

DATABASE

Open Access



# BioOne: a national-scale platform for integrated discovery and utilization of diverse biological resources in South Korea

Tae-Ho Kang<sup>1†</sup>, KyoungSoo Ha<sup>1†</sup>, WonHyun Kyung<sup>1</sup>, Yong-Jin Jeon<sup>1</sup>, Geun-Hyoung Jo<sup>1</sup>, KangMin Park<sup>1</sup>, KwangHee Lee<sup>1</sup> and Tae-Eun Jin<sup>1\*</sup>

## Abstract

The rapid growth of biological resources and associated research outputs has increased the complexity of resource discovery, access, and reuse across heterogeneous repositories. However, fragmented metadata schemas, limited interoperability, and siloed access mechanisms continue to hinder the integrated exploration of biological resources and their associated knowledge. Herein, we present BioOne (Biological resources One-Stop service platform), a unified informatics framework designed to address the fragmentation of national biological resources by integrating heterogeneous metadata from 14 distinct clusters and establishing a seamless resource-to-knowledge pipeline to enhance the discoverability and practical utilization of biological assets.

BioOne is a national-scale web-based discovery platform that harmonizes biological resource metadata across 14 domain-specific biological resource clusters in Korea and systematically links these resources with external knowledge objects, including research papers, patents, biological datasets, and disease–drug–target information, through a unified discovery interface. To achieve this, we adopted a standard-aligned metadata integration framework, interoperable identifier mapping, and a modular system architecture to support scalable indexing, cross-domain search, and association-based navigation.

By extending conventional catalog-based biological resource databases with an integrated discovery and access layer connecting distributed biorepositories to evidence-oriented knowledge resources, BioOne provides an informatics infrastructure for data-driven discovery, translational research, and coordinated utilization of biological resources at the national scale. The BioOne also offers a transferable implementation model for the large-scale integration of distributed biological resource systems.

**Keywords** BioOne, Biological resources, Metadata integration, Interoperability, Biological database

## 1 Background

Biological resources and their associated experimental and analytical data are foundational assets for modern life science research [1, 2]. Advances in high-throughput technologies have substantially increased the volume and diversity of biological data, intensifying the need for systematic management, integration, and use of biological resources [3–5]. These resources underpin diverse research activities, including functional genomics,

<sup>†</sup>Tae-Ho Kang and KyoungSoo Ha are co-first authors.

\*Correspondence:

Tae-Eun Jin  
tejin@kribb.re.kr

<sup>1</sup> Korea Bioinformation Center (KOBIC), Korea Research Institute of Bioscience & Biotechnology (KRIBB), Daejeon, Republic of Korea



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

disease mechanism studies, drug discovery, and agricultural biotechnology [6–8].

Despite the proliferation of biological resource centers and repositories worldwide, biological resource information remains fragmented across independent platforms with heterogeneous metadata structures and access policies, limiting cross-domain discovery, hampering reproducibility, and increasing the effort required to identify and evaluate suitable resources [9, 10]. Researchers often need to navigate multiple portals to supplement inconsistent metadata and manually integrate contextual evidence from research papers, patents, and datasets [11, 12]. The NCBI (National Center for Biotechnology Information) in the USA [13], ELIXIR (European life-sciences infrastructure for biological information) infrastructure in Europe [14], and NBDC (National Bioscience Database Center) in Japan [15] exemplify global efforts to integrate biological resources and data, all emphasizing standardized metadata and unified access.

To address these challenges, we developed BioOne (Biological resources One-Stop service platform), a national-scale platform that integrates distributed biological resource metadata and associated research outputs into a unified discovery environment [4, 9, 16–19]. By harmonizing heterogeneous metadata and associating biological resources with relevant scientific and technological evidence, BioOne facilitates efficient resource discovery and knowledge-driven utilization across research domains.

## 2 Construction and content

In accordance with the Third National Strategy for the Management and Utilization of Biological Research Resources of Korea, national biological resources are organized into 14 major biological resource clusters to enable systematic management and domain-specific utilization [17, 19]. As described in Supplementary Table S1, these clusters include animal models, cell lines, chemical compounds, fishery resources, human brain resources, human derived resources, livestock, marine resources, microbes, natural products, pathogens, seeds, stem cells, wildlife resources [19]. These 14 clusters represent a strategic reorganization of national biological assets into specialized domains, each governed by unique curation standards and quality controls tailored to specific resource types. Although the pathogen cluster prioritizes bio-safety and epidemiological data, the seed cluster focuses on agricultural traits and genetic diversity. BioOne bridges these distinct domains by harmonizing heterogeneous metadata into a standardized discovery layer. This integration enables cross-disciplinary research, which was previously hindered by siloed management, by systematically linking biological

resources with associated knowledge objects, including research papers, patents, and biological datasets.

BioOne integrates metadata and associated research outputs across all 14 clusters, providing a unified discovery environment while preserving the resource information, curation practices, and distribution workflows unique to each cluster. BioOne is a centralized, web-based platform designed to integrate heterogeneous biological resource information through standardized interfaces in accordance with the Korean Standard for Biological Resource Data Integration, as described in Supplementary data 1. The platform adopts a multimodal data collection strategy to construct a comprehensive database encompassing biological resource metadata, scientific research papers, patents, biological datasets, disease information, and drug data [20–23].

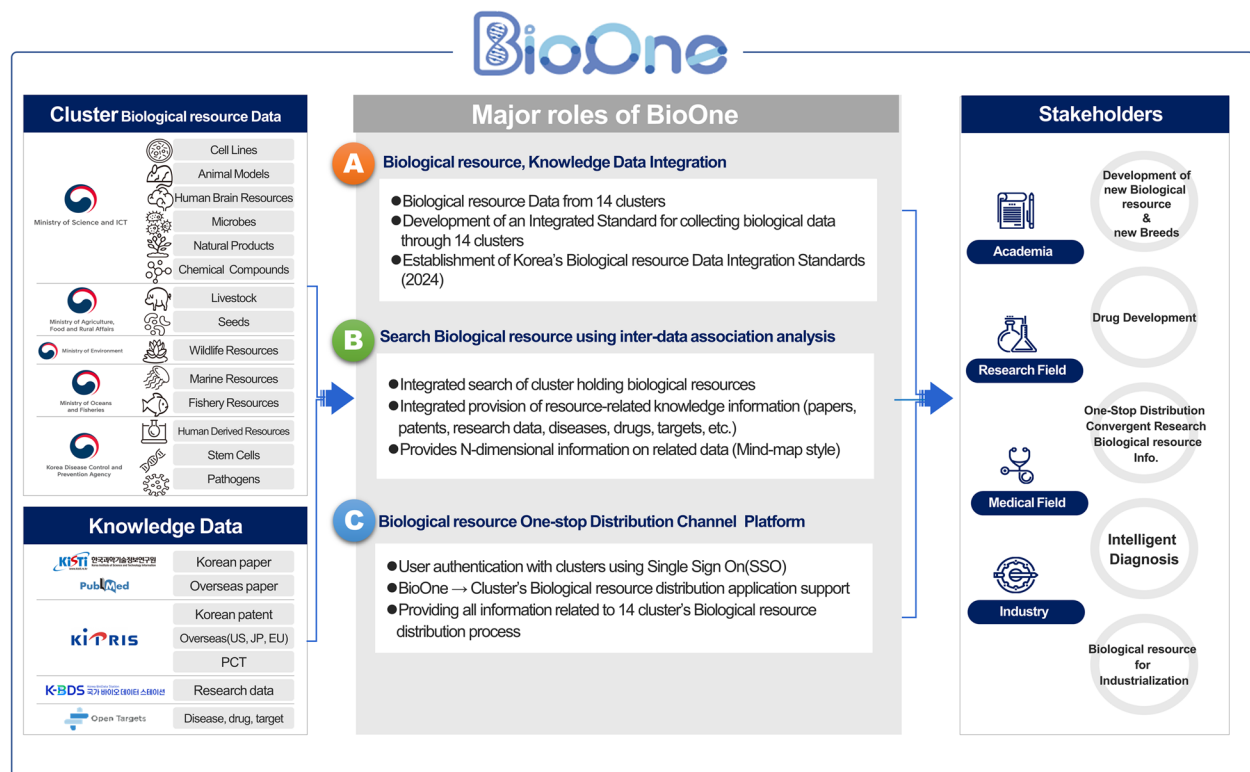
As illustrated in Fig. 1, the conceptual framework and core functionalities of BioOne include a biological resource data integration system, an integrated search environment, semantic enrichment, and an integrated access-and-distribution workflow interface. BioOne provides a robust integrated search engine spanning all indexed data types, including biological resources, research papers, and patents, allowing researchers to explore cross-domain connections using a single query [9, 24].

Simultaneously, the platform supports cluster-specific navigation, enabling experts to perform in-depth searches within a particular biological domain while maintaining the unique metadata context of each cluster [20, 25]. Elasticsearch, which facilitates real-time indexing and complex query processing, powers the search capabilities of the platform [25].

### 2.1 System architecture

BioOne is a multilayered informatics framework designed to bridge the gap between fragmented biological resource repositories. As shown in Fig. 2, the system architecture consists of three primary functional layers. First, the Data Acquisition Layer harvests heterogeneous metadata from 14 national resource clusters via RESTful (Representational State Transfer) application programming interfaces (APIs) [26]. Second, the Integration and Processing Layer transforms raw data into a unified schema, ensuring semantic interoperability. Finally, the Service Layer automatically links integrated resources with external knowledge objects, such as PubMed, GenBank, and patent databases. This systematic workflow allows users to navigate from a biological resource to its functional genomic data and associated research outcomes within a single interface [6, 27, 28].

Additional data sources, including biomedical research papers and disease–target databases, were integrated



**Fig. 1** Conceptual framework and core functionalities of BioOne. BioOne integrates fragmented biological resource metadata from 14 national clusters into a standardized framework. **A** Data integration: heterogeneous metadata were harvested and normalized using automated pipelines and the Korea Biological Resource Data Integration Standard. **B** Knowledge enrichment: resources were systematically linked to external knowledge objects, including publications, patents, and disease–drug–target datasets. **C** Discovery services: the platform provides a unified search interface (Elasticsearch-based), category-specific navigation, and association-based discovery tools. This architecture enables researchers to explore resources within broader functional and scientific contexts through a single entry point

through API-based pipelines and pre-processed using Logstash before indexing in Elasticsearch [26, 27, 29]. Patent data were extracted from XML (eXtensible Markup Language)-based sources and transformed into structured JSON (JavaScript Object Notation) formats using Python-based pipelines, enabling the integration of complex intellectual property metadata.

Direct database-to-database synchronization via validated database views was used to ensure data consistency and prevent duplication [30]. This architecture minimizes manual intervention and enhances the long-term maintainability.

## 2.2 Data and knowledge integration

BioOne functions as a centralized gateway that integrates multiple specialized biological resource portals and external knowledge bases. Core biological resource meta-data—including taxonomy, origin, storage conditions, availability, and application procedures—are harmonized

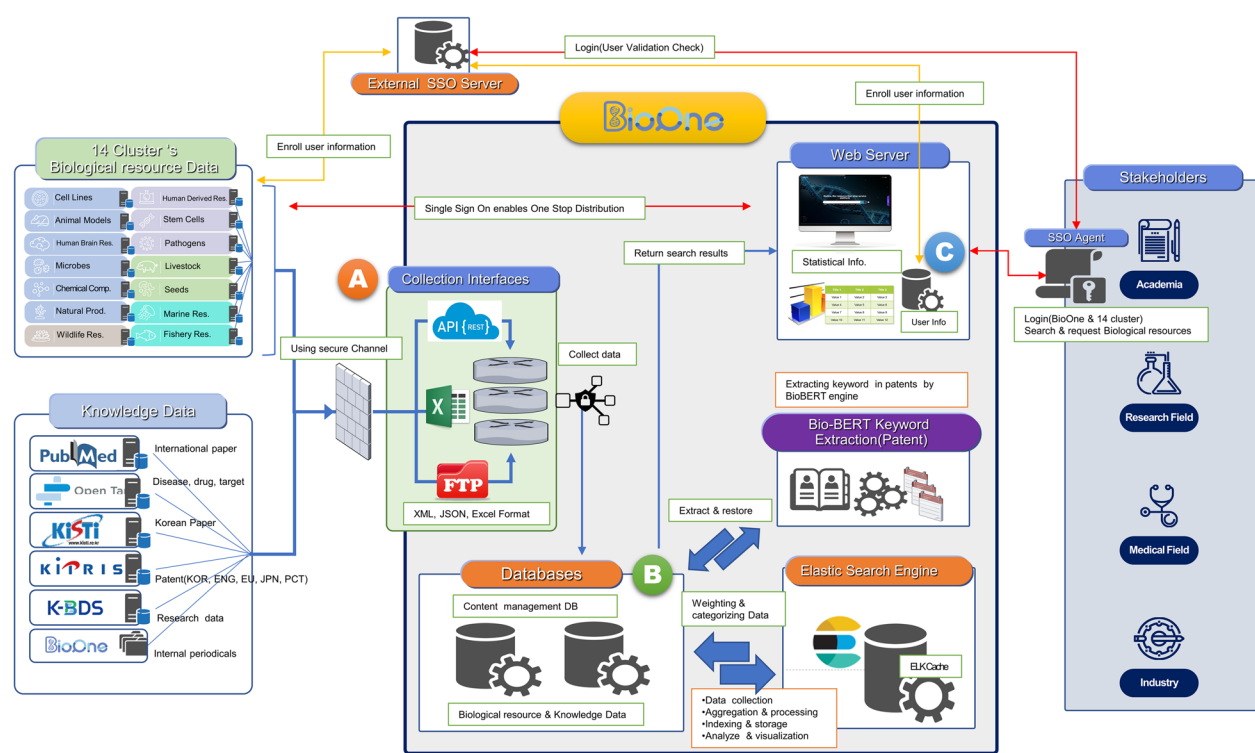
to ensure consistent representation across resource types [5, 9, 18, 31].

Table 1 summarizes the external data sources integrated with BioOne to support an integrated discovery environment. The sources are grouped into four functional domains: biological resource information, literature and patents, drug and disease informatics, and research results.

As described in Table 1, the platform integrates international and Korean research paper databases, patents, and biological data, enabling users to explore biological resources within a broader scientific and technological context. BioOne supports the comprehensive evaluation of resource relevance, novelty, and application potential by linking biological resources to research papers, patents, and research datasets [9, 24].

## 2.3 Search, semantic enrichment, and visualization

BioOne leverages the Elasticsearch–Logstash–Kibana (ELK) stack to support scalable indexing, real-time retrieval, and interactive visualization [16]. Advanced



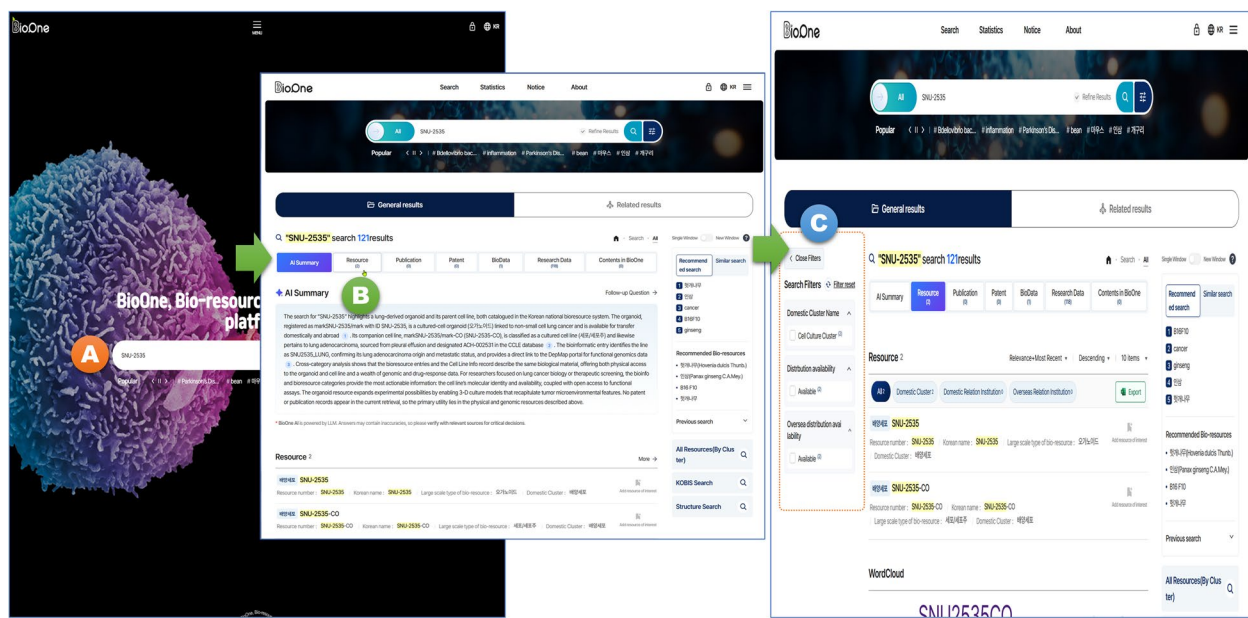
**Fig. 2** Technical system architecture of BioOne. This architecture is designed with a logical flow, moving from raw data acquisition to refined user services. It comprises three functional layers. **A** Data acquisition layer: this layer collects standardized metadata for biological resources provided by 14 national resource clusters. It collates domestic and international knowledge data, including academic papers, patents, research data, drugs, targets, and genetic information. **B** Standardization and integration layer: this layer standardizes heterogeneous data into a unified schema. It performs search indexing using Elasticsearch to improve the search efficiency. **C** Service layer: this layer provides a user interface for searching and establishing knowledge connection. The arrows in the diagram represent the data flow, beginning with raw data collection and leading to an integrated knowledge output

**Table 1** Categorization and functional roles of external data sources integrated into BioOne

| Category                        | Data source                            | URLs                                                                            | Information provided                                                                    | Records                 |
|---------------------------------|----------------------------------------|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|-------------------------|
| Biological resource information | Specialized biological resource portal | Cluster-specific portals for 14 national biological resource domains            | Each cluster provides access to its own biological resources through a dedicated portal | 1,167,776 (2025.12.22)  |
| Literature and patents          | PubMed (International Papers)          | <a href="https://pubmed.ncbi.nlm.nih.gov">https://pubmed.ncbi.nlm.nih.gov</a>   | International biomedical literature access through NCBI APIs                            | 39,827,329 (2025.12.22) |
|                                 | KISTI (Korean Papers)                  | <a href="https://scienceon.kisti.re.kr">https://scienceon.kisti.re.kr</a>       | Korean bio-related scientific publications                                              | 290,053 (2025.12.22)    |
|                                 | KIPRIS                                 | <a href="https://www.kipris.or.kr">https://www.kipris.or.kr</a>                 | Integration with the patent data                                                        | 2,122,338 (2025.12.22)  |
| Drug and disease                | Open Target Platform                   | <a href="https://platform.opentargets.org">https://platform.opentargets.org</a> | Integration with disease, target, drug data                                             | 136,376 (2025.12.22)    |
| Research results                | K-BDS                                  | <a href="https://kbds.re.kr">https://kbds.re.kr</a>                             | All data generated and used through Korean Bio R&D                                      | 156,982 (2025.12.22)    |

text analysis techniques, including language-specific tokenizers and synonym dictionaries, improve the search accuracy across multilingual and heterogeneous datasets [25]. Text-based retrieval is optimized using analyzer-level configurations rather than relevance-score tuning. BioOne applies the Nori tokenizer for Korean-language

processing [26], an N-gram analyzer for partial matching, and a curated synonym dictionary for term normalization across data types. The search results and associated relationship-based visualizations were ranked using the default BM (Best Matching) 25 relevance scoring of Elasticsearch.



**Fig. 3** Stepwise workflow of the BioOne discovery interface. **A** Step 1: the user enters a search query (e.g., “SNU-2535”) within the BioOne search panel to retrieve relevant biological resource records. **B** Step 2: on the search results screen, users can select the resource menu to activate the detailed search panel (left panel), and then use it to limit the search results for biological resources to the cluster or specific knowledge data of their choice. **C** Step 3: BioOne displays a consolidated results page integrating biological resource records with linked knowledge objects, including related publications, patents, and biological datasets. The side menu provides various filtering options to refine the results by resource type and source, enabling users to identify and utilize relevant resources more effectively

BioOne applies BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text Mining), a biomedical domain-specific language model, to extract representative keywords from patent claims and related texts to address the limited keyword metadata typically associated with patent documents [32]. The extracted terms were indexed as supplementary metadata, enabling improved discoverability of patent records and supporting context-aware keyword recommendations for related publications and research datasets. Semantic enrichment is used to improve recommendations and keyword expansion services across research papers and datasets [24, 32, 33].

Search results are presented through tabular views and interactive radial visualization interfaces, enabling the intuitive exploration of relationships among biological resources, research papers, patents, and datasets [25, 34].

Figure 3 summarizes the major roles of BioOne in the national biological resource ecosystem to clarify the scope and positioning of the platform beyond its technical architecture. In addition to aggregating standardized metadata across 14 cluster-specific portals, BioOne links biological resources with evidence-oriented knowledge objects, including research papers, patents, and disease–drug–target information, to support discovery-driven utilization. This conceptual view highlights how

integrated search and association-based exploration in BioOne facilitates stakeholder workflows from resource identification to integrated access and distribution. A comprehensive user manual detailing the search functionalities and association-based navigation is available through the BioOne website to support accessibility and ease of use for diverse researchers [<https://www.bioone.re.kr/help/en/index.html>].

## 2.4 Performance and security

BioOne uses a single sign-on (SSO) authentication framework to provide seamless access across integrated biological resource clusters [35, 36]. Centralized identity management ensures secure and uniform authentication through a trusted identity provider (IdP), simplifying security policy enforcement, user provisioning, and access monitoring [14, 37]. This architecture is intended to improve system security, operational manageability, and user convenience.

The implementation of SSO reduces administrative overhead by eliminating the need for individual authentication systems at each biological resource cluster, thereby enabling centralized and consistent user access management. Consequently, IT support costs and login-related inquiries were minimized [37]. SSO systems rely on standard authentication protocols, such as Security Assertion Markup

Language or OpenID Connect, allowing secure token exchanges between BioOne and the central IdP [36, 37].

BioOne streamlines biological resource procurement through integrated authentication and one-stop access to specialized biological resource distribution portals. This integration reduces administrative overhead and supports a more efficient biological resource acquisition process through centralized identity management. [37, 38].

As described in Supplementary data 2, benchmark testing demonstrated stable performance under concurrent user access within an on-premises, secure research network [30]. The system operation complies with national information security policies, with future plans for secure external deployment to improve global accessibility [17, 38].

### 3 Utility and discussion

BioOne differs from general-purpose information retrieval tools in that it supports validated resource discovery together with access pathways. For example, when a user searches for “SNU-2535,” BioOne can present the resource record together with linked research outputs, the corresponding repository or cluster information, and the access route for distribution through the relevant center. By contrast, general-purpose language models or web search tools may summarize textual information related to “SNU-2535,” but they do not natively provide validated provider information or integrated distribution procedures within the same interface. In this sense, BioOne supports a resource-oriented workflow that extends beyond information retrieval to practical resource identification and acquisition.

BioOne addresses persistent challenges in biological resource discovery through an integrated platform based on metadata harmonization and an interoperable system design. Unlike conventional biological resource portals that operate as isolated, resource-type-specific repositories, BioOne functions as a cross-domain integration layer that connects biological resources with relevant scientific, technological, and biomedical evidence [9, 29]. The platform supports exploratory discovery workflows that extend beyond simple catalog-based access by situating biological resources within a broader data-driven research environment.

A key contribution of BioOne is its standardized metadata integration framework based on the Korean Standard for Biological Resources Data Integration described in Supplementary data 1. The platform reduces structural inconsistencies that impede cross-platform discovery, a challenge common in multi-institution and multidisciplinary data integration settings, by harmonizing heterogeneous metadata across biological resource clusters and aligning them with internationally

recognized principles [5, 24, 28]. BioOne preserves cluster-native identifiers while assigning unified internal identifiers for indexing and cross-linking [5, 9]. This dual-identifier strategy enables seamless integration without disrupting the existing cluster-level distribution and management workflows [9, 28].

From an informatics perspective, BioOne extends beyond conventional catalog-style databases by incorporating semantic enrichment and relationship-driven exploration [24, 32, 33]. The integration of research papers, patents, diseases, and drug information enables users to contextualize biological resources within the existing knowledge landscape [21, 22, 24]. The application of domain-specific language models to under-annotated text sources, particularly patent documents, enhances discoverability by mitigating the limitations of keyword-sparse metadata [32, 33]. Although automated semantic extraction may introduce noise, this approach substantially improves recall in exploratory search scenarios, which are common in early-stage biological resource utilization [24, 32].

Compared with conventional integration approaches, BioOne adopts a standardized metadata harvesting model to support more consistent data integration across diverse biological domains [5]. Unlike link-based systems that primarily provide redirection, mediator/wrapper-based federation models that depend on source-specific query translation, and real-time API-based approaches that may be affected by source-side variability, BioOne relies on pre-integrated and indexed metadata within a unified schema [30]. This design supports more stable query performance, reduces dependency on source-side interface changes, and improves cross-domain consistency in resource discovery.

In relation to platforms such as NCBI (Entrez) and ELIXIR, BioOne has a distinct functional scope in the context of biological resource utilization [14, 20]. Whereas NCBI and ELIXIR primarily support broad life science data access and analysis, BioOne is specifically designed to manage provenance and utilization metadata associated with biological resources. Through a unified discovery interface, BioOne links biological resources with related knowledge objects, including publications, patents, and disease–drug–target information, to support resource-oriented exploration and utilization.

BioOne is complementary to existing Korean biological data infrastructure, particularly the Korea BioData Station (K-BDS). Although the K-BDS focuses on the long-term archiving and sharing of raw biological research data generated through Korean R&D projects, BioOne focuses on the discovery and utilization of biological resources and their associated contextual knowledge [17, 19, 23]. Linking biological resources to datasets archived

in K-BDS allows researchers to trace experimental evidence across biological samples and digital data, clarifying functional separation while enabling synergistic interoperability [5, 33].

Currently, full integration has been achieved for a subset of Korean biological resource clusters, and metadata consistency varies across resource types due to differences in curation practices and update frequencies [9]. Semantic enrichment based on automated language models requires continuous validation and refinement to improve precision [32, 33]. Addressing these challenges requires continuous standardization efforts, the incorporation of ontology-driven annotations, and iterative evaluation using curated reference datasets [5, 24].

BioOne functions as an implementation-level infrastructure that translates cluster-based governance into an operational discovery platform by integrating biological resources across 14 nationally designated clusters defined in the Third National Strategy [17, 19]. This design enables cross-cluster exploration and comparative analysis while maintaining cluster-specific management and distribution workflows, thereby supporting both centralized visibility and decentralized stewardship of the biological resources. In this context, BioOne is well positioned to support the Fourth National Strategy for the Management and Utilization of Biological Research Resources by providing a scalable foundation for expanded cluster integration, ELK-based discovery services, and strengthened linkages between biological resources and data-driven research ecosystems.

## 4 Conclusion

BioOne presents a practical implementation of a national-scale biological resource discovery platform that consolidates distributed biological resource metadata and associated research objects into a unified access framework. The platform improves biological resource discoverability and supports a better understanding of relationships among resources by harmonizing heterogeneous metadata, applying interoperable identifier mapping, and linking biological resources with research papers, patents, and biological datasets. This integrated approach reduces fragmentation in research workflows and supports more efficient identification and utilization of biological resources.

Beyond the initial 14 biological resource clusters, BioOne provides an architectural basis for the future integration of additional resource centers. New sites can be incorporated through schema mapping to the BioOne core metadata model, automated harvesting through standardized interfaces, and identifier normalization for cross-resource linking. This modular design provides

a scalable framework for expanding into additional domains while maintaining existing search functions, metadata consistency, and association-based discovery services.

In addition to its immediate technical contributions, BioOne establishes an infrastructure model for coordinated biological resource utilization at the national level. The modular architecture, standard-aligned metadata framework, and complementary integration with the K-BDS provide a scalable foundation for long-term sustainability and international interoperability. With the accumulation of biological data and resources in terms of scale and complexity, platforms such as BioOne will be essential for data-driven research, fostering interdisciplinary collaboration, and maximizing the scientific value of biological resources.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s44342-026-00070-x>.

Supplementary Material 1: Data 1. Korea Biological Resource Data Integration Standard.

Supplementary Material 2: Data 2. Benchmarking test results of BioOne.

Supplementary Material 3: Table S1. Classification and functional overview of the 14 national biological resource clusters integrated into BioOne.

## Acknowledgements

Not applicable.

## Availability and requirements

[Project name] BioOne.

[Project home page] <https://www.bioone.re.kr/en>.

[User manual] <https://www.bioone.re.kr/help/en/index.html>.

[Platform] web-based.

[Programming languages] Python, Java, JavaScript.

[License] CC BY-NC-SA 4.0.

[Any restriction to use by non-academics] No restrictions.

## Use of large language models

Large language models (LLMs) were used to assist with manuscript editing and formatting. An LLM (ChatGPT, OpenAI) was used to improve the clarity and consistency of the English expressions, refine the section structure, and align the manuscript with the target journal style. All LLM-assisted texts were reviewed and verified by the authors, who took full responsibility for the content, accuracy, and integrity of the manuscript. LLMs were not listed as authors and were not used to generate or interpret scientific results.

## Authors' contributions

All authors conceptualized the study. WK and KH prepared figures. TK and WK developed BioOne; TK, KH, WK, and TJ wrote the original manuscript. All authors contributed to the review and editing of the manuscript. TJ supervised the manuscript preparation. All authors have read and agreed to the published version of the manuscript.

## Funding

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (RS-2021-NR057628).

**Data availability**

BioOne is accessible online, and the list of integrated data sources is provided in Table 1.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors declare no competing interests.

Received: 29 December 2025 Accepted: 24 March 2026

Published online: 08 April 2026

**References**

- Courtot M, Gupta D, Liyanage I, Xu F, Burdett T. BioSamples database: FAIRer samples metadata to accelerate research data management. *Nucleic Acids Res.* 2022;50:D1500–7. <https://doi.org/10.1093/nar/gkab1046>.
- Zhang S, Liu K, Liu Y, Hu X, Gu X. The role and application of bioinformatics techniques and tools in drug discovery. *Front Pharmacol.* 2025;16:1547131. <https://doi.org/10.3389/fphar.2025.1547131>.
- Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov.* 2011;10:188–95. <https://doi.org/10.1038/nrd3368>.
- Marx V. Biology: the big challenges of big data. *Nature.* 2013;498:255–60. <https://doi.org/10.1038/498255a>.
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>.
- Luo Y, Zhao C, Chen F. Multiomics research: principles and challenges in integrated analysis. *BioDesign Res.* 2024;6:0059. <https://doi.org/10.34133/bdr.0059>.
- Ruperao P, Rangan P, Shah T, Thakur V, Kalia S, Mayes S, et al. The progression in developing genomic resources for crop improvement. *Life.* 2023;13:1668. <https://doi.org/10.3390/life13081668>.
- Vitorino R. Transforming clinical research: the power of high-throughput omics integration. *Proteomes.* 2024;12:25. <https://doi.org/10.3390/proteomes12030025>.
- Amorim RC, Castro JA, Rocha da Silva J, Ribeiro C. A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Univ Access Inf Soc.* 2017;16:851–62. <https://doi.org/10.1007/s10209-016-0475-y>.
- Lermen D, Gwinner F, Bartel-Steinbach M, Mueller SC, Habermann JK, Balwir MB, et al. Towards harmonized biobanking for biomonitoring: a comparison of human biomonitoring-related and clinical biorepositories. *Biopreserv Biobank.* 2020;18:122–35. <https://doi.org/10.1089/bio.2019.0092>.
- Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, et al. Big data: the future of biocuration. *Nature.* 2008;455:47–50. <https://doi.org/10.1038/455047a>.
- Gligorijević V, Pržulj N. Methods for biological data integration: perspectives and challenges. *J R Soc Interface.* 2015;12:20150571. <https://doi.org/10.1098/rsif.2015.0571>.
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2022;50:D20–6. <https://doi.org/10.1093/nar/gkab1112>.
- Harrow J, Drysdale R, Smith A, Repo S, Lanfear J, Blomberg N. ELIXIR: providing a sustainable infrastructure for life science data at European scale. *Bioinformatics.* 2021;37:2506–11. <https://doi.org/10.1093/bioinformatics/btab481>.
- Kawashima S, Katayama T, Hatanaka H, Kushida T, Takagi T. NBDC RDF portal: a comprehensive repository for semantic data in life sciences. Database (Oxford). 2018;2018:bay123. <https://doi.org/10.1093/database/bay123>.
- O'Donoghue SI. Grand challenges in bioinformatics data visualization. *Front Bioinform.* 2021;1:669186. <https://doi.org/10.3389/fbinf.2021.669186>.
- Ministry of Science and ICT. Third national strategy for management and utilization of biological research resources. Seoul: Ministry of Science and ICT; 2020.
- National Academies of Sciences, Engineering, and Medicine. Biological collections: ensuring critical research and education for the 21st century. Washington, DC: National Academies Press; 2020.
- Nature. Riding the biological resource boom. Nature Publishing; 2022. <https://www.nature.com/articles/d42473-022-00020-5>. Accessed 29 Dec 2025.
- National Center for Biotechnology Information. PubMed. <https://pubmed.ncbi.nlm.nih.gov/>. Accessed 29 Dec 2025.
- Korean Intellectual Property Office, KIPRIS (Korea Intellectual Property Rights Information Service). <https://www.kipris.or.kr>. Accessed 29 Dec 2025.
- Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malagone C, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* 2023;51:D1353–9. <https://doi.org/10.1093/nar/gkac1046>.
- Lee B, Hwang S, Kim PG, Ko G, Jang K, Kim S, et al. Introduction of the Korea BioData Station (K-BDS) for sharing biological data. *Genomics Inform.* 2023;21:e12. <https://doi.org/10.5808/gi.22073>.
- Sfakianaki P, Koumakis L, Sfakianakis S, Iatraki G, Zacharioudakis G, Graf N, et al. Semantic biomedical resource discovery: a Natural Language Processing framework. *BMC Med Inform Decis Mak.* 2015;15:77. <https://doi.org/10.1186/s12911-015-0200-4>.
- Gormley C, Tong Z. Elasticsearch: the Definitive Guide. Sebastopol, CA: O'Reilly Media; 2015. <https://www.elastic.co/guide/en/elasticsearch/guide/index.html>. Accessed 29 Dec 2025.
- Elastic. Nori tokenizer: Korean language analysis plug in. Elastic documentation; 2025. <https://www.elastic.co/guide/en/elasticsearch/plugins/current/analysis-nori-tokenizer.html>. Accessed 29 Dec 2025.
- Fielding RT, Taylor RN. Principled design of the modern web architecture. *ACM Trans Internet Technol.* 2002;2:115–50. <https://doi.org/10.1145/514183.514185>.
- Child AW, Hinds J, Sheneman L, Buerki S. Centralized project-specific metadata platforms: toolkit provides new perspectives on open data management within multi-institution and multidisciplinary research projects. *BMC Res Notes.* 2022;15:106. <https://doi.org/10.1186/s13104-022-05996-3>.
- Moreddu R. Integrated data-driven biotechnology research environments. Database (Oxford). 2025;2025:baaf064. <https://doi.org/10.1093/database/baaf064>.
- Silberschatz A, Korth HF, Sudarshan S. Database system concepts. 7th ed. New York: McGraw-Hill; 2019.
- Merino-Martinez R, Norlin L, van Enckevort D, Anton G, Schuffenhauer S, Silander K, et al. Toward global biobank integration by implementation of the Minimum Information about Biobank Data Sharing (MIABIS 2.0 Core). *Biopreserv Biobank.* 2016;14:298–306. <https://doi.org/10.1089/bio.2015.0070>.
- Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics.* 2020;36:1234–40. <https://doi.org/10.1093/bioinformatics/btz682>.
- Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans Comput Healthcare.* 2021;3:1–23. <https://doi.org/10.1145/3458754>.

34. Keim DA. Information visualization and visual data mining. *IEEE Trans Vis Comput Graph*. 2002;8:1–8. <https://doi.org/10.1109/2945.981847>.
35. Linden M, Prochazka M, Lappalainen I, Bucik D, Vyskocil P, Kuba M, et al. Common ELIXIR Service for Researcher Authentication and Authorisation. *F1000Res*. 2018;7:ELIXIR-1199. <https://doi.org/10.12688/f1000research.15161.1>.
36. Alaca F, Oorschot PCV. Comparative analysis and framework evaluating web single sign-on systems. *ACM Comput Surv*. 2020;53:1–34. <https://doi.org/10.1145/3409452>.
37. Joshi U, Cha S, Esmaili-Sardari S. Towards adoption of authentication and authorization in identity management and single sign on. *Adv Sci Technol Eng Syst J*. 2018;3:492–500. <https://doi.org/10.25046/aj030556>.
38. Cybersecurity and Infrastructure Security Agency. Barriers to Single Sign-On Adoption for small and medium-sized businesses: Identifying challenges and opportunities. Washington, DC: CISA; 2024. <https://www.cisa.gov>. Accessed 29 Dec 2025.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.