Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Global and National Dynamics of Scopus-Indexed Data Paper Publishing: A Bibliometric Analysis With Evidence From Ukraine

Pavlo Kuznietsov

p.m.kuznietsov@nuwm.edu.ua

National University of Water and Environmental Engineering

Olha Biedunkova

National University of Water and Environmental Engineering

Oleg Pinchuk

National University of Water and Environmental Engineering

Research Article

Keywords: Open Science, FAIR principles, research data management, data article, scientific publishing, data publishing

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11256137/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This article examines the global evolution and national differentiation of Data Papers in the context of the development of open science, with a particular focus on Ukraine. The aim of the study was to determine the temporal dynamics, geographic and disciplinary distribution, major publishing platforms, and the characteristics of Ukraine's positioning within the global research data publication landscape. The analysis was conducted based on Scopus data using the distinct "Data Paper" document type. Publications from 2014 to 2026 were analyzed on a global scale, and those from 2016 to 2026 were analyzed for Ukraine; subject areas, journals, open access models, and mechanisms for waiving publication fees were examined separately. An overall upward global trend in Data Papers was observed, with a peak annual figure of approximately 2,300 documents in 2024. At the same time, Ukrainian activity remains low in volume and volatile: only 57 documents were identified during the study period, with a maximum of 10 publications in 2022. Global publication activity is concentrated primarily in the United States, China, and leading European research systems. The Ukrainian landscape is characterized by a unique combination of computer science, mathematics, social sciences, management and analytics, and natural sciences. The results indicate that Ukraine is already integrated into the Data Paper segment; however, its trajectory differs significantly from the global one in terms of scale and stability.

MSC: 62R07 - Statistical aspects of big data and data science; 68P20 - Information storage and retrieval of data

JEL: O30 - Innovation, Research and Development, Technological Change, Intellectual Property Rights: General; O32 - Management of Technological Innovation and Research and Development.

Introduction

Open Science and the Transformation of Research Publishing

The transformation of scientific communication is changing not only the ways in which research findings are disseminated, but also the very understanding of what constitutes a scientific publication. In the traditional publication model, a scientific finding is primarily represented by a research article, which brings together the problem statement, methods, results, and their interpretation. At the same time, the development of the Open Science paradigm is gradually shifting the focus from a publication as a static text document to a collection of interconnected research objects, including the article, raw and processed data, program code, metadata, methods, protocols, the computational environment, and other digital components necessary for verifying, reproducing, and reusing the results (Hitzler et al., 2021; Ioannidis, 2025). The research article (Table 1) illustrates a fundamental difference between traditional and Open Science models: the article retains its central role in scientific communication but ceases to be the sole carrier of scientific results. Research data acquire the status of an independent, citable resource; code becomes a separate tool for analysis and reproduction; metadata provide a structured description of digital objects; methods and protocols form part of the evidence base; the computational environment ensures the reproducibility of software-based analysis, and semantic links combine

individual resources into a unified digital structure (Hitzler et al., 2021; Ramachandran et al., 2021; Zakharova, 2023; Umbach, 2024). While in the traditional system data were primarily viewed as an internal foundation for obtaining results, in the modern Open Science environment they are increasingly acquiring the status of an independent scientific product. This entails their identification, documentation, deposit, citation, long-term preservation, and ensuring their reusability. Accordingly, research data is no longer merely a supporting element of a journal article but has become a distinct component of scientific communication (Thibault et al., 2023; Ramachandran et al., 2021; Umbach, 2024; Nygren, 2024).

Table 1. Main components of a decomposed scientific publication in the Open Science paradigm

Component	Function in traditional publishing	Role in the Open Science paradigm	Key requirement
Research article (Article)	Presentation of methods, results, and their interpretation	Central textual component linked to other digital objects	Links to data, code, methods, and supplementary materials
Research data	Primarily the basis for obtaining research results	Independent research object suitable for citation and reuse	Repository, DOI/identifier, metadata
Code / software	Often not published or described only in textual form	Tool for analysis, modelling, and reproduction of results	Openness, versioning, documentation, persistent link
Metadata	Limited description within the article	Structured description of a digital object, its provenance, and relationships	Completeness, standardization, interoperability, machine processing
Methods / protocols	Described in the Methods section	Independent component of the evidentiary basis required for reproducibility	Detailed and transparent description of procedures
Computational environment	Usually not part of the publication	Enables reproduction of computational analyses and models	Description of dependencies, software versions, and the execution environment
Semantic links	Largely absent or represented through bibliographic references	Connect the article, data, code, and other resources into an integrated digital structure	Linked Data, graphs, machine-readable relationships

Documents Containing Data as a Separate Research Result

The decomposition of a scientific publication into a single study does not mean that the “Research article” loses its significance, but rather that its function changes. The research article increasingly

serves as a structured interpretation of the research question and key findings, while raw data, code, metadata, and methodological materials can exist as separate digital objects (Thibault et al., 2023; Gomez-Diaz & Recio, 2023; Nygren, 2024). Gomez-Diaz and Recio (2023) characterize articles, software, and data as an interconnected system of research outputs that requires comprehensive analysis within the framework of contemporary scientific communication. Nygren (2024) emphasizes that open data is a necessary prerequisite for a complete assessment of scientific results and the actual reproducibility of research. One of the most telling manifestations of this transformation is the emergence of the “data paper” a publication specifically focused on describing a research dataset, its origin, structure, acquisition methods, quality, and potential for reuse. Unlike a traditional research article, the primary focus of a data paper is not a new interpretation of a scientific problem, but rather a systematic presentation of a dataset as a reproducible and citable scientific resource. At the same time, the proliferation of data papers cannot be viewed as an isolated editorial innovation; rather, it is part of the transformation from “Open Science by Publication” to “Open Science by Design,” in which the principles of openness are established as early as the planning and execution stages of research (Wittenburg, 2021). In such a system, the article, data, software, metadata, and methods form an integrated digital structure, and their interconnections enable machine search, interoperability, and reuse. A distinct level of this transformation is the shift from simply making resources open to semantically linking them. The concept of “enhanced publication” involves establishing connections between articles, datasets, and other research resources (Sales & Sayão, 2018). In this model, a publication ceases to be an isolated PDF document and becomes a network of digital objects that can be jointly discovered, analyzed, and reused. A further development of this concept is represented by the “publication beyond the PDF” approach, which focuses on a structured and semantically rich representation of scientific information (Procko & Ochoa, 2024). The practical feasibility of this approach is demonstrated by the work of Palmblad et al. (2022), in which semantic annotation of experimental methods was used to construct semantic graphs based on open-access publications in analytical chemistry.

The operational foundation of this transformation is the FAIR principles: Findable, Accessible, Interoperable, Reusable which mean that an open resource must not only be accessible for viewing but also findable, retrievable, integrable with other systems, and reusable (Umbach, 2024; Nygren, 2024). A FAIR-oriented research resource requires a stable DOI identifier, complete metadata, a standardized format, information about its provenance, and terms of use. Umbach (2024) views Open Access, Open Data, and FAIR publishing as interconnected components of the transformation of modern data-driven research. Wittenburg (2021) emphasizes that these principles must be integrated not only at the publication stage but already at the research design stage. Cloern and Soranno (2019) demonstrated the practical implementation of this format within the ASLO system, where “Data Articles” were introduced as a mechanism for disseminating and reusing valuable datasets in the aquatic sciences. Empirical studies generally confirm the gradual strengthening of the role of open research resources. Dos Santos Rocha et al., 2023, highlight the gap between stated commitment to Open Science and actual data sharing, while Cao et al., 2023, after analyzing over a million publications, found a long-term increase in the proportion of articles citing data and methods. At the same time, the authors identified a

concentration of such resources in a relatively limited number of domains and demonstrated a link between the presence of references to data and methods and higher citation rates, with this effect being more pronounced for active links. Thus, modern scientific communication is gradually shifting from a model in which a paper is a self-contained and largely isolated outcome to a model of a network of interconnected research resources. Articles, data, code, metadata, and methods form a structured evidence ecosystem, and the data paper is one of the most prominent institutionalized formats for presenting data as an independent scientific product (Hitzler et al., 2021; Ramachandran et al., 2021; Thibault et al., 2023; Gomez-Diaz & Recio, 2023; Nygren, 2024; Cloern & Soranno, 2019; Pinto, 2024). The next stage of this transformation involves not only open access to individual components but also their semantic, standardized, and machine-processable integration (Zakharova, 2023; Sales & Sayão, 2018; Procko & Ochoa, 2024; Palmblad et al., 2022; Wittenburg, 2021).

Research Gaps, Purpose, and Research Questions

Despite the rapid spread of Open Science principles and the growing importance of open research data, the current literature still does not provide a comprehensive overview of the data paper as a distinct type of scientific publication. Research indicates a rise in the open sharing of data and methods, but at the same time points to a significant gap between declarative support for Open Science and the actual accessibility of research resources (Thibault et al., 2023; Dos Santos Rocha et al., 2023; Cao et al., 2023). A significant portion of the research focuses on the conceptual foundations of Open Science, the FAIR principles, open data, repositories, reproducibility, and data-sharing practices, while the actual dynamics of data paper publications remain under-researched. A data paper is an institutionalized publishing format whose emergence and dissemination depend on journal editorial policies, the emergence of specialized data-centric journals, indexing criteria, and the recognition of data as a distinct scientific product requiring separate analysis. Global trends in Open Science do not necessarily replicate at the level of individual national research systems, as the number of publications may be concentrated in a limited group of countries with developed research infrastructure, a high proportion of international collaboration, well-established repositories, and established mechanisms for funding open access. At the same time, countries with fewer resources may follow a different trajectory even when they formally support the principles of Open Science. For most bibliometric studies, the author’s country of origin is just one of many attributes of a publication, while the specific trajectories of individual countries remain under-researched.

One of the most significant factors influencing researchers’ ability to participate in Gold Open Access is the Article Processing Charge (APC). In the APC model, the costs of open publication are partially or fully shifted from the reader to the author, their institution, a grant program, or another funding source; this issue is particularly relevant for Ukraine. The aim of this study is to identify the global evolution and national differentiation of data papers in the context of the development of Open Science, with a particular focus on the dynamics, structure, and positioning of Ukraine. Moreover, an analysis of Ukrainian publication activity in the field of data papers makes it possible to assess not only the country’s quantitative presence in this segment of global scientific communication but also the nature of its thematic, journal-based, and temporal differentiation.

Materials and Methods

For the empirical analysis of publication activity in this study, the Scopus database was used as the primary source of bibliometric data, since this type of publication is indexed under the article type “Data Paper.” Therefore, the Scopus scientometric database is considered in this study not as a complete registry of all scientific publications, but as a standardized source for comparative analysis within a single indexed information environment. In Scopus, the “Data Paper” document type has a separate designation, “dp” (Data Paper), which allows for the creation of a targeted sample without mixing it with other document types, such as articles, reviews, and conference proceedings. The study analyzed records indexed in Scopus as “Data Papers.” In this study, a Data Paper refers specifically to a Scopus record assigned the document type “Data Paper (dp),” irrespective of the terminology used by the publishing journal. Accordingly, the analysis concerns the bibliographic classification of publications in Scopus rather than all publications that may be described by publishers or authors as data papers, data articles, data descriptors, or related formats. The article separately examines the dynamics of publications of this type on a global scale and for Ukraine. TThe Scopus search was conducted using the Advanced Search interface with the document-type field DOCTYPE(dp), which identifies records classified by Scopus as “Data Paper.” No additional keyword restriction was applied to the retrieval of Data Papers. The search was performed on 27 August 2026, and the retrieved records were exported for subsequent bibliometric analysis. The search strategy was applied consistently to the global dataset and to the subset of publications associated with Ukraine. After forming the initial sample, the records underwent structured processing based on key bibliometric characteristics. In the first stage, the uniqueness of bibliographic records was verified; duplicates that may have arisen due to repeated export or overlapping search results were removed based on document identifiers. The geographic origin of the publications was determined based on the authors’ country of affiliation as indicated in the Scopus metadata; documents related to Ukraine were identified separately, which made it possible to calculate Ukraine’s share in the global publication corpus and compare Ukraine’s publication volume with that of other countries. For each document, we used the subject areas defined and indexed by Scopus; for each subject area, we determined the number and share of documents in the study sample. Furthermore, publication sources were grouped by journals in which Data Paper-type documents were indexed. We determined the absolute number of Data Paper-type documents published in each calendar year. This indicator was used to construct time series and assess the development of this type of scientific publication. Publications from 2014 to 2026 were analyzed on a global scale, and those from 2016 to 2026 were analyzed for Ukraine.

Descriptive statistics were used to characterize the temporal distribution of Data Papers associated with Ukraine. For the complete years 2016–2025, the arithmetic mean, median, standard deviation (SD), minimum, maximum, range, and coefficient of variation (CV) were calculated. Interannual variability was additionally characterized by the mean absolute year-to-year change. A simple linear regression model was used to estimate the temporal trend in annual publication output, with calendar year as the independent variable and the number of Data Papers as the dependent variable. The slope, coefficient of determination (R^2), and p-value were reported. The partial-year 2026 observation was excluded from

these calculations. An important limitation of the dataset concerns the use of the Scopus “Data Paper” document-type classification. Previous comparative research has demonstrated inconsistencies in the coverage and assignment of the Data Paper document type across scholarly databases and journals, including within Scopus. Therefore, the present dataset should be interpreted as a corpus of Scopus-indexed records classified as Data Papers rather than as an exhaustive census of all data papers published worldwide or in Ukraine. This limitation may result in both undercoverage and classification discrepancies and should be considered when interpreting absolute publication counts and temporal trends.

The Global Evolution of Data Paper Over Time and in Ukraine

The distribution of Data Papers over time (Fig. 1) demonstrates a clear global trend toward the emergence and subsequent growth of this type of scientific communication. As shown in Fig. 1a, in 2014, only a few publications were recorded in the Scopus sample; subsequently, over the following years, there was a general, though uneven, increase in the number of Data Papers. By 2018, the annual figure had reached approximately 1,500 documents, and in 2020 it exceeded 2,000. In 2021, there was a temporary decline, after which publication activity rose again in 2022 and reached about 2,300 documents in 2024. An important prerequisite for the observed growth in the number of Data Papers was the creation in 2014 of several specialized platforms for publishing research data. In particular, Scientific Data was launched by the Nature Publishing Group in 2014 as a peer-reviewed open-access journal, with Data Descriptors becoming its primary publication type (Welcome, 2014). At the same time, the Geoscience Data Journal (Allan, 2014) and Data in Brief began operations in 2014. The study by Chavan and Penev (2011) proposed the data paper as a mechanism to encourage the publication and reuse of data, defining it as a scientific publication whose primary function is to provide a structured description of an accessible dataset, while the study by Candela et al. (2015) documented the emergence of a distinct class of data journals and their role in developing new models for publishing research data. Data journals created an additional mechanism for researchers to gain scientific recognition for preparing, documenting, and disseminating datasets—that is, they provided an incentive that the traditional publication model often failed to offer. A study by Walters (2020) showed that data journals integrate access to data and its documentation into the standard system of scientific communication through peer review, publication, and citation. Particularly telling is the acceleration of global trends after 2018, with a shift from 1,500 documents in 2018–2019 to over 2,000 in 2020 and a new high of 2,300 in 2024. (Fig. 1a), this trend aligns with a broader shift toward data-intensive research and the growing institutional significance of research data. Another important methodological factor is the change in the coverage and classification of bibliographic databases; in particular, a study by Jiao et al. (2023) found that while data journals are indexed in Scopus, Web of Science, Dimensions, and OpenAlex, the “Data Paper” document type is assigned inconsistently. The figure for 2026 in Fig. 1a does not represent a complete annual figure, as the sample was compiled as of August 2026; accordingly, the decline to approximately 0.85–0.9 thousand documents in 2026 cannot be interpreted as an actual reduction in global publication activity.

Compared to the global trend, the Ukrainian trend is on a fundamentally different scale (Fig. 1b). The first recorded publications by Ukrainian authors date back to 2016, when only 2 documents were registered; in 2017, the figure dropped to 1, and then rose to 4 documents in 2018. Subsequent trends are characterized by fluctuations: 3 publications in 2019, 7 in 2020, 5 in 2021, a sharp increase to 10 in 2022, a decline to 6 in 2023, a renewed increase to 9 in 2024, and a further decline to 6 documents in 2025. Thus, unlike the global trend, where absolute volumes are measured in thousands of publications, Ukrainian activity remains low in volume and volatile. The peak value for the entire period under review 10 papers in 2022 represents less than 0.5% of the global level; consequently, Ukraine does not make a significant quantitative contribution to global data paper production, though its presence in this segment has been stable since 2016. The appearance of the first Ukrainian documents in 2016 does not mean that the practice of open data publication emerged in Ukraine specifically that year, since national research data had been published earlier in other formats, including through repositories, thematic databases, and traditional scientific articles. However, Fig. 1b specifically reflects the point at which such activity began to be recorded in Scopus's established search results for the "Data Paper" document type. At the same time, there are grounds to believe that the actual scope of Ukrainian work with open data is broader than what the Data Paper sample in Scopus reveals. Boiko et al. (2021), in analyzing the dissemination of open research data in Ukraine and globally, highlighted differences between Ukraine and leading European countries in the development of research data repositories and data-sharing practices. Additionally, the number of "Data Paper" documents cannot be used as the sole indicator of the maturity of Ukraine's open data system. These results allow us to draw an important interim conclusion: the global increase in Scopus-indexed Data Papers indicates the expanding presence of this publication format within the indexed scholarly communication system, whereas the Ukrainian trajectory is characterized by early, yet quantitatively limited and unstable, integration into this segment. At the same time, Fig. 1 reveals another fundamental feature: the Ukrainian trend is not a scaled-down version of the global one; the global series after 2018 shows a high level of productivity with peaks in 2020 and 2024, whereas the Ukrainian series exhibits distinct local peaks, primarily in 2020, 2022, and 2024.

The geographic distribution of documents indexed in Scopus as Data Papers reveals a pronounced concentration of publication activity in a limited group of countries. As shown in Fig. 2a, the United States is the undisputed leader, followed by China; Germany, the United Kingdom, Italy, France, Spain, India, Canada, and Australia rank next with significantly lower figures. Thus, global Data Paper output is characterized by significant concentration within leading scientific systems. This structure generally corresponds to the broader geography of contemporary scientific output. The concentration of Data Paper output in the leading countries may partly reflect differences in the scale of national research systems, publication capacity, institutional infrastructure, international collaboration, and access to research funding. However, these factors were not directly measured in the present study; therefore, the observed geographical distribution should be interpreted descriptively rather than as evidence of specific causal mechanisms. It is particularly telling that the vast majority of leading countries belong to the group of high-income countries with developed scientific infrastructure (the United States, Germany, the United Kingdom, France, Italy, Spain, Canada, and Australia). At the same time, Ukraine occupies a

fundamentally different position. As shown in Fig. 2b, over the entire period, the Scopus sample identified 57 documents classified as Data Papers. This means that Ukraine's absolute figure is approximately 65 times smaller than that of the United States and approximately 44 times smaller than that of China. Ukraine is not among the global leaders in terms of the absolute volume of Data Paper publishing, and its share of the global total remains very small (Fig. 2). Thus, unlike the global trend, where annual publication volumes are measured in thousands of documents, Ukrainian activity remains quantitatively limited and shows substantial interannual variability. For 2016–2025, the mean annual output was 5.30 documents (SD = 2.91), corresponding to a coefficient of variation of 54.86%. At the same time, the Ukrainian series exhibited a positive linear trend of approximately 0.73 documents per year ($R^2 = 0.583$, $p = 0.010$). The Ukrainian trajectory therefore combines a gradual long-term increase with pronounced year-to-year fluctuations rather than a stable, uniform growth pattern. Despite considerable year-to-year fluctuations, the annual series demonstrated a positive linear temporal trend. The estimated regression slope was 0.73 documents per year, with the model explaining 58.3% of the variance in annual publication output ($R^2 = 0.583$; $p = 0.010$). Thus, the observed Ukrainian trajectory combined an overall increase in Data Paper production with pronounced interannual variability. The highest annual output was recorded in 2022 (10 documents), whereas the lowest was observed in 2017 (1 document). The 2026 observation was excluded from the statistical calculations because it represents only the records indexed by 27 August 2026 and therefore does not constitute a complete annual observation (Table 2).

Table 2
Descriptive statistics and temporal trend of Scopus-indexed Data Papers associated with Ukraine, 2016–2025

Indicator	Value
Number of complete observation years, n	10
Total number of Data Papers	53
Mean annual output, documents	5.30
Median annual output, documents	5.50
SD, documents	2.91
Minimum annual output, documents	1
Maximum annual output, documents	10
Range, documents	9
Mean absolute year-to-year change, documents/year	2.89
Linear regression slope, documents/year	0.73
CV, %	54.86
R^2	0.583
p-value	0.010

One of the factors that must be taken into account when interpreting this international disparity is the APC, since in Gold Open Access, the APC is one of the main mechanisms for funding professional scholarly publishing. A study by Solomon and Björk (2012) revealed significant variation in APCs depending on the journal, discipline, and publisher, as well as higher fees in high-impact journals published by major international publishers. Importantly, APCs are not uniform for all authors; publishers are increasingly applying differential pricing based on the author's country of affiliation, as well as waiver and discount mechanisms, and the author's country of origin may be factored into the pricing structure as a means of lowering financial barriers for researchers from lower-income countries. This mechanism is particularly important for the analysis of Ukraine; Elsevier's current policy provides for Geographical Pricing for Open Access (GPOA) for approximately 300 fully Gold OA journals (https://www.elsevier.com/en-gb/about/policies-and-standards/pricing). Under the current scheme, which is based on the World Bank's 2026 classification, Ukraine is categorized as Upper-Middle-Income Group 1, for which the standard geographic price is 45% of the list APC. This means that in journals covered by the GPOA, the standard price for the corresponding group of authors from Ukraine effectively provides a 55% discount off the list price. Under Springer Nature's current policy for Ukraine, the standard APC waiver was temporarily increased to 100% following the start of the full-scale Russian invasion (https://www.springernature.com/jp/open-science/policies/journal-policies/apc-waiver-countries). This allows Ukraine to be viewed as a special case within the global APC-based Open Access system; on the

one hand, its research system is significantly smaller in scale than those of the United States, China, Germany, or the United Kingdom; and on the other hand, Ukrainian authors have received special support mechanisms since 2022 that substantially reduce or completely eliminate the APC barrier for international publishing portfolios. The study by Kaliuzhna and Hauschke (2024) confirms that the financial component is of particular importance for the Ukrainian system; after analyzing 186,000 publications by Ukrainian authors from 2012 to 2021, the authors found a significant prevalence of Open Access, but noted that, prior to the introduction of large-scale specialized support mechanisms, the lack of targeted national funding for APCs limited Ukrainian researchers' ability to publish in Gold OA. Demeter & Istratii (2020) showed that the higher scientific prestige of journals may be associated with higher APCs, which potentially exacerbates inequality in access to prestigious Open Access publishing channels. In this regard, countries with large research budgets not only have more researchers and publications but also greater capacity to cover the costs of publishing in highly ranked Open Access journals. Therefore, special APC support mechanisms create more favorable conditions for Ukrainian authors to access Gold OA than would be expected based solely on the country's overall economic situation.

Publishing Platforms, Journal Specializations, and Accessibility of Publication

The distribution of papers across specialized journals shows that global publication activity for Data Papers is uneven and largely concentrated around a limited group of data-centric journals. In the Scopus sample, these main platforms include Data (MDPI), Chemical Data Collections (Elsevier), Data in Brief (Elsevier), Scientific Data (Nature Portfolio/Springer Nature), and Earth System Science Data (Copernicus Publications) (Fig. 3). The most pronounced global trajectory in Fig. 3a is that of *Data in Brief*; following rapid growth in the second half of the 2010s, the journal reached the highest levels among the publications presented, with local peaks around 2018 and 2020, after which its annual output declined but remained substantial. Data in Brief is a multidisciplinary peer-reviewed open-access journal; the current stated APC (as of August 27, 2026) is USD 1,600; the actual APC amount may be adjusted in accordance with the GPOA; for authors affiliated with institutions in Ukraine, APC waivers are currently available at a rate of 45%. Publications in the Scientific Data journal have been growing rapidly since the early 2020s and will peak in 2024; this is a multidisciplinary, peer-reviewed open-access journal with a current stated APC of USD 2,690 (as of August 27, 2026); and 100% APC waivers are provided for authors affiliated with institutions in Ukraine (Research4Life countries). The Chemical Data Collections represent a more narrowly specialized model focused on chemical research data; however, the journal is hybrid, so publication can occur either via open access or through a subscription model; as of August 27, 2026, the stated APC for making an article openly accessible is USD 2,080 (August 27, 2026). Chemical Data Collections has a significantly smaller global output compared to Data in Brief and Scientific Data; however, its share in the national sample is disproportionately high (Fig. 3b), at the same time, it is important to consider the access model for these publications: all articles by Ukrainian authors in The Chemical Data Collections identified in the sample were published under a subscription model, meaning

without open access to the full text for an unlimited audience. The journal Data, published by MDPI, is a multidisciplinary peer-reviewed open-access journal; the base APC for Data is CHF 1,600 (as of August 27, 2026), and APC waivers for authors from Ukraine may be granted upon request. The Earth System Science Data journal specializes in Earth and environmental science data; the official APC for ESSD is EUR 1,400 (as of August 27, 2026), and 100% APC waivers are provided for authors affiliated with institutions in Ukraine (Research4Life countries). The distribution of publication output trends for documents indexed as “Data Paper” by authors from Ukraine (Fig. 3b) shows that Data and Data in Brief were prominent in the early stages; the role of Chemical Data Collections increased significantly in 2020–2024; and in 2025–2026, Scientific Data makes an appearance, while Earth System Science Data is represented by a much smaller number of documents (Fig. 3b).

In Data Paper publications from data-centric journals, the FAIR principles of open science are implemented through specific publishing mechanisms. Findability is ensured by assigning a DOI to the publication and a persistent identifier to the dataset, as well as through structured metadata, indexing, and formal data citation. Accessibility is supported by mandatory or recommended deposit of the dataset in an accessible repository and the establishment of a direct link between the article and the data. Interoperability is enhanced through the use of standardized metadata, disciplinary formats, and machine-readable data structures. Reusability is ensured by describing provenance, collection and processing methods, technical validation, licensing terms, and usage notes. The publication of a Data Paper under a subscription model in the context of Open Science is not sufficient grounds for concluding that research data is fully open, as it is necessary to distinguish between at least three components: (1) the openness of the full text of the article; (2) the accessibility of the dataset itself; (3) the possibility of reusing it in accordance with licensing and technical conditions. Thus, the subscription-based publication of Data Papers can ensure citability, bibliographic identification, and scientific documentation of the data, but does not guarantee their actual accessibility to all potential users; therefore, in the case of a subscription model for Data Paper publication, it is more appropriate to speak not of the full implementation of Open Science, but of the partial implementation of data-centric publishing. The journal creates an infrastructure for the formal presentation, citation, and documentation of research data; however, subscription-based access to the article and potential restrictions on access to associated datasets may reduce the actual openness of such resources. This differs fundamentally from a fully open model, in which the article and the corresponding dataset are available without authorization or financial barriers.

Subject Structure

The distribution of documents indexed in Scopus as “Data Papers” by subject area reveals significant differences between the global and Ukrainian models of data-centric publishing (Fig. 4). At the global level (Fig. 4a), the dominant subject area is “Multidisciplinary,” accounting for 43.5% of documents; a share that significantly exceeds that of any single discipline and indicates the fundamentally interdisciplinary nature of a significant portion of global Data Paper publishing; this is typical for data-

centric publications, since a single dataset can be relevant to several scientific fields simultaneously. Following the multidisciplinary segment, the global distribution is characterized by nearly equal shares of Computer Science, Social Sciences, and Mathematics. The Ukrainian distribution profile has a different configuration (Fig. 4b); unlike the global structure, in Ukraine the three largest subject areas—Computer Science, Multidisciplinary, and Social Sciences—each have an equal share of 17.3% and do not show a concentration in the multidisciplinary category. At the same time, these data require cautious interpretation due to the small absolute size of the Ukrainian sample. Thus, the disciplinary structure of Data Papers demonstrates a quantitative difference between the global and Ukrainian samples, with distinct distributions of subject areas. These differences should be interpreted as differences in the disciplinary composition of Scopus-indexed Data Papers rather than as evidence of fundamentally different models of integrating data-centric publishing into national research systems. At the global level, the dominance of “Multidisciplinary” in combination with “Computer Science,” “Social Sciences,” and “Mathematics” indicates the widespread institutionalization of data papers as an interdisciplinary mechanism for presenting and reusing research data. The Ukrainian structure is less multidisciplinary in focus and is characterized by the nearly equal roles of “Computer Science,” “Multidisciplinary,” and “Social Sciences.”

The concept of decomposing a scientific publication of a single study by publishing a Data Paper takes on practical significance when integrity is ensured and a set of interrelated scientific objects-datasets, methodological descriptions, analytical results, and full-text research articles—is presented. A practical example of such decomposition is provided by the studies conducted by the Ukrainian authors of this paper. In each of the two examples, one of the research datasets was presented not only as a traditional full-text research article but as a collection of separate, functionally interrelated publications, including a Data Paper. Therefore, the first example concerns a study of the physicochemical characteristics of water in a nuclear power plant (NPP) cooling system and the processes of scale formation and corrosion activity. The first level of decomposition was achieved by separately presenting the original dataset in the journal Data in Brief, and the article by Kuznietsov (2024) was published as a Data Paper and contains data on the physicochemical monitoring of the cooling system, specifically pH values, temperature, total dissolved solids, total hardness, and total alkalinity, as well as calculated indices. This data—which, in the traditional publication model, might have remained primarily supplementary material to the main article, was presented as an independent scientific resource with its own DOI.The second level of decomposition was implemented through a separate methodological presentation of the research in the journal MethodsX; the article by Biedunkova et al., 2025, systematized the methodological approaches to the comprehensive assessment of the cooling system. The third level is represented by the complete research results in the article by Biedunkova et al., 2026, published in Energy Science & Engineering, in which the research results are no longer considered in isolation as a dataset or methodological protocol, but rather in the context of the operation of an open-loop nuclear power plant cooling system, the behavior of dissolved inorganic salts, and the formation of water discharge. The second example concerns a study of potentially toxic elements in the water of the Styr River; at the first level, the research dataset was presented as a Data Paper, the article by Biedunkova

and Kuznietsov, 2024, published in the journal Scientific Data. The second level of presentation for this example is realized in a full-length scientific research article by Biedunkova et al., 2025a, published in the Chemosphere.

The examples discussed demonstrate the practical value of the Data Paper as an independent format for presenting scientific results within the Open Science paradigm. Unlike the traditional model of scientific publication, the Data Paper enables the dataset itself to be presented as a separate research result, complete with the ability to describe, cite, and reuse it. The authors' practical experience confirms the feasibility of this approach. Accordingly, the Data Paper creates conditions for data to be independently discovered by other researchers, verified, reanalyzed, and used to generate new scientific results, since a full-text article typically focuses on the scientific problem, methodology, results, and their interpretation. For environmental research, this capability is particularly important, as observational results can be reused for comparative, spatial, temporal, or statistical analysis.

Conclusions

The analysis demonstrates a sustained expansion of Scopus-indexed Data Paper publishing at the global level, although the observed 2026 values represent only a partial year and should not be interpreted as a complete annual output. The geographical distribution is concentrated among countries with high volumes of indexed research output, while Ukraine accounts for a quantitatively limited share of the identified Data Paper corpus. For Ukraine, 57 Data Papers were identified for 2016–August 2026, including 53 documents published during the complete years 2016–2025 and 4 records available by 27 August 2026. Thus, Ukrainian participation in the Data Paper segment can be characterized as persistent but quantitatively limited and temporally variable. The disciplinary structure of the Ukrainian sample differs from the global distribution, with Computer Science, Multidisciplinary, and Social Sciences representing prominent components of the national sample. These differences should be interpreted as differences in the disciplinary composition of Scopus-indexed Data Papers rather than as evidence of fundamentally different national models of Open Science development. The findings also demonstrate the importance of distinguishing Data Paper publishing from Open Data and Open Science as broader concepts. The presence of a Data Paper in Scopus does not by itself establish unrestricted access to the associated dataset, its reusability, or the full implementation of Open Science practices. Finally, the results should be interpreted within the methodological limitations of Scopus-based document-type classification. The study measures participation in the Scopus-indexed Data Paper segment rather than the total volume or maturity of research-data publishing and Open Science practices in Ukraine.

Declarations

Ethical approval:

Not applicable. This study did not involve human participants or animals, and therefore ethical approval was not required.

Consent to participate:

Not applicable.

Consent to publish:

Not applicable.

Competing interests:

The authors declare no competing interests.

Clinical trial number

not applicable.

Funding statement:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

P. K.: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review and editing. O. B.: Conceptualization, Methodology, Investigation, Validation, Supervision, Writing – review and editing. O. P.: Investigation, Data curation, Formal analysis, Validation, Writing – review and editing. Ihor Statnyk: Investigation, Methodology, Data curation, Visualization, Writing – review and editing. All authors reviewed the manuscript.

Data Availability

Data is provided within the manuscript.

References

1. Allan, R. (2014). Geoscience data. Geoscience Data Journal, 1(1), 1–1. DOI: https://doi.org/10.1002/gdj3.3
2. Björk, B.-C., & Solomon, D. J. (2012). Pricing principles used by scholarly open access publishers. Learned Publishing, 25(2), 132–137. https://doi.org/10.1087/20120207

3. Boiko, A., Kramarenko, O., & Shabanov, S. (2021). The state of dissemination of open research data in Ukraine and the world: Bibliometric analysis. European Journal of Management Issues, 29(2), 82–92. https://doi.org/10.15421/192120
4. Biedunkova, O., & Kuznietsov, P. (2024). Dataset on heavy metal pollution assessment in freshwater ecosystems. Scientific Data, 11, 1241. https://doi.org/10.1038/s41597-024-04116-z
5. Biedunkova, O., Kuznietsov, P., Grokhovska, Y., Konontcev, S. (2026). Influence of Technological and Environmental Factors on the Major Ion Composition in Cooling Water Discharges of a Nuclear Power Plant En. Sci. & Eng. 1 1–17. https://doi.org/10.1002/ese3.70437
6. Biedunkova, O., Kuznietsov, P., & Pinchuk, O. (2025). Comprehensive assessment of scale formation, corrosion, and biological pollution in a cooling water supply system. MethodsX, 14, 103154. https://doi.org/10.1016/j.mex.2025.103154
7. Biedunkova, O., Kuznietsov, P., & Mandryk, O. (2025a). Study of the dominant modes of formation and variability of potentially toxic element concentrations and their impact on environmental quality. Chemosphere, 388, 144688. https://doi.org/10.1016/j.chemosphere.2025.144688
8. Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. Journal of the Association for Information Science and Technology, 66(9), 1747–1762. https://doi.org/10.1002/asi.23358
9. Cao H., Dodge J., Lo K., McFarland D.A., Wang L.L. The Rise of Open Science: Tracking the Evolution and Perceived Value of Data and Methods Link-Sharing Practices. arXiv. 2023. DOI: https://doi.org/10.48550/arXiv.2310.03193
10. Chavan, V., & Penev, L. (2011). The data paper: A mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics, 12(Suppl 15), S2. https://doi.org/10.1186/1471-2105-12-S15-S2
11. Cloern J., Soranno P. ASLO Takes a Next Step toward Open Science: Introducing Data Papers, a New Article Type in Limnology & Oceanography Letters. Limnology and Oceanography Bulletin. 2019;28(4):142–143. DOI: https://doi.org/10.1002/lob.10341
12. Demeter, M., & Istratii, R. (2020). Scrutinising what Open Access journals mean for global inequalities. Publishing Research Quarterly, 36, 505–522. https://doi.org/10.1007/s12109-020-09771-9
13. Dos Santos Rocha A., Albrecht E., El-Boghdadly K. Open science should be a pleonasm. Anaesthesia. 2023;78(5):551–556. DOI: 10.1111/anae.15962. https://doi.org/10.1111/anae.15962
14. Gomez-Diaz T., Recio T. Articles, software, data: An Open Science ethological study. Maple Transactions. 2023;3(4):Article 17132. DOI: https://doi.org/10.5206/mt.v3i4.17132
15. Hitzler P., Janowicz K., Shimizu C., Zhou L., Eells A. Open Science data and the Semantic Web journal. Semantic Web. 2021;12(3):401–402. DOI: 10.3233/SW-210427. https://doi.org/10.3233/SW-210427
16. Ioannidis Y. Open Science: A New Paradigm for the Research Lifecycle. Proceedings of the VLDB Endowment. 2025;18(12):5512. DOI: 10.14778/3750601.3760518.

https://doi.org/10.14778/3750601.3760518

17. Jiao, C., Li, K., & Fang, Z. (2023). How are exclusively data journals indexed in major scholarly databases? An examination of four databases. Scientific Data, 10, 737. https://doi.org/10.1038/s41597-023-02625-x
18. Kaliuzhna, N., & Hauschke, C. (2024). Open access in Ukraine: Characteristics and evolution from 2012 to 2021. Quantitative Science Studies, 5(4), 1022–1041. https://doi.org/10.1162/qss_a_00324
19. Kuznietsov, P. (2024). Evaluation of the scaling and corrosive potential of the cooling water supply system of a nuclear power plant based on the physicochemical control dataset. Data in Brief, 54, 110347. https://doi.org/10.1016/j.dib.2024.110347
20. Nygren P. Advancing the Everyone's Right to Forest Science: opening research data and computational codes in Silva Fennica. Silva Fennica. 2024;58(1):Article 24016. DOI: https://doi.org/10.14214/sf.24016
21. Palmblad M., Asein E., Bergman N.P., Ivanova A., Ramasauskas L., Mohammed Reyes H., Ruchti S., Soto-Jácome L., Bergquist J. Semantic Annotation of Experimental Methods in Analytical Chemistry. Analytical Chemistry. 2022;94(44):15464–15471. DOI: https://doi.org/10.1021/acs.analchem.2c03565
22. Pinto F.A.M.A.G. Publicação de dados de pesquisa no contexto da comunicação científica: cenário terminológico-conceitual. Ciência da Informação. 2024;53. DOI: 10.18225/ci.inf.v53i.6668. https://doi.org/10.18225/ci.inf.v53i.6668
23. Procko T., Ochoa O. Semantic Science: Publication Beyond the PDF. Proceedings of IEEE SoutheastCon 2024. 2024. DOI: https://doi.org/10.1109/SoutheastCon52093.2024.10500258
24. Ramachandran R., Bugbee K., Murphy K. From Open Data to Open Science. Earth and Space Science. 2021;8(5):e2020EA001562. DOI: https://doi.org/10.1029/2020EA001562
25. Sales L.F., Sayão L.F. Use of conceptual relations for semantic integration of scientific publications and research data. In: Challenges and Opportunities for Knowledge Organization in the Digital Age. 2018. p. 711–719. DOI: https://doi.org/10.5771/9783956504211-711
26. Solomon, D. J., & Björk, B.-C. (2012). A study of open access journals using article processing charges. Journal of the American Society for Information Science and Technology, 63(8), 1485–1495. https://doi.org/10.1002/asi.22673
27. Thibault R.T., Amaral O.B., Argolo F., Bandrowski A.E., Davidson A.R., Drude N.I. Open Science 2.0: Towards a truly collaborative research ecosystem. PLOS Biology. 2023;21(10):e3002362. DOI: https://doi.org/10.1371/journal.pbio.3002362
28. This H. Articles de résultats, articles d'interprétations: des modes nouveaux de publication des résultats scientifiques / Articles for results, articles for interpretations: new ways of publishing scientific results. Notes académiques de l'Académie d'agriculture de France. 2026. https://www.academie-agriculture.fr/publications/notes-academiques/articles-de-resultats-articles-dinterpretations-des-modes-nouveaux-de

29. Umbach G. Open Science and the impact of Open Access, Open Data, and FAIR publishing principles on data-driven academic research: Towards ever more transparent, accessible, and reproducible academic output? Statistical Journal of the IAOS. 2024;40(1):59–70. DOI: https://doi.org/10.3233/SJI-240021
30. Verma, R., Kumar, A., Pabreja, K. et al. A decade of data science research: insights from a bibliometric study. Discov Data 3, 61 (2025). https://doi.org/10.1007/s44248-025-00097-5
31. Walters, W. H. (2020). Data journals: Incentivizing data access and documentation within the scholarly communication system. Insights, 33, 18. https://doi.org/10.1629/uksg.510
32. Welcome, Scientific Data!. Nature 509, 534 (2014). https://doi.org/10.1038/509534a
33. Wittenburg P. Open Science and Data Science. Data Intelligence. 2021;3(1):1–11. DOI: https://doi.org/10.1162/dint_a_00082
34. Zakharova O.V. Scientific documents metadata as a component of the system of the “open science” information resources. PROBLEMS IN PROGRAMMING. 2023;(4):27–38. DOI: https://doi.org/10.15407/pp2023.04.027

Figures

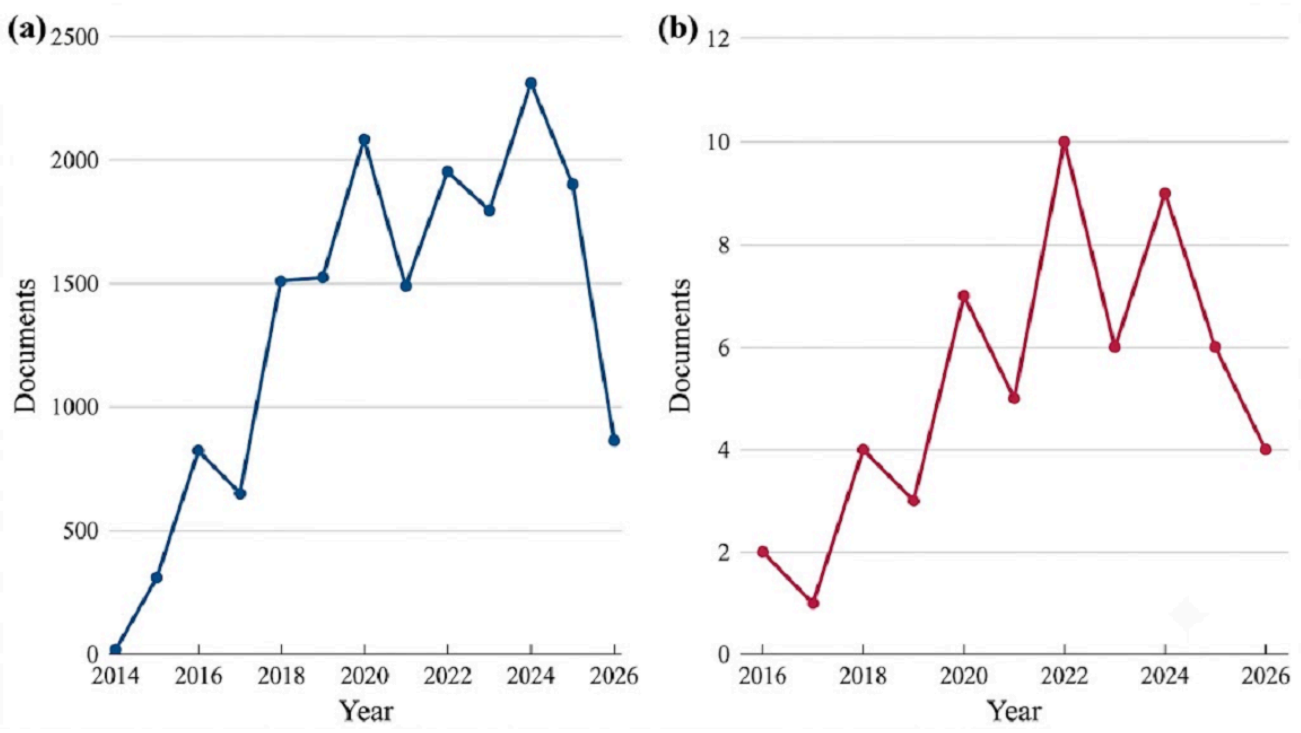


Figure 1

Annual publication output for the "Data Paper" document type: (a) worldwide dynamics (2014 – August 2026); (b) publication activity in Ukraine (2016–2026). Data retrieved from Scopus (https://www.scopus.com/ accessed on 27 August 2026).

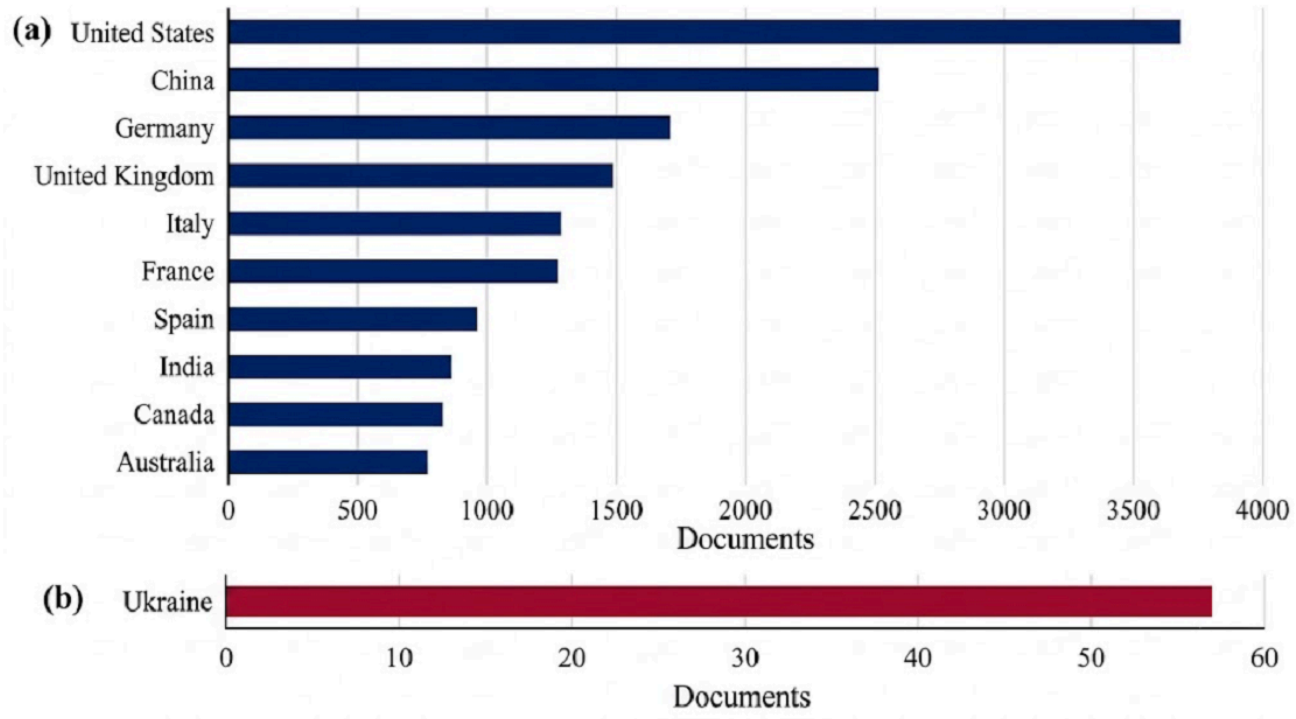


Figure 2

Geographical distribution of documents indexed as "Data Paper" (2014 – 2026): (a) leading countries by publication volume; (b) national publication output of Ukraine. Data retrieved from Scopus (https://www.scopus.com/ accessed on 27 August 2026).

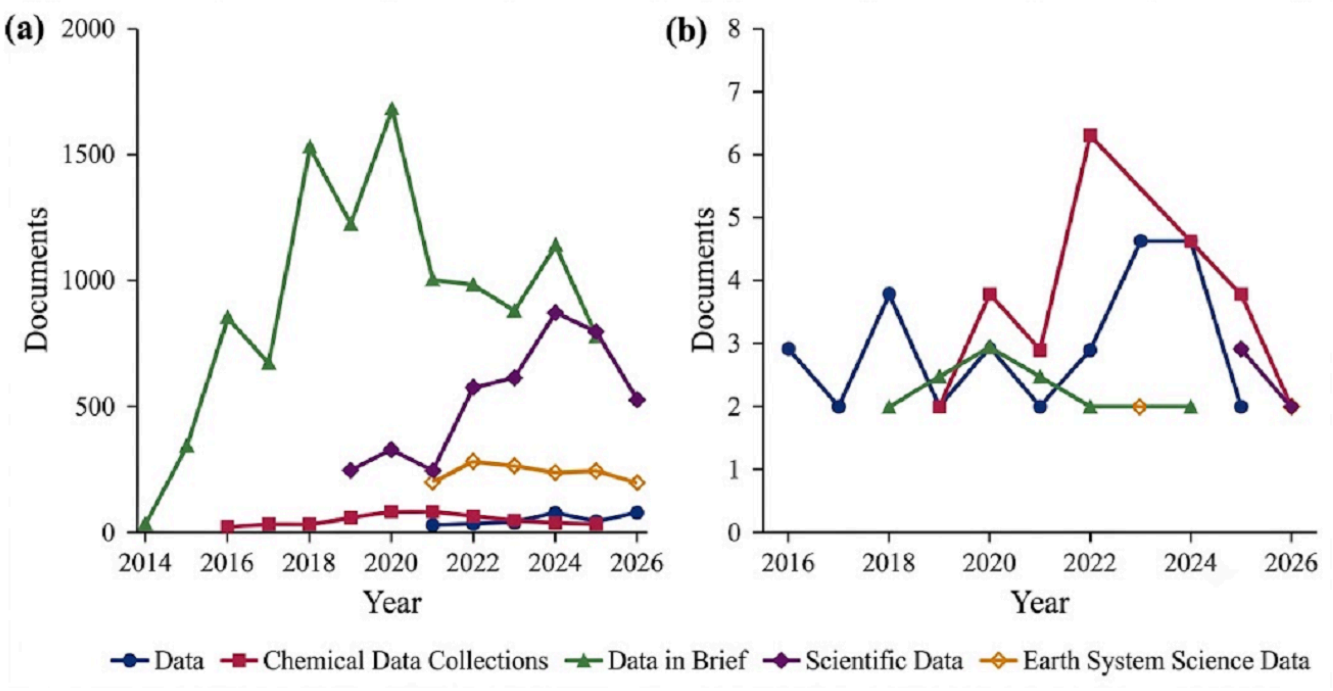


Figure 3

Publication output trends of documents indexed as "Data Paper" in data-centric journals : (a) annual number of published documents across major data journals (2014– 2026); (b) detailed distribution of document output for of Ukraine (2016– 2026). Data retrieved from Scopus (https://www.scopus.com/ accessed on 27 August 2026).

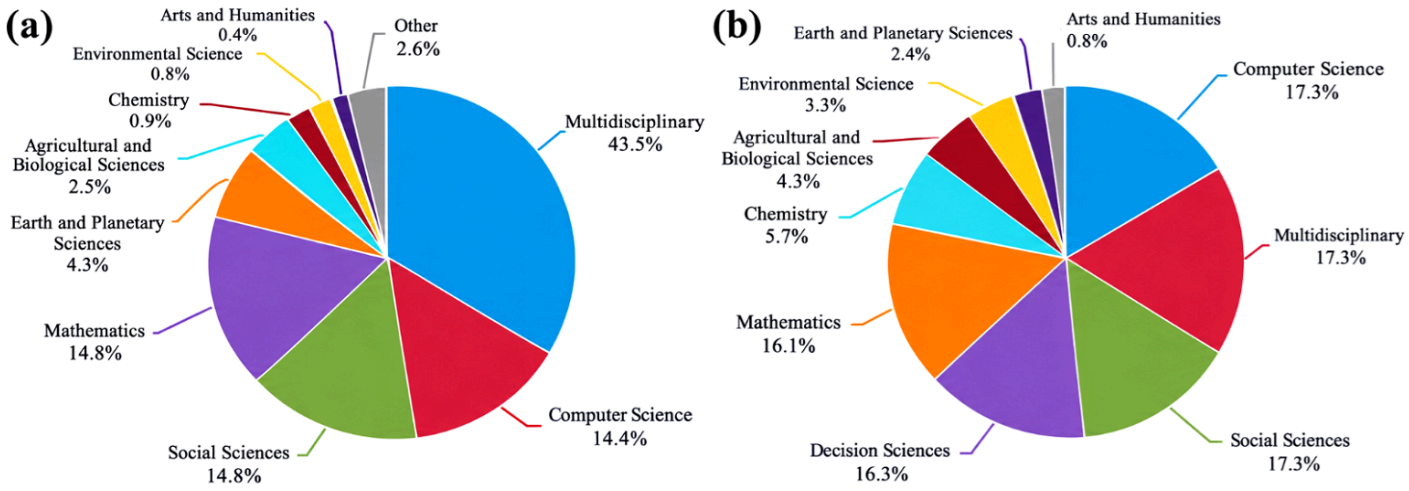


Figure 4

Comparative subject area distribution of documents indexed as "Data Paper": (a) worldwide trends (2014– 2026); (b) Ukraine (2016– 2026). Data retrieved from Scopus (https://www.scopus.com/ accessed on 27 August 2026).