 평가-연차점검

사업 진행에 따라 연차별로 점검이 이루어지며, 이때 제출된 연차보고서를 IPM에 업로드하게 된다. 연차보고서는 사업 수행 기관에서 제출한 것과 KOICA의 연차점검 보고서가 각각 존재하는데, 점검 사항은 각 보고서 내 문서로만 기입되어 있다. 또한, 연차점검 시기별 보고서 및 주요 내용의 직관적인 확인이 어렵다. 따라서 연차보고서 내 산출물 별 지표 및 조사내용을 파악하기 위해서는 이를 추출하여 관리하여야 한다. 각 보고서 내용의 차이점 혹은 향후 계획의 실제 이행 여부 등 파악하기 쉽게 해당 내용의 입력란이 마련되어야 하며, 기술된 지표 및 수치가 누락되지 않고 입력될 필요가 있다.

(6) 사업수행: 점검 및 평가-종료보고

종료보고에서 제출된 문서에서 비교적 자세한 성과지표 데이터를 확인할 수 있으며, M&E

매트릭스와 PDM을 통해 성과지표 선정 배경 및 논리에 대해 파악할 수 있다. 하지만 여전히 이 지표의 데이터가 신뢰할 만한 데이터인지 출처가 분명히 제시되지 않았다는 한계가 있다. 또한 이전 단계에서 성과지표 선정과 기초선, 중간선 데이터 수집에 대한 내용을 확인할 수 없다가 종료 단계(종료보고 및 종료평가)에서 모든 데이터가 제시되었다는 점에서 신뢰하기 어려운 측면이 있다.

(7) 사업수행: 점검 및 평가-종료평가

먼저 종료평가의 경우도 별도의 용역계약을 체결하여 실시하기 때문에, 평가단 구성 인적 정보 및 평가기간이 평가 기본정보 데이터로 입력될 필요가 있다.

앞서 사업종료 단계에서 언급했듯이, 성과지표가 어떻게 선정되었고 데이터가 어떻게 수집되었는지에 대한 정보를 IPM 상 확인하기 어렵다. 보다 상세한 성과 평가를 위해서는 데이터가 세부적으로 수집될 필요가 있다. 보건 사업의 예를 들면, 초진 외래환자 수나 타 병원에서의 후송환자 수의 경우 인원 총합이 아닌, 진단병명, 성별, 연령별, 거주 지역별로 세부적으로 분리하여 수집해야 보다 의미있는 성과평가가 가능할 것으로 보인다. 또한 사업 수행기관 혹은 별도로 파견되는 모니터링단은 데이터 제출 시 기존의 보고서 문서 내에 수록하는 것이 아니라 별도의 엑셀파일 혹은 데이터 입력 시스템에 원자료(raw data)의 형태로 제출해야 하며, 조사 대상의 범위, 데이터 수집 방법론과 책임 소재 등을 명확히 기재하여야 한다.

(8) 사업변경 이력

사업변경 이력은 IPM의 사업변경 탭에서 해당 사업을 검색하면, 변경일자, 변경구분(사업기간, 사업비, 사업명), 변경전과 후, 변경사유, 변경자에 대한 데이터를 확인할 수 있다. 그러나 변경사항이 발생한 단계에서 해당 내용을 직관적으로 파악할 수 없다는 점에서 관리에 어려운 측면이 있다. 사업관리의 효율성을 위해 각 단계에서 변경이력 추적이 용이하게 개선할 필요가 있다.

사업이력의 경우 현 IPM 상에서 PCP 등록, 집행계획 수립 탭에서 사업이력 버튼을 클릭하면 PCP-예비조사-기획조사-R/D 체결-집행계획에 이르는 단계별 간단한 정보를 확인할 수 있다. 그러나 관련 문서를 비롯한 상세한 내용 파악을 위해서는 결국 사업단계별 탭마다 들어가 사업번호 또는 기타 기본정보를 입력하여 해당 사업을 재조회해 일일이 확인해야 하는 구조를 가지고 있다. 이에 해당 사업을 한번 검색하면, 해당 단계 탭 클릭만을 통해 단계별 데이터를

파악할 수 있도록 시스템 개선이 필요하다.

또한 데이터 입력자 외에 정확한 데이터 값이 입력될 수 있도록 데이터 품질 검토 절차가 필요할 것으로 사료된다.

3) KOICA 프로젝트 데이터 분석

본 장에서는 프로젝트 사업을 중심으로 KOICA의 데이터 관리 현황을 살펴보았다. KOICA 프로젝트의 경우, 각 시스템 상의 유기적인 연계 부족으로 인해 동일한 정보값이 시스템별로 상이하게 입력되거나, 동일한 정보값을 여러 차례 반복적으로 입력해야 하거나 하는 문제가 있음이 확인되었다. 데이터를 활용한 분석을 위해서는 데이터가 동일한 규칙과 양식에 따라 저장되는 것이 가장 중요하며, 이는 품질관리의 기초가 된다. 따라서 각 시스템별로 반복적으로 동일하게 입력되는 정보값들을 확인하고 이 값들을 표준화하여 입력하도록 하는 개선이 필요하다. 주요 정보의 경우 입력 시 자율성을 최소화하기 위해 드롭박스 기능을 활용하는 등 시스템 디자인 자체가 통일성을 갖도록 지원해야 한다. 일반관리 데이터의 경우, 프로젝트 사업 추진에서 핵심정보가 되는 지역정보와 사업정보들을 사업시작 단계에서 입력 후, 이후 단계에서는 입력된 정보가 초기값으로 그대로 나타나고 수정이 되는 경우는 적절한 이력 관리와 변경관리가 시스템 상으로 이루어질 수 있어야겠다. 국제통계의 경우, IATI와 OECD/DAC에 대한 보고항목²⁴⁾이 보고 형식대로 입력되고 이후 동일한 이력관리와 변경관리가 이루어지도록 해야겠다. 다만, 사업성과 데이터의 경우, 사업별로 상이한 성과의 표준화가 유의미한 집합적 정보로 작동하기 어렵다는 점을 고려하여, 엑셀형식의 논리모형과 논리모형의 근거가 되는 데이터들을 필수 첨부로 저장하게 하계하여 한글문서가 아닌 엑셀문서로 관리할 것을 제안한다. 분기보고서나 종료보고서의 첨부가 아니라, 시스템상의 성과관리 항목에서 통합적으로 관리하게 하여, 엑셀문서라 하더라도 변경 및 이력관리가 되도록 하는 것이 중요하겠다.

24) 자세한 항목은 부록1 참고

IV. KOICA 데이터 거버넌스 구축을 위한 방향성 제안 및 결론

데이터 거버넌스는 데이터의 성격과 활용 방향성에 따라 누가 어떤 역할과 책임을 갖고 어떤 기술과 솔루션을 통해 재현해 낼 것인지에 대해 정책과 규정을 마련하고 로드맵을 그리는 과정이라고 볼 수 있다. 본 보고서에서는 데이터 거버넌스의 주요 요소로 표준화, 품질, 생애주기를 언급하였으며, 생성하고 있는 데이터의 성격과 활용 방향성을 고려하여 표준화 지침을 마련하는 것이 급선무라 판단하였다.

KOICA의 데이터 관리는 표준화 시작 단계라고 볼 수 있다. 데이터의 현황을 파악하기 위해 본 연구에서는 현재 KOICA 업무 분야와 업무 시스템을 임의로 구분하여 살펴보고, 국별협력사업을 중심으로 부서간 관계 및 생성 데이터도 살펴보았다. KOICA의 업무 분야를 일부 살펴본 결과, 다양한 시스템을 활용하여 관련 데이터를 수집 및 저장관리 하고 있으나, 체계적이라고 하기엔 다소 미흡한 부분이 많이 발견되었다. 현재 KOICA는 많은 데이터를 생성하고 수집하고 있으나, 이에 대한 합의된 표준은 미비한 상황이며, 다수의 시스템에 중복 수집 및 저장되고 있다. 더욱이 중복되는 데이터의 상당 부분은 표준화가 되어 있지 않기에, 이에 대한 가공 작업이 추가로 필요한 상황이다. 또한 품질 및 생애주기 관련하여 가이드라인 및 규정 마련이 요구된다.

데이터의 중요성이 강조되며 이를 자산으로 인지하고 관리할 필요성이 대두되고 있는 가운데, 2021년 디지털 전환을 발표하며 주요 과제 중 하나로 데이터 관리 체계 수립을 선정하였다. KOICA의 경우, 기관 내에서 활용될 데이터와 대외적으로 공개해야 하는 데이터가 있다. KOICA는 공공기관으로 공공데이터 포털에 개방해야 할 의무가 주어져 있어, 현재는 OECD DAC에 보고하는 데이터를 중심으로 공공데이터로 공개하고 있다. 데이터 거버넌스에 따라 표준화, 품질 및 생애주기 관리가 이루어진다면, 대내적으로는 업무 효율화를, 대외적으로는 데이터의 체계적인 공개 및 이를 활용한 국제개발협력 저변 확대까지도 기대해 볼 수 있을 것이다.

1. KOICA 프로젝트 단계별 핵심 문서 및 데이터 제안

KOICA는 프로젝트 단계별로 다양한 문서 및 데이터가 생성되고 있으며, 이를 IPM을 중심으로 전자결재시스템, 조달시스템, 예산재무시스템 등에 수집 및 저장하고 있다. 다만, 앞서 살펴본

바와 같이 시스템 간의 연계가 한정적이며, 각 담당자별 다른 기준에 따라 수집되고 있다고 볼 수 있다. 이는 결과적으로 다수의 정형 및 비정형 데이터가 많이 있으나, 실제 활용 관점에서는 가치가 있다고 보기 어렵다.

현재 IPM을 통해 수집 및 저장되고 있는 프로젝트 단계별 필수 문서 및 데이터 현황을 확인해본 결과, 데이터는 상당하지만 효율적인 관리가 필요함을 드러냈다. 사업의 기본 정보 데이터이자 OECD DAC 보고대상인 통계 데이터의 경우 상대적으로 일관된 기준 하에 관리가 이루어지고 있었으나, 그 외의 사업 단계별 핵심 문서 및 데이터의 경우 특히 이를 검증하는 단계가 미흡한 것으로 나타났다.

KOICA 프로젝트 데이터는 전체 데이터 중 일부로 많은 종류의 데이터가 문서 혹은 값의 형태로 생성되고 있다. 이를 토대로 앞서 살펴보았던 데이터 거버넌스 요소들을 개발협력 프로젝트를 중심으로 살펴보면 다음과 같다.

우선, 표준화 차원에서는 각 단계에서 입력되어야 하는 데이터를 정의하고, 입력 방식에 대해 규칙을 수립해야 한다. 예를 들어 사업발굴 단계에서 수요요청 사업구성 요소를 데이터로 지정할 경우 관련 문서를 저장하는 것 외에도 시스템 상 각 요소를 입력할 수 있는 칸이 마련되어야 할 것이며, 표준 용어를 활용한 입력 규칙까지도 지정해야 한다. IPM을 통해 많은 문서 및 데이터는 어느 정도 표준화되어 입력 및 저장되고 있다. IPM은 각 사업별로 일정한 문서 및 데이터를 입력하도록 틀을 제시하고 있으며, 일부 조달시스템 및 재무시스템과 연동되어 다양한 정보를 확인 가능하도록 하고 있다. 하지만, 디지털화된 데이터는 아직 미흡한 단계이며 이에 대한 표준화는 걸음마 단계라고 볼 수 있다. 위 표에서 제안한 핵심 문서와 데이터가 기관 내 합의된 것은 아니며, 기존에 저장되고 있는 문서를 바탕으로 데이터로 입력 가능한 값을 제안한 것이다. 따라서 본 연구에서 제안하는 핵심 문서 및 데이터에 대해 검토하여 최종적으로 수집 및 저장·관리할 문서 및 데이터를 정의하고, 이에 관련된 값의 입력 방식에 대한 지침이 마련되어야 한다. 특히, 기초선 조사 등으로 수집되는 데이터의 경우, 현재 명확한 지침 없이 개별적으로 이루어지고 있으며 결과만 문서로 저장되고 있다. 조사를 통해 수집될 수 있는 데이터는 사업 분야별 차이가 있을 것이기에 통합적인 지침이 마련되기는 어려울 것이나, 기본적인 틀은 통합적으로 유지하되, 세부 지침은 분야별로 작성되는 것이 합리적이라 할 수 있다. 현재 사업 형성 단계에서 분야별 또는 SDG별 전문가들의 검토가 이루어지고 있는 바, 조사 대상 및 수집 데이터에 대한 분야별 표준을 마련이 가능할 것으로 사료된다. 또한 데이터 입력 시 직접 입력이 아닌 보기를 주고 선택하도록 하는 방식도 표준화에 효과적일 수 있다. 기자재 목록, 주요 지표

등의 경우 표준안을 수립하여 드롭다운으로 선택입력하는 방식도 고려해 볼 수 있다.

그 다음 품질의 경우, 표준 지침에 맞추어 데이터의 입력이 제대로 이루어졌는지 검토하는 단계라고 볼 수 있다. 각 사업 담당 부서에서 입력되고 있는 정형 및 비정형 데이터에 대해 입력하고 1차적인 품질관리를 수행한다면, 데이터 품질 총괄 담당자를 통해 각 목적에 맞게 입력되었는지 검토하고 관리되어야 한다. 현재 KOICA는 OECD DAC 통계 보고를 위해 해당 항목에 대한 값은 각 사업 담당 부서에서 입력하고 데이터혁신팀의 통계담당자들이 각 사업담당 부서에서 입력된 값을 검토하며 관리하고 있다. 즉, 통계 데이터 외에 일반 관리 데이터 및 성과 데이터에 대해서도 유사한 품질관리 역할의 확대가 필요하다. EU 등 타원조기관의 경우, 데이터를 입력한 뒤, 전문인력에 의한 확인절차를 거치도록 하는 등의 데이터 품질관리 절차가 마련되어 있다. 이러한 생성단계에서의 품질관리 뿐 아니라, 데이터 저장관리 시스템을 마련한 후에는 생산 이후 단계에서 데이터 관리 규칙의 변화 등에 따라 데이터가 유효한 값을 계속 유지할 수 있도록 노후화를 방지하는 장치의 마련이 필요하다. 또한 품질관리에도 기준이 요구되는바, 이를 위한 가이드라인 수립도 뒤따라야 할 것이다.

셋째, 데이터의 생애주기 관점에서 데이터의 중요도를 판단하여 이에 맞는 저장 기한 및 방식에 대한 합의가 필요하다. 현재 KOICA의 문서는 그 성격에 따라 보존기간이 정해져 있다. 데이터 또한 이와 유사한 체계 마련이 필요하다. KOICA 데이터의 경우 경영관리, 사업관리 모두에서 관리(management)시스템과 기록관리(archiving)시스템이 분리되어 있지 않아, 정형 및 비정형 데이터 생산 이후, 체계적이고 용이한 활용에 제약이 크다. 데이터가 생산된 이후 체계적으로 마련된 분류 규칙에 따라 저장되어, 이후 내, 외부 사용자가 검색을 통해 손쉽게 이용할 수 있도록 하는 시스템적인 개선이 반드시 필요하다. 특히, 비정형 데이터의 경우, 비공식 및 공식 보고서와 문서들이 정확하게 적기에 손쉽게 검색, 활용되도록 하는 기록관리 시스템을 통한 관리가 절실하다. 나아가 데이터의 중요도를 판단하는 기준을 마련하여 이에 따른 입력, 수정 및 검토 등의 권한 부여가 필요하다. 또한, 일정 기간 혹은 일정 사용 빈도에 기준에 따라 활용 가치가 떨어진 데이터에 대해 보존을 위해 저장소로 이관하거나 폐기함으로써 데이터 활용의 편의성을 확보할 필요가 있다.

이외에도 함께 고려해봐야 할 대상으로는 데이터의 지적재산권이 있다. 현재 사업수행기관과의 계약에는 원자료 및 관련 소유권에 대한 구체적인 명시는 없다. 하지만, 향후 데이터의 체계적인 관리를 위해선 데이터의 소유권에 대해 명확히 할 필요가 있다. 더욱이 공공데이터로 개방 시 데이터의 자유로운 활용을 위해서도, 소유권을 KOICA에 귀속할 필요가 있다. 따라서 기초선,

중간선, 종료선 데이터 수집 및 분석 결과를 제출, 원자료 산출물에 포함하여 제출, 데이터의 소유권에 대해 KOICA 및 사업수행기관이 반드시 인지하고 이러한 내용이 계약서에 명시될 필요가 있다.

더불어, 시스템 개선 측면에서 현재 엑셀 혹은 별도의 문서로 관리되고 있으나 시스템에 수집되지 못하는 데이터를 시스템으로 포함시킬 필요가 있다. 인사이동이 잦은 구조 하에 시스템으로 저장되지 못하는 데이터는 관리 대상임에도 불구하고 결국 언젠가는 누락되거나 삭제될 가능성이 높기 때문이다. 또한 프로젝트의 효율적인 관리를 위해 연동이 필요한 데이터를 [표 24]에 별도로 기술하였다. 일반관리에 있어 사업변경이력, 계약정보, 예산집행현황은 사업 수행기간 동안 유동적으로 변하며, 관리자가 프로젝트의 일정, 예산, 범위를 점검하고 관리하기 위해서 숙지하고 있어야 하는 요소이다. 현재 IPM 상에서 사업변경이력은 별도의 탭에서 확인할 수 있으며, 계약정보, 예산집행 현황은 사업실시-데이터 관리-사업별 데이터 관리를 순차적으로 클릭한 후 사업 검색을 하면 확인할 수 있으나, 여러 번의 단계를 거치는 번거로움이 있다. 성과 데이터의 경우, PDM 및 성과지표는 사업 수행기간 동안 모니터링 및 평가를 위한 데이터 수집, 분석을 위해 연동되어야 하는 데이터이다. IPM 상에도 이를 고려하여 사업 수행 및 평가 단계에서 PDM과 연동하였다. 그러나 실제 PDM 메뉴로 가면 데이터값이 없는 등 활용이 이루어지지 않고 있으며, 집행계획 첨부문서 상에서만 확인 가능하다.

2. 향후 과제

본 보고서에서는 데이터 거버넌스의 정의를 살펴보고, 주요 요소로 데이터의 표준화, 품질, 생애주기에 대해 논의하였다. 데이터는 중요한 자산으로 인식되어야 하며, 데이터 거버넌스는 전사적인 전략으로 전사의 참여와 지원이 필요하다. 앞서 언급한 내용을 정리하면 다음과 같다. 우선 표준화를 위해 표준용어 수립 및 관련 규정 수립이 필요하다. 데이터 품질을 위해 데이터 검증을 위한 부서 신설이 필요하며, 품질기준에 따라 검토가 이루어질 필요가 있다. 데이터의 성격에 따라 보안등급 및 보관 주기 등의 구분이 필요하며, 기존의 기록관리 체계와 연계 및 통합까지 고려되어야 할 것이다. 특히, 개발협력 프로젝트 데이터를 데이터 거버넌스 구축의 시발점으로 한다면, 앞서 살펴본 3가지 요인에서의 관점 외에도 산출물 관리를 위한 분야별 데이터 수집 가이드라인, 데이터 소유권의 명시화 및 시스템 간의 연동 등의 개선이 함께 논의되어야 할 것이다.

데이터의 가치는 누가 어떠한 관점에서 살펴보느냐에 따라 달라진다. 데이터의 중요성에 대해 무수히 강조하지만, 앞서 언급한 바와 같이 데이터의 생성, 저장·관리, 활용 등의 각 단계별 참여자가 다르기에, 모든 참여자가 체감하는 데이터의 중요성은 다를 것이다. 활용자에 있어서는 데이터가 매우 중요하겠으나, 생성자 및 저장·관리자는 상대적으로 그 중요성을 체감하기 어려울 수도 있다. 기관의 입장에서는 데이터는 중요하며, 각각의 참여자의 역할이 모두 필요하다. 따라서 기관 내 데이터 흐름과 활용도에 대해 모두가 인지하고 각자의 역할에 적극적으로 참여할 수 있도록 독려할 필요가 있다.

개발협력 데이터는 특수하며, 국제개발협력 분야의 난제로 SDG17에서 강조하고 있는 통계역량 강화와도 깊이 연관되어 있다. 따라서 KOICA 사업을 통해 얻어지는 데이터를 체계적으로 관리한다는 것은 그만큼 큰 의미가 있다. 프로젝트 데이터를 시작으로 기관에서 생성되는 데이터는 사업 현황을 파악하고 사업 관리부터 영향력 평가, 사업 방향성 수립에 있어 과학적 근거에 기반하는 토대를 마련하고, 기관 운영에 있어서도 적재적소에 필요한 재원을 투입할 수 있는 근거를 제시한다. KOICA의 경우 사업을 통해 얻어지는 데이터는 내부적으로는 사업 방향성 및 전략 수립에 활용이 가능하며, 사업결과에 대한 평가에도 활용 가능할 것이다. 분야별·지역별·유형별 평균 사업기간 및 예산, 성과 및 평가결과 등이 향후 사업 계획에 참고자료로 활용이 될 수 있을 것이다. KOICA의 경쟁 우위 사업을 쉽게 확인할 수 있을 것이며, 이를 근거하여 다양한 혁신 사업도 발굴할 수 있을 것이다. 또한 대외적으로는 개발도상국에서의 사업과 관련한 데이터를 개방함으로써 개발도상국으로의 접근성을 확대하고 국제개발협력 분야의 다양한 정보를 공유할 수 있는 계기를 마련할 수 있다. 또한, KOICA는 공공기관으로 데이터를 외부에 공개할 의무가 주어진 바, 양질의 공공데이터 개방을 통해 데이터경제 활성화에 기여할 수 있을 것이다.

단기간에 모든 데이터를 포함하는 데이터 거버넌스 수립은 있을 수 없다. 체계적으로 준비하였어도 다양한 이유로 실패할 가능성이 높다는 것을 앞서 살펴보았다. KOICA의 핵심 분야인 프로젝트 데이터를 기반으로 데이터 거버넌스 수립을 시작하여, 점진적으로 확대해 나아가야 할 것이다. 데이터의 활용을 위해 거버넌스를 수립하고 표준화된 데이터를 체계적으로 관리하여 일정 수준의 품질을 유지 관리하는 데에는 상당한 시간과 자원, 인력의 투입을 요구한다. 더욱이 개발도상국이라는 특수한 환경 속에 언어 장벽과 시스템의 한계를 극복하고 양질의 데이터를 생성하고 수집하여 저장한다는 것은 일반적으로보다 더 많은 노력이 요구되리라 예상된다. 따라서 단기적으로는 프로젝트 데이터를 중심으로 기관 내 업무 효율성을

높이기 위한 데이터 거버넌스를 확립하고 관련 정책 및 규정, 기준을 마련하여 내재화하여야 한다. 장기적으로는 개발도상국 현장에서의 프로젝트 데이터 수집이 직접 이루어질 수 있도록 디지털 인프라 구축이 필요하며, 프로젝트 데이터에서 시작하여 향후 전사적 확대를 위한 기반 마련 차원에서 데이터의 흐름을 중심으로 충분한 자원과 인력의 투입이 반드시 요구된다. KOICA는 개발협력 사업을 통해 다양한 종류의 데이터에 노출되어 있으며, 이는 경제적인 관점에서 가치 창출 및 활용 가능성이 매우 높은 데이터를 생성 및 수집할 수 있는 기회를 갖고 있는 것이기도 하다. 데이터 거버넌스 수립을 통해 향후 데이터 기반 의사결정 및 데이터 경제 실현에 한걸음 더 다가갈 수 있을 것으로 기대한다.

참고문헌

- 고요한 · 윤유리 · 박종남 · 조예은. 2020. 개발협력 프로젝트 데이터 활용 방안 연구. 성남:한국국제협력단
- 김석수. 2011. EA를 위한 데이터 거버넌스 구축 사례 연구. 정보기술아키텍처연구. 8(3):255-265
- 김유승. 2014. 기록으로의 공공데이터 관리를 위한 제도적 고찰. 한국기록관리학회지. 14(1):53-73
- 김은진 · 김민수 · 김희웅. 2019. 활용성 제고를 위한 공공데이터 표준화 연구. 지식경영연구. 20(4):23-38
- 박수영. 2009. (프로젝트 관리와 평가) 프로젝트 기획, 모니터링 및 평가 방법론. 성남:한국국제협력단
- 부창완. 2020. 데이터로 경영하라-기업 경쟁력을 높이는 데이터 거버넌스의 변화와 추진방안. 삼성SDS 인사이트 리포트 2020.11.20.
https://samsungsds.com/kr/insights/data_governance.html 접속일: 2021.10.10
- 윤선웅. 2021. 차세대 빅데이터 플랫폼 DATA LAKE. 서울:좋은땅
- 이영상 · 박원환 · 신동선, 원유재. 2019. 개인정보의 활용과 보호를 위한 데이터 거버넌스 성숙도 모형과 종합이행절차에 관한 연구. 정보보호학회논문지. 29(5):1117-1132
- 이춘열 · 황철현. 2014. 일반 관리자를 위한 데이터 관리 및 활용 전략. 서울:컴윈미디어
- 장경애 · 김우제. 2016. 데이터 거버넌스 프레임워크의 구성요소 개발 및 상세화. 한국정보기술학회논문지. 14(9):93-107
- 장경애 · 김우제. 2016. Data Governance 구성요소 개발과 중요도 분석. 한국경영과학회지 41(3):45-58
- 조완섭. 2017. 빅데이터 거버넌스와 표준화 동향. OSIA Standards & Technology Review. 30(2): 26-29
- 조진희. 2019. 디지털 전환 시대의 데이터 경제 거버넌스 구축 전략. 한국지역정보화학회 학술발표대회 논문집, 3-17.
- 최영훈 · 김동원 · 김두현 · 김석주 · 김선경 · 명승환 · 서진환 · 송충근 · 윤상오 · 이강춘 · 이태영 · 정익재 · 정진우 · 한세익. 2014. 정보통신기술과 행정. 고양:대영문화사

- 최명석 · 이상환. 2020. 데이터 관리 계획의 국내 현황 및 과제. 한국콘텐츠학회논문지. 20(6):220-229
- 하수옥 · 이강찬 · 인민교 · 이승윤. 2015. 국내외 빅데이터 표준화 현황 및 전망. 전자통신동향분석. 30(2):32-39
- 한국국제협력단. 2021. 내부문서 ‘핵심데이터 선정 현황 및 IPM 개선방향 중간보고’
- 한국국제협력단. 2020. 내부문서 ‘2021년도 신규사업 PCP 개정(안)’
- 한국국제협력단. 2018. 국제개발협력 프로젝트 실행과 관리. 성남:한국국제협력단
- 한국데이터산업진흥원. 2020. 2020 데이터산업 백서. 서울: 한국데이터산업진흥원
- 한국데이터산업진흥원. 2021. 데이터 표준화 개념.
dataonair.or.kr/db-guide/da-guide/?pageid=2&mod=document&keyword=표준화&uid=269, 접속일: 2021.10.10
- 한국정보화진흥원. 2018. 공공데이터 품질관리 매뉴얼 v2.0. 대구:한국정보화진흥원
- 한국데이터베이스진흥센터. 2006. 데이터품질관리 지침. 서울:한국데이터베이스진흥센터
- 한국데이터베이스진흥원. 2009. 데이터 품질관리 시리즈 4 데이터 품질진단 절차 및 기법. 서울: 한국데이터베이스진흥원
- 행정안전부. 2019. 공공데이터 관리지침. 행정안전부 고시 제2019-71호
- 행정안전부. 2020. 공공데이터 공통표준용어. 행정안전부고시 제2020-42호
- Abraham, R., Schneider, J. and Brocke J. 2019. “Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management,” International Journal of Information Management, Vol 49, 424-438
- Baum, W.C. and Tolbert, S.M. 1985. “Investing in Development: Lessons of World Bank Experience.” World Bank
- Cahill, J. 2018. “Transforming Data into Action: The Business Outlook for Data Governance.” The Economist Intelligence Unit
- DAMA International. 2017. DAMA-DMBOK Data Management Knowledge System Guide 2nd Ed. NJ: Technics Publications
- IBM. 2007. The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance. NY:IBM Corporation

- O'Neal, K. 2015. "Top 10 Artifacts Needed for Data Governance. First San Francisco Partners
- Khatri, V. Brown, C.V. 2010. "Designing Data Governance." Communications of the ACM, Vol. 53, no.1, 148–152, January 2010
- Ladley, J. 2012. Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program, Morgan Kaufmann, Print. The Morgan Kaufmann Series on Business Intelligence.
- Laudon K,C. and Laudon J.P. 2018. Management Information Systems: Managing the digital firm, 15th ed, Essex: Pearson Education
- Merkus,J., Helms, R. and Kusters, R. 2019. "Data Governance and Information Governance: Set of Definitions in Relation to Data and Information as Part of DIKW." ICEIS, 143–154
- OECD. 2021. Going digital: The "Norwegian development aid - Statistics and results" portal. Development Co-operation Tips-Tools Insights Practices
- OECD/DAC. 2013. Types of Aid - Frequently Asked Questions. OECD/DAC
- Olavsrud, T. 2021. "Data Governance: A best practices framework for managing data assets." <https://www.cio.com/article/3521011/what-is-data-governance-a-best-practices-framework-for-managing-data-assets.html>, last cited: Oct.10,2021
- Otto,B. 2011. "Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers." Communications of the Association for Information Systems.
- Power, D. 2011. "A 4-D Approach to Data Governances". Information Management, 14(6):28–30
- Project Management Institute. 2021. PMBOK GUIDE 7th edition.
- SAS. 2018. The SAS Data Governance Framework: A Blueprint for Success. https://sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-data-governance-framework-107325.pdf last cited: Oct.10,2021
- Soares, S. 2010. The IBM Data Governance Unified Process: Driving Business Value with IBM Software and Best Practices, Ketchum: MC Press

USAID. 2020. ADS Chapter 579 USAID Development Data. Washington D.C.:US Agency for International Development

Weber, K. Otto, B. and Osterle, H. 2009. "One size does not fit all– A Contingency Approach to Data Governance." ACM Journal of Data and Information Quality. Vol. 1. no.1. June 2009

한국데이터산업진흥원 데이터 온에어 플랫폼 <https://dataonair.or.kr>

Data Governance Institute <https://datagovernance.com>

부록 1. IATI 공개항목 및 DAC 보고항목

〈표 부록-1〉 IATI 공개항목

1	Organization Identifier	기관 식별자	18	Collaboration Type	협력형태
2	Organization Name	기관명	19	Default Flow Type	재원흐름
3	Reporting Organization	보고 기관	20	Default Finance Type	재원종류
4	Total Budget	전체 예산(3년간)	21	Default Aid Type	원조형태
5	Recipient Organization Budget	수원기관 예산(3년간)	22	Default Tied Status	구속성상태
6	Recipient Region Budget	수원지역 예산(3년간)	23	Budget	예산
7	Recipient Country Budget	수원국 예산(3년간)	24	Planned Disbursement	집행 계획
8	Total Expenditure	전체 지출(3년간)	25	Capital Spend	자본 지출
9	Document Link	문서링크	26	Transaction	거래
10	IATI Identifier	IATI 식별자	27	Document Link	문서링크
11	Reporting Organization	보고 기관	28	Related Activity	관련 사업
12	Title	사업명	29	Legacy Data	운영계 데이터
13	Description	사업 설명	30	Conditions	조건
14	Participating Organization	참여 기관	31	Result*	결과
15	Other Identifier	기타 식별자	32	CRS-ADD	CRS-ADD
16	Activity Status	사업 진행 상황	33	FSS	FSS
17	Activity Date	사업 날짜			

〈표 부록-2〉 DAC 보고항목

1	Reporting year	보고 연도	33	Associated financing	혼합신용
2	Commitment Date	약정일자	34	Biodiversity	생물다양성
3	Reporting country	보고 국가	35	Climate change mitigation	기후변화 완화
4	Extending agency	실행기관	36	Climate change adaptation	기후변화 적응
5	CRS Identification no.	CRS ID번호	37	Desertification	사막화 방지
6	Donor Project no.	공여국 프로젝트 번호	38	Currency	통화
7	Nature of submission	제출 성격	39	Commitments	약정액
8	Receipient country	수원국	40	Amounts extended	지출액(집행금액)
9	Channel of delivery_name	전달채널명칭	41	ODA grant equivalent	ODA증여증가액
10	Channel code	채널코드	42	Amounts received (for loans: principal only)	회수액(차관의 경우 원리금만 해당)
11	Bi/multi	양자/다자	43	Amount untied	비구속성 금액
12	Type of flow	자금흐름유형	44	Amount partially untied	부분구속성 금액
13	Type of finance	자금유형	45	Amount tied	구속성 금액
14	Type of aid	원조유형	46	Amount of IRTC	투자 연계 기술 원조
15	Short description/ Project title	사업명	47	If project type, amount of experts-commitments	프로젝트 중 전문가 파견 약정액
16	Sector/Purpose code	분야/목적코드	48	If project type, amount of experts-extended	프로젝트 중 전문가 파견 집행액
17	Geographical target area	사업대상지역 코드	49	Amount of export credit in AF package	혼합신용 패키지 중 수출신용금액
18	Exp. Starting date	사업착수계획년월	50	Leveraging mechanism and role/postion	레버리지 메커니즘
19	Exp. Comp. date	사업완료계획년월	51	Amount mobilised from the private sector	민간동원금액
20	Description	사업상세설명	52	Origin of the funds mobilised	동원한 자금의 출처
21	SDG focus	중점SDG	53	Type of repayment	상환유형

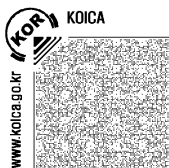
22	Gender equality	양성평등	54	Number of repayments per annum	연 상환 횟수
23	Aid to environment	환경 원조	55	Interest rate	금리
24	PD/GG	참여적개발/굿거버넌스	56	Second interest rate	제2금리
25	Trade developmnet	무역 개발	57	First repayment date	최초 상환일자
26	Disaster Risk Reduction	재난위험 경감	58	Final repayment date	최종 상환일자
27	Inclusion and Empowerment of persons with Disabilities	장애포괄	59	Interest received	수취 이자액
28	Basic Nutrition	기초영양	60	Principal disbursed and still outstanding	지불 원리금 및 잔액
29	RMNCH	모자보건증진	61	Arrears of principal (included in item 46)	연체 원금
30	FTC (Free-stading technical co-operation)	독립적 기술협력	62	Arrears of interest	연체 이자
31	PBA(Sector programme)	PBA 활용여부	63	Future debt service: First year, principal	채무구제서비스: 첫해, 원리금
32	Investment project	투자 프로젝트	64	Future debt service: First year, interest	채무구제서비스: 첫해, 이자

부록 2. 데이터 관리체계 구축방안 연구: KOICA 개발협력 프로젝트 데이터 중심으로 보이스아이코드

표지 제목



목차 1



목차 2



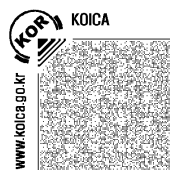
요약 1



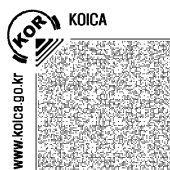
요약 2



요약 3



요약 4



요약 5



요약 6



연구정보 2021-16-194

데이터 관리체계 구축방안 연구:
KOICA 개발협력 프로젝트 데이터 중심으로

발 행 2021년 12월

발 행 인 손혁상

발 행 처 한국국제협력단

편 집 인 ODA연구센터

주 소 경기도 성남시 수정구 대왕판교로 825(우: 13449)

전 화 1588-0434

팩 스 031)740-0260

홈페이지 <http://www.koica.go.kr>

인 쇄 (주)명진씨앤피

I S B N 978-89-6469-591-3 93320

[본 보고서의 저작권은 한국국제협력단에 있으며,
한국국제협력단의 허락 없이 무단 전재와 복제를 금합니다.]

연구자료

연구정보 2021-16-194

978-89-6469-591-3 93320



데이터 관리체계 구축방안 연구 : KOICA 개발협력 프로젝트 데이터 중심으로



KOICA 한국국제협력단

13449 경기도 성남시 수정구 대왕판교로 825
Tel. 1588-0434 Fax. 031-740-0260
<http://www.koica.go.kr>