DCPAPERS

Developing Standardized Metadata Schemas for Multidisciplinary Research Data: A National Platform Approach

Hea Lim Rhee[1,*]

[1]Korea Institute of Science and Technology Information, Daehangro 245, Daejeon, South Korea

Abstract

The Korea Institute of Science and Technology Information (KISTI) is developing standardized metadata schemas for the National Research Data Platform (DataON) to address heterogeneity issues arising from Korea's "Research Data Act," which mandates centralized data deposition. Through systematic analysis and stakeholder engagement, KISTI aims to establish a unified metadata framework that enables cross-disciplinary data discovery and interoperability.

Keywords

metadata, DataON, research data, Korea

1. Background and Motivation

The Korea Institute of Science and Technology Information (KISTI) has been at the forefront of Korea's scientific advancement for over six decades, leading innovation in science and technology information services and supercomputing infrastructure. In the current era of artificial intelligence and high-performance computing convergence, KISTI operates the National Research Data Platform (DataON) to systematically manage and provide long-term access to research data as strategic knowledge assets for national R&D innovation. DataON represents Korea's commitment to transforming research data from isolated institutional resources into a unified national asset that drives innovation-based economic growth and enhances the efficiency and competitiveness of government-funded research activities.

Korea is pioneering the world's first comprehensive "Research Data Act," which fundamentally restructures the national approach to research data management. This landmark legislation shifts responsibility from individual researchers to research institutions, creating a systematic framework for accumulating, sharing, and utilizing research data to generate new scientific, technological, and socio-economic value. The act mandates that all government-funded research institutes deposit their data into DataON, establishing unprecedented requirements for national-scale data sharing and accessibility.

DCMI-2025 International Conference on Dublin Core and Metadata Applications

*Corresponding author.

rhee.healim@kisti.re.kr (H. L. Rhee)

0000-0002-4171-5710 (H. L. Rhee)

https://doi.org/10.23106/dcmi.952587362

2. Problem Statement

The implementation of mandatory data deposition presents significant technical and organizational challenges for DataON. Currently, the platform manages heterogeneous research data from diverse scientific disciplines, including the Korea Institute of Geoscience and Mineral Resources, the Korea Polar Research Institute, the Korea Research Institute of Bioscience and Biotechnology, the Korea Institute of Oriental Medicine, and the Korea Research Institute of Chemical Technology. Each contributing institute operates within distinct disciplinary paradigms, employing specialized data collection methodologies, analytical frameworks, and documentation practices that result in fundamentally different metadata structures and vocabularies.

This metadata heterogeneity creates substantial barriers to effective data discovery, integration, and cross-disciplinary utilization. Users attempting to locate relevant datasets across multiple domains encounter inconsistent descriptive frameworks, incompatible terminology systems, and varying levels of documentation completeness. The lack of standardized metadata schemas undermines the platform's core mission of facilitating knowledge-based innovation and limits the potential for serendipitous discoveries that often emerge from interdisciplinary data exploration.

3. Methodology

KISTI's Data Platform Center has initiated a comprehensive metadata standardization program employing systematic analysis and stakeholder engagement methodologies. The approach encompasses four critical phases: First, conducting exhaustive evaluation of existing metadata schemas currently implemented within DataON, documenting their structural characteristics, semantic coverage, and operational effectiveness. Second, performing detailed comparative analysis of metadata schemas submitted by participating research institutes, identifying common elements, unique requirements, and potential integration pathways. Third, investigating contemporary international metadata standards and emerging best practices in scientific data management, including examination of successful implementations in similar national research infrastructures globally. Fourth, synthesizing findings to develop a unified, extensible metadata schema framework that accommodates disciplinary specificity while ensuring cross-domain interoperability and semantic consistency.

The methodology emphasizes evidence-based decision-making through quantitative analysis of metadata usage patterns, qualitative assessment of user experience requirements, and iterative refinement through stakeholder feedback mechanisms. Particular attention is devoted to maintaining backward compatibility with existing data deposits while establishing forward-looking frameworks that can accommodate emerging research paradigms and technological developments.

4. Expected Outcomes and Impact

The standardized metadata schema framework will fundamentally transform Korea's national research data ecosystem by enabling seamless data discovery, enhanced interoperability, and

https://doi.org/10.23106/dcmi.952587362

systematic reusability across disciplinary boundaries. Implementation across participating institutes will catalyze cross-disciplinary research collaboration, accelerate scientific discovery processes, and maximize return on investment for government-funded research activities. The framework seeks Telecommunications Technology Association (TTA) certification to ensure alignment with national technical standards and facilitate broader adoption across Korea's research infrastructure.

Beyond immediate operational improvements, this standardization initiative will establish Korea as an international leader in national research data management, providing a replicable model for countries developing similar legislative frameworks and technical infrastructures.

5. Significance

This research addresses fundamental challenges in large-scale, multidisciplinary research data management while contributing to the broader scientific understanding of metadata standardization in heterogeneous research environments. The work provides critical insights for policymakers, research administrators, and technical professionals engaged in developing national research data infrastructures, offering practical solutions that balance standardization benefits with disciplinary autonomy requirements. The standardized approach will position Korea's research ecosystem for enhanced productivity and innovation in an increasingly data-driven scientific landscape.

Acknowledgments

This paper was conducted as part of the 2025 Basic Research Project of the Korea Institute of Science and Technology Information (KISTI). (Project Number: (KISTI) K25L3M1C2)

https://doi.org/10.23106/dcmi.952587362