

<https://doi.org/10.62405/osiconf.2026.54>

Tsybenko Iryna,

State Scientific and Technical Library of Ukraine, Kyiv, Ukraine,
ORCID: 0000-0002-6408-9011, e-mail: tsyb2012@ukr.net

Zherebchuk Sofia,

State Scientific and Technical Library of Ukraine, Kyiv, Ukraine,
ORCID: 0000-0002-0212-7107, e-mail: zerebcuksofia@gmail.com

DEVELOPMENT OF A UKRAINIAN E-INFRASTRUCTURE CLASSIFIER AS A TOOL FOR INTEGRATION INTO THE EUROPEAN RESEARCH AREA

ABSTRACT. *The study aims to develop a methodological framework for the classification of Ukraine's research e-infrastructure as a prerequisite for its integration into the European Research Area. The research is based on the analysis of European approaches to the organization and cataloguing of digital research infrastructures, including practices used within European research and open science initiatives. The methodology combines comparative analysis, systematization, and classification methods to identify key categories of e-infrastructure resources and define the structure of a national database. As a result, a conceptual model of the classifier was developed, covering digital research services, repositories, data infrastructures, computing resources, communication networks, and open science platforms. The proposed approach enables the inventory, monitoring, and coordinated development of national e-infrastructure resources. The findings demonstrate that a unified classification framework can improve interoperability, support strategic decision-making, and facilitate the integration of Ukrainian research infrastructures into the European research and open science ecosystem.*

KEYWORDS: *e-infrastructure; research infrastructure; digital research ecosystem; classification system; research data management; European Research Area.*

INTRODUCTION. The rapid digitalization of science has fundamentally changed the way research is conducted, shared, and managed. Today, scientific progress increasingly depends not only on researchers and institutions but also on the digital infrastructures that support data-intensive and collaborative research. E-infrastructure provides access to computing resources, research data repositories, communication networks, cloud services, and digital platforms that enable researchers to work across disciplines, institutions, and national borders. Only over the last decade, the European research landscape has moved toward a more integrated model in which digital resources are viewed as a common foundation for scientific collaboration. Initiatives such as the European Open Science Cloud (EOSC) and the European Strategy Forum on Research Infrastructures (ESFRI) have promoted the development of interoperable research environments that facilitate data sharing, resource discovery, and cross-border cooperation. In this context, the ability to identify, describe, and systematically organize research e-infrastructure has become an important prerequisite for their effective management and integration.

For Ukraine, the development of a coherent e-infrastructure ecosystem is closely linked to the country's broader efforts to integrate into the European Research Area.

The need for a systematic approach has become even more evident in recent years. Beyond supporting integration into European research frameworks, Ukraine faces the task of assessing, preserving, and modernizing its research infrastructure under exceptionally difficult conditions. Reliable information about existing digital resources is therefore not only a matter of coordination but also an important element of long-term recovery and development strategies. The creation of a national e-infrastructure classifier can be viewed as an essential step toward building a more transparent, interoperable, and sustainable research ecosystem.

The aim of this study is to develop a methodological framework for the creation of a classifier of Ukrainian research e-infrastructures that can support the systematic organization of digital research resources and facilitate their integration into the European Research Area. The proposed classifier is intended for implementation within the National Electronic Scientific Information System (URIS) and will serve as a tool for collecting, organizing, and maintaining information about national e-infrastructure assets.

To achieve this aim, the study addresses several objectives:

- examines existing European and national approaches to the classification and description of research e-infrastructures;
- identifies the principal categories and characteristics that can be used to describe different types of digital research resources;
- proposes a conceptual classification model and a set of criteria for assigning infrastructure objects to specific categories;
- outlines the structure of a database designed to support the long-term maintenance, analysis, and practical use of the classifier.

A national classifier can address these challenges by creating a structured inventory of existing resources and services. Such an inventory would support more informed planning and investment decisions, help avoid duplication of functionality, and provide a clearer understanding of infrastructure capacities across different research fields. Equally important, it would establish a foundation for monitoring infrastructure development and evaluating its contribution to scientific activity.

From an international perspective, a standardized classification framework would improve the compatibility of Ukrainian systems with European research infrastructures and digital platforms. This is particularly relevant in the context of Open Science and the FAIR principles, where interoperability and standardized metadata play a central role. In practice, integration into European initiatives requires not only technical connectivity but also a shared understanding of how infrastructures are described, categorized, and represented within larger information ecosystems.

Also, the development of a national e-infrastructure classifier should build on existing European approaches to the organization and description of research infrastructures. In this regard, the experience of ESFRI and EOSC is particularly relevant, as both initiatives play an important role in shaping the European research ecosystem. ESFRI provides a framework for identifying and coordinating research infrastructures across different scientific domains, including life sciences, environmental sciences, physical sciences, social sciences, and digital infrastructures. Its approach demonstrates how a structured classification system can improve the visibility of research resources and support strategic planning at both national and international levels.

EOSC focuses on the digital dimension of research by creating an environment where researchers can access data, services, and computing resources across institutional and national boundaries. A key aspect of this initiative is the promotion of interoperability and the implementation of FAIR principles, which enable data and services to be more easily discovered, accessed, and reused.

METHODOLOGY OF RESEARCH. The classification of e-infrastructures is based on a combination of approaches that ensure a clear and structured system.

A taxonomic approach is used to create a hierarchical structure. It allows grouping resources from broad categories, such as computing resources, data storage systems, and collaboration platforms, down to specific services and their characteristics.

A semantic data model is also important. It helps not only to group similar objects but also to define relationships between them. This is especially useful for interdisciplinary research and complex data use across different scientific domains.

The inclusion of e-infrastructures is based on several key criteria.

First is functional purpose, which defines the role of an infrastructure in research. This includes computing systems (e.g., HPC clusters and cloud platforms), data repositories, analytics tools (including AI and big data systems), and research data management platforms.

Second is disciplinary coverage, which identifies the scientific fields supported by the infrastructure, such as natural sciences, engineering, medicine, social sciences, and digital humanities.

Another criterion is the level of integration into the research ecosystem. Priority is given to infrastructures connected to national or international networks such as EOSC, GEANT, or OpenAIRE.

Compliance with international standards is also essential. This includes adherence to FAIR principles, use of standard metadata formats, and compatibility with open science frameworks. Finally, technological architecture is considered, including whether the resource is local or distributed, its access model (open, restricted, or closed), and the technologies used for its operation.

The classification system is designed to align with European standards, especially the European Open Science Cloud (EOSC) and FAIR principles.

EOSC promotes open, connected research infrastructures. Therefore, the classification supports open access to research outputs, integration with European and national systems, and the use of common metadata standards such as CERIF, DCAT, and Schema.org.

FAIR principles guide the management of research data:

- Findable: each resource should have a unique identifier and structured metadata for easy discovery.
- Accessible: access conditions must be clearly defined, including open or restricted availability and authentication rules.
- Interoperable: systems should use common formats and protocols (e.g., RDF, JSON-LD, REST APIs) to ensure data exchange.
- Reusable: resources must include clear usage licences and documentation to support reuse.

LITERATURE REVIEW. The development of classification systems for research e-infrastructures is strongly influenced by current international approaches to open science, research data management, and interoperability of digital research environments. The literature in this field highlights the need for integrated, standardised, and transparent systems that support both national research ecosystems and their integration into global infrastructures.

A central conceptual foundation for this study is the FAIR data principles, which define how research data should be managed in modern scientific environments. According to (Wilkinson et al., 2016), data and digital resources should be findable, accessible, interoperable, and reusable. These principles are widely recognised as a core framework for designing research infrastructures and metadata systems that support long-term usability and cross-platform integration. In the context of e-infrastructure classification, FAIR principles provide the basis for defining metadata structure, access rules, and interoperability requirements.

The importance of interoperability and integrated research environments is also strongly emphasised within the European policy framework for open science. The European Commission (European Commission, 2018) highlights the development of the European Open Science Cloud (EOSC Association, 2017) as a key initiative aimed at creating a trusted and federated ecosystem for research data and services across Europe. EOSC is designed to support open access, data sharing, and reuse of research outputs, which directly influences how national e-infrastructures should be structured and classified.

The strategic vision of EOSC is further developed in the EOSC Declaration, which defines the need for a federated infrastructure that connects existing national and thematic research systems into a unified European

environment (EOSC Association, 2017). This approach underlines the importance of compatibility between national classification systems and broader European frameworks, ensuring that research infrastructures can be discovered and used across institutional and national boundaries. The EOSC platform itself continues to serve as a central reference point for the implementation of these principles in practice (EOSC Association, n.d.).

In addition to European policy initiatives, international organisations provide technical and methodological standards that support data management and security. ISO/IEC 27001:2013 establishes requirements for information security management systems, which are relevant for ensuring the protection and controlled access to research infrastructures (International Organization for Standardization, 2013). Similarly, ISO 9001:2015 defines quality management principles that are applicable to the organisation and governance of research data systems, ensuring consistency, reliability, and continuous improvement (International Organization for Standardization, 2015).

Another important aspect of the literature is the development of open research communication infrastructures. OpenAIRE plays a significant role in supporting open access and research data sharing in Europe. Its guidelines for data management in research projects provide practical recommendations for structuring and publishing research outputs in a way that ensures visibility and reuse (OpenAIRE, 2021). The OpenAIRE platform further contributes to the aggregation and dissemination of research information across Europe (OpenAIRE, n.d.).

Persistent identifiers are also widely recognised as a key component of modern research infrastructures. ORCID provides a global system for identifying researchers, enabling clear attribution and improved integration of research outputs across systems (ORCID, n.d.). In parallel, the Research Data Alliance (RDA) develops recommendations and best practices for data sharing and open science, supporting global coordination in research data management (Research Data Alliance, n.d.).

Finally, national legislation provides the legal framework for access to research and public information. The Law of Ukraine on Access to Public Information (Law of Ukraine № 2939-VI, 2011) establishes principles of openness and transparency, which are essential for the development of national research data systems. The Law of Ukraine on Scientific and Scientific-Technical Activity (Law of Ukraine № 848-VIII, 2015) defines the organisational and legal basis for scientific research, including the use and management of research infrastructures. These legal foundations ensure that the development of classification systems aligns not only with international standards but also with national regulatory requirements.

Stages of Developing the Classification System

The development of a classification system for Ukrainian e-infrastructures is a multi-stage process. It requires analysis of existing resources, design of a structured classification model, alignment with international standards, and implementation of unique identification mechanisms. The final system is intended to support a unified national register of e-infrastructures and ensure compatibility with European approaches to research data management.

The first stage involves a detailed review of current Ukrainian information systems and databases that contain data on digital research resources.

Key sources include:

- Ukrainian Research Information System (URIS), which collects metadata on research activities and can help identify infrastructures used in scientific projects.
- National geospatial data infrastructure, which provides spatial data relevant for environmental, urban, and land-use research.
- Institutional repositories of universities and research organisations, which store open-access publications and datasets.

- Registers of objects of national heritage, which include important research infrastructures requiring special classification.
- Open data and research communication systems, including international services such as OpenAIRE and DataCite.

This analysis helps identify existing infrastructures, their status, and their level of integration into the research ecosystem. It also reveals areas where information is incomplete or not yet structured.

Based on the collected data, the classification system is organised as a hierarchical structure with several levels.

Main classification dimensions include:

- Type of e-infrastructure, such as computing resources, data storage systems, analytics platforms, and research repositories.
- Scientific domain, aligned with national and international discipline classifications.
- Level of integration, ranging from local to international infrastructures.
- Access model, including open, restricted, or hybrid access.

A functional grouping is also applied:

- Computing resources: HPC systems, clusters, and cloud platforms.
- Data storage and management: repositories, data centres, and digital archives.
- Network infrastructures: research and education networks such as URAN and secure data exchange systems.
- Open science platforms: open repositories, journals, and metadata aggregation services.

This structure ensures clear organisation and efficient search of resources within the system.

To ensure compatibility with global research systems, the classification is aligned with key international frameworks:

- EOSC (European Open Science Cloud), which supports open and interconnected research data environments.
- CERIF (Common European Research Information Format), a standard for structured research information management.
- FAIR principles.

Applying these standards improves integration of Ukrainian infrastructures into the European and global research ecosystem and increases their visibility and usability.

A key element of the classification system is the use of Persistent Identifiers (PID), which ensure unique and stable identification of all objects.

The main types of identifiers include:

- DOI (Digital Object Identifier) for publications and datasets
- ROR (Research Organization Registry) for research institutions
- Handle System for digital resources
- ORCID for researcher identification, linked to infrastructures and outputs

The use of PIDs improves traceability, reduces duplication, and strengthens interoperability between Ukrainian and international research systems. It also supports long-term management and integration of digital research resources.

Building the Classification Database

The development of a database for the classification of Ukrainian e-infrastructures is a key stage that determines the overall efficiency, scalability, and ability to integrate with other research information systems. The database should provide structured storage of information about national e-infrastructures, support automated data updates, and comply with modern international standards.

The choice of database architecture is essential for performance and long-term sustainability. Two main approaches are considered: relational and non-relational databases.

Relational systems such as PostgreSQL or MySQL are suitable for structured data, including identifiers, infrastructure names, functional types, and relationships between entities. Non-relational systems such as MongoDB or Elasticsearch are more effective for storing large volumes of metadata, especially semi-structured or dynamic information.

For optimal performance, a hybrid architecture combining both models is recommended. This approach allows flexible management of structured metadata while also supporting fast processing of unstructured or evolving data. It also ensures scalability as the volume of information grows. Reliability should be supported through clustering, backup mechanisms, and distributed processing.

The database should be populated using verified national sources covering the full range of Ukrainian research infrastructures.

These include national registers maintained by the Ministry of Education and Science of Ukraine, particularly those listing objects of national heritage and other officially recognised research infrastructures. Important input also comes from institutional systems such as the URAN network, which connects scientific and educational organisations.

Additional sources include government and institutional research reports describing funded scientific projects, as well as university repositories containing information on digital infrastructures used by researchers.

Combining these sources ensures a complete and up-to-date overview of existing e-infrastructures, their characteristics, and access conditions.

To maintain data accuracy, the system should include automated mechanisms for data collection and updating. This can be achieved through the use of web scraping and API-based data retrieval from open repositories and institutional websites.

An administrative interface should allow organisations to update their own infrastructure records directly. In addition, artificial intelligence and natural language processing techniques can be used to analyse scientific publications and identify references to specific digital research resources.

Such automation reduces administrative workload and helps prevent outdated or inconsistent information.

Interoperability is a core principle of the database design. Integration with national systems such as URIS ensures access to up-to-date information on research infrastructures used in Ukraine.

At the international level, connection with EOSC supports inclusion in the European Open Science Cloud, increasing visibility and reuse of Ukrainian research resources. Integration with platforms such as OpenAIRE and DataCite further strengthens data discoverability and citation. Alignment with international metadata standards, particularly CERIF, ensures compatibility with global research information systems.

To support openness and reuse, the database should provide API-based access. This enables automated data exchange with other systems, creation of analytical dashboards, and integration into external research platforms.

The API should support standard protocols such as REST and GraphQL, ensuring flexibility for different

use cases. At the same time, authentication and access control mechanisms are required to manage different levels of data sensitivity.

CONCLUSIONS. At present, a unified classification system for research e-infrastructures in Ukraine has not yet been established. However, its development is an important step toward creating a single integrated database that can systematically describe the main components of the national research infrastructure. Including elements such as scientific platforms, data repositories, computing resources, research networks, and other digital tools will enable the formation of a transparent and well-organised system for managing research resources.

Without a unified classification framework, it is difficult to ensure effective integration into international research initiatives such as the EOSC. This limits Ukraine's level of participation in global scientific collaboration. The absence of structured classification also leads to fragmentation of information and makes it harder to identify and reuse available resources. In contrast, a structured system would improve resource management, reduce duplication, and simplify access to relevant infrastructures for researchers. It would also increase the visibility of Ukrainian e-infrastructures for international users and strengthen the country's position in the global research environment.

The proposed methodology for building the database of research e-infrastructures shows strong potential, but several areas still require further improvement. First, the classification system should be expanded with additional categories that reflect emerging types of infrastructure, including cloud-based data processing platforms, large-scale data storage systems, and modern open science tools. Second, more advanced automated solutions should be introduced for data collection and updating in order to improve the accuracy and timeliness of information. This would help reduce dependence on manual updates and ensure that the database remains up to date. Third, stronger integration with both national and international systems such as EOSC and OpenAIRE is necessary to support efficient data exchange and improve collaboration between research institutions. Finally, the system should include a more user-friendly interface that allows researchers to easily search, filter, and analyse information according to different criteria.

Integration with international initiatives is a key factor for Ukraine's participation in the global research space. EOSC provides a shared environment for research data and services, and connecting the Ukrainian classification system to this framework would allow national infrastructures to be used within a broader European context while also making them accessible to international partners. OpenAIRE plays an important role in the dissemination of open research outputs in Europe, and its integration would improve access to funded research projects and support better data sharing between countries. Participation in broader initiatives such as GESIS and other global open science networks would further strengthen international cooperation and increase the visibility of Ukrainian research infrastructures. In this context, compliance with international standards for open data, particularly those defined within EOSC and related frameworks, is essential for successful integration.

The methodology has several strengths. It is flexible and can be adapted to different types of e-infrastructures, which makes it suitable for long-term development and scaling. It also improves transparency by making information about research infrastructures more accessible to a wide range of users, including researchers, government institutions, and international partners. In addition, it provides a solid basis for interoperability with global systems, supporting the development of open science and international collaboration.

At the same time, the methodology has certain limitations. The current process of data collection and updating is still partly manual, which increases the risk of outdated information and requires additional administrative effort. Another limitation is the lack of a centralised system for monitoring how e-infrastructures are used, which reduces the ability to assess their efficiency and impact.

Declaration on the Use of AI Tools in Text Preparation. During the preparation of this manuscript, artificial intelligence-based tools were used to support the editing, language refinement, and further methodological development of the content. Specifically, the Claude AI system was applied as a language editing assistant in order to improve the clarity, coherence, and

grammatical accuracy of the text. In this process, prompts were used to request the revision of sentence structures, correction of linguistic and stylistic errors, and improvement of overall readability while preserving the original scientific meaning and authorial intent.

In addition, the Gemini AI system was used as a supplementary analytical tool during the research and conceptual development of the classification framework for e-infrastructures. Its role was related to supporting the structuring of ideas, refinement of methodological approaches, and assistance in the conceptual design of the classification model. The use of this tool contributed to the systematic organisation of ideas and helped ensure consistency with current approaches in research data management and digital infrastructure classification.

It is important to note that all outputs generated with the support of AI tools were critically reviewed, revised, and validated by the author. The final responsibility for the content, accuracy, and scientific validity of the manuscript remains entirely with the author.

REFERENCES

- EOSC Association. (n.d.). European Open Science Cloud (EOSC). <https://www.eosc.eu>
- European Commission. (2018). European Open Science Cloud: Building a trusted, open system for science in Europe. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en
- European Open Science Cloud (EOSC) Association. (2017). EOSC declaration: Towards a European Open Science Cloud. <https://www.eosc.eu>
- International Organization for Standardization. (2013). ISO/IEC 27001:2013 information security management systems — Requirements. <https://www.iso.org/standard/54534.html>
- International Organization for Standardization. (2015). ISO 9001:2015 quality management systems — Requirements. <https://www.iso.org/standard/62085.html>
- Law of Ukraine «On access to public information». № 2939-VI (2011). [in Ukrainian]. <https://zakon.rada.gov.ua/laws/show/en/2939-17>
- Law of Ukraine «On Scientific and Scientific-Technical Activity» № 848-VIII (2015). [in Ukrainian]. <https://zakon.rada.gov.ua/laws/show/848-19>
- OpenAIRE. (2021). OpenAIRE guidelines for data management in Horizon 2020 projects. <https://www.openaire.eu>
- OpenAIRE. (n.d.). OpenAIRE platform. <https://www.openaire.eu>
- ORCID. (n.d.). ORCID: Connecting research and researchers. <https://orcid.org>
- Research Data Alliance. (n.d.). RDA recommendations on data sharing and open science. <https://www.rd-alliance.org>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Цибенко Ірина,

Державна науково-технічна бібліотека України, Київ, Україна,
ORCID: 0000-0002-6408-9011, e-mail: tsyb2012@ukr.net

Жеребчук Софія,

Державна науково-технічна бібліотека України, Київ, Україна,
ORCID: 0000-0002-0212-7107, e-mail: zerebcuksofia@gmail.com

РОЗРОБЛЕННЯ КЛАСИФІКАТОРА Е-ІНФРАСТРУКТУРИ УКРАЇНИ ЯК ІНСТРУМЕНТУ ІНТЕГРАЦІЇ ДО ЄВРОПЕЙСЬКОГО ДОСЛІДНИЦЬКОГО ПРОСТОРУ

АНОТАЦІЯ. Метою дослідження є розроблення методологічних засад класифікації е-інфраструктури України як передумови її інтеграції до Європейського дослідницького простору. У роботі проаналізовано європейські підходи до організації та каталогізації цифрових дослідницьких інфраструктур, а також практики, що використовуються в межах європейських ініціатив у сфері досліджень та відкритої науки. Методологія дослідження ґрунтується на застосуванні методів порівняльного аналізу, систематизації та класифікації для визначення основних категорій е-інфраструктурних ресурсів і формування структури національної бази даних. У результаті запропоновано концептуальну модель класифікатора, що охоплює цифрові дослідницькі сервіси, репозитарії, інфраструктури даних, обчислювальні ресурси, комунікаційні мережі та платформи відкритої науки. Розроблений підхід створює основу для інвентаризації, моніторингу та узгодженого розвитку національної е-інфраструктури. Встановлено, що впровадження єдиної системи класифікації сприятиме підвищенню інтероперабельності, підтримці управлінських рішень та інтеграції української дослідницької інфраструктури до європейського наукового простору.

КЛЮЧОВІ СЛОВА: е-інфраструктура; дослідницька інфраструктура; цифрова дослідницька екосистема; класифікатор; управління дослідницькими даними; Європейський дослідницький простір.