

# Mapping the Global Terrain of OpenDOAR Utilization: Insights from Data Visualization Approach

Manash Esh

Saptarshi Ghosh

## Abstract

*Open-access repositories have become essential platforms for distributing scholarly knowledge globally, fostering accessibility and collaboration within the academic community. Understanding the landscape of OARs and use patterns across different regions is vital for advancing scholarly communication and research dissemination practices. This study maps the global terrain of OpenDOAR (Directory of Open Access Repositories) through a comprehensive analysis using data visualization techniques. Leveraging data from OpenDOAR, insights into the geographical distribution, software preferences, content types, and growth trends of OARs are provided. Advanced data visualization techniques, including dynamic pie charts, line plots, and choropleth maps, offer a nuanced understanding of OAR usage patterns. The analysis reveals significant disparities in repository distribution across continents, with Europe leading in repository count, followed by Asia, North America, South America, and Africa. The United States, Japan, and Germany emerged as top contributors to the repository count. Longitudinal analysis indicates consistent repository growth in the USA, Japan, and Germany, while other countries show varying trends. The analysis of software usage across continents reveals variations in adoption and use within scholarly repository ecosystems, with DSpace emerging as the most widely adopted software globally. This study provides valuable insights into the evolving landscape of OAR usage, offering implications for researchers involved in scholarly communication and research dissemination. Understanding these trends can inform strategic decision-making regarding software selection, resource allocation, and collaborative opportunities within the global scholarly community.*

**Keywords:** OpenDOAR, OAR, Geographical Data, Growth Trends, Data Visualization

## 1. Introduction

Open access repositories (OARs) serve as pivotal platforms for the dissemination of scholarly knowledge on a global scale, fostering accessibility and collaboration within the academic community. Understanding the landscape of OARs and use patterns across different regions is essential for advancing scholarly communication and research dissemination practices. In this context, the present study endeavours to map



the global terrain of OpenDOAR (Directory of Open Access Repositories) through a comprehensive analysis using data visualization techniques. The proliferation of OARs has garnered significant scholarly attention in recent years, reflecting a collective endeavour to democratize access to knowledge and promote collaborative research endeavours. OpenDOAR, a reputable directory cataloguing a diverse array of OARs worldwide, is a rich data source for understanding these repositories' distribution, features, and use dynamics. Leveraging the wealth of information available through OpenDOAR, this study aims to provide insights into the geographical distribution, software preferences, content types, and growth trends of OARs. By using advanced data visualization techniques, including dynamic pie charts, line plots, and choropleth maps, this study offers a nuanced understanding of the global landscape of OpenDOAR usage. Through meticulous aggregation and categorization of data sourced from OpenDOAR, the study endeavours to elucidate the relative contributions of different continents, countries, and institutions to the overall repository landscape. Longitudinal analyses will shed light on temporal trends in repository counts, software usage, and content types, providing valuable insights into the evolving nature of scholarly communication ecosystems. The findings of this study hold implications for researchers, policymakers, and institutional stakeholders involved in scholarly communication and research dissemination by summarizing OAR usage patterns.

## **2. Literature Review**

Exploring open-access repositories (OARs) within the realm of technology has become a focal point of scholarly inquiry, driven by the imperative to enhance global knowledge dissemination. In their seminal work, Wani et al. (2023) started an insightful journey to evaluate the proliferation and maturation of OARs, meticulously investigating attributes such as geographical distribution, software preferences, and content types. Leveraging the reputable OpenDOAR directory, their study illuminated the landscape of technology-focused repositories, revealing intriguing patterns and trends. They discovered a rich tapestry of repositories spanning 125 nations, with the United States emerging as the leader in repository abundance. This critical contribution underscores the pivotal role of OARs in advancing scholarly communication and highlights the importance of multidisciplinary institutional collaborations in repository stewardship.

Similarly, Norris, Oppenheim, and Rowland (2008) conducted a study to discern the efficacy of various search tools in navigating the open-access scholarly literature landscape. Through an exhaustive analysis encompassing ecology, economics, and sociology domains, they shed light on the dominance of general-purpose search engines, notably Google and Google Scholar, in uncovering open-access articles. Their findings advocate for the primacy of these search engines in scholarly research endeavours, emphasizing the importance of search engine optimization in helping with knowledge discovery. Nazim et al. (2022) dug into the nuanced contours of open-access publishing within the Indian scholarly milieu, elucidating the intricate interplay of policy frameworks, institutional dynamics, and technological infrastructures. A comprehensive array of data sources posits that the study unveiled India's burgeoning footprint in open-access publishing. Despite commendable strides, critical gaps in policy implementation and institutional

compliance were identified, highlighting the need for concerted efforts to bridge these fissures. Sofi and Mir (2023) pioneered unraveling the idiosyncratic features of open-access repositories housing patents across Asian nations. Leveraging data from OpenDOAR, their study illuminated unique features and trends underpinning patent repositories in the Asian continent, offering novel insights into repository typologies, software preferences, and language interfaces. Posigha and Eseivo (2022) investigated the impediments obstructing the effective curation of institutional repository contents within the Nigerian and South African academic landscapes. Through a rigorous survey-based inquiry, their study identified challenges ranging from faculty awareness deficits to systemic policy lacunae, emphasizing the need for institutional interventions to ameliorate these barriers. Finally, Pinfield et al. (2014) conducted a comprehensive retrospective analysis of open-access repositories' global evolution, leveraging insights gleaned from the OpenDOAR project. Their seminal work unraveled multifaceted drivers and inhibitors shaping repository adoption trajectories, providing invaluable insights into past growth, current features, and future trajectories of open-access repositories. Collectively, these seminal works offer a rich tapestry of insights into the multifaceted dimensions of open-access repository landscapes, underscoring their pivotal role in advancing scholarly communication, knowledge dissemination, and research stewardship.

### **3. Methodology**

We employed a comprehensive method to scrutinize repositories and software usage dissemination across diverse continents and countries, leveraging Python programming and advanced data visualization techniques. The dataset was sourced from an Excel file, meticulously curated from the URL <https://v2.sherpa.ac.uk/opensoar/> on March 23, 2024, capturing pertinent details encompassing repository tallies, country demographics, and software usage metrics. Our analysis began with meticulous aggregations and categorizations based on continent and country, meticulously tabulating the cumulative repository counts for each entity. This methodological approach helped identify prominent contributors to the overall repository landscape at both the continental and national levels. Then, we translated our findings into visually compelling representations, utilizing dynamic pie charts and line plots to depict the geographical distribution of repositories across continents and spotlight the top ten countries. These visualisations provided lucid insights into evolving trends, effectively illustrating the relative contributions of each continent and country to the cumulative repository count. A longitudinal examination was conducted to scrutinize the top ten countries' repository count trends over time, using line plots to discern nuanced growth patterns, stability, or regression across diverse regions. This analytical endeavour offered valuable insights into the idiosyncratic dynamics of repository proliferation within specific countries. Last, our investigation extended to scrutinize the distribution of software usage across continents, using line plots to elucidate the adoption and use dynamics of varied software platforms. This discerning analysis unearthed pronounced disparities in software adoption across different continents, shedding light on regional predilections and evolving trends within scholarly repository ecosystems.

#### 4. Research Questions

**RQ1:** How does the distribution of repositories of OpenDOAR vary across different continents and top countries, and what are the relative contributions of each continent and country to the overall repository count?

**RQ2:** How have the trends in repository counts evolved for the top ten countries with the highest repository counts, namely the USA, Japan, Germany, Peru, the UK, Turkey, Spain, Croatia, Brazil, and Indonesia?

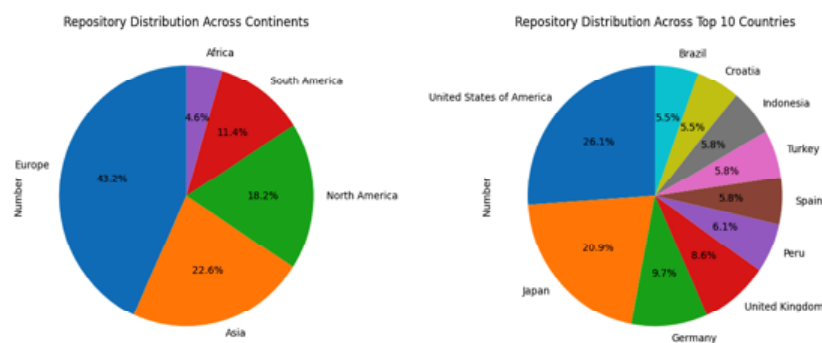
**RQ3:** How does the geographical distribution of repository counts vary across different continents, and what factors contribute to repository establishment and use disparities?

**RQ4:** How does the distribution of software usage vary across different continents, and what insights can be gleaned from the disparities in software adoption and use within scholarly repository ecosystems?

#### 5. Analysis and Discussion

**RQ1:** How does the distribution of repositories of OpenDOAR vary across different continents and top countries, and what are the relative contributions of each continent and country to the overall repository count?

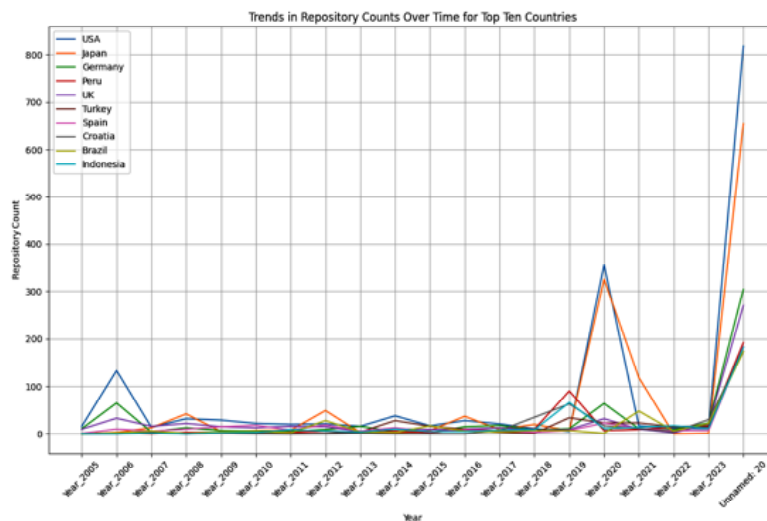
The analysis of the OpenDOAR data provides valuable insights into the distribution of repositories across different continents and top countries. By examining the repository counts, we aim to understand the relative contributions of each continent and country to the overall repository count—the first step in our analysis involved aggregating the repository counts by continent and country. From Figure 1, we can observe significant variations in the distribution of repositories across continents and top countries. In the pie chart representing repository distribution across continents, we identified the top five continents contributing to the overall repository count. This visualization lets us grasp the proportional representation of repositories globally. Also, it highlights the continents with the highest concentration of repositories, providing insight into the geographic distribution of scholarly resources.



**Figure 1: OpenDOAR repository distribution across Continents and the top 10 Countries**

Figure 1, illustrating repository distribution across the top 10 countries, lets us identify the leading contributors to the overall repository count. By examining each country's percentage share of repositories, we can discern the countries with the most substantial presence in the scholarly repository landscape. After analysing the data from OpenDOAR, we observe significant disparities in repository counts across different continents. Europe is the leading continent with the highest repository count, totaling 2500 repositories. Asia follows, with 1310 repositories, while North America and South America stand at 1054 and 659 repositories, respectively. Africa trails behind with 264 repositories. Digging deeper into country-specific data reveals the top contributors to the repository count. The United States of America leads the pack with many repositories, totaling 818. Japan and Germany follow with 655 and 304 repositories, respectively. The United Kingdom, Peru, Spain, Turkey, Indonesia, Croatia, and Brazil also boast significant repository counts, contributing to 3133 repositories. These findings underscore the diverse distribution of repositories globally, reflecting varying levels of research and academic activity across different regions and countries. Further exploration into the factors driving repository establishment and use in these regions could provide valuable insights into scholarly communication trends and research dissemination practices worldwide.

**RQ2:** How have the trends in repository counts evolved for the top ten countries with the highest repository counts, namely the USA, Japan, Germany, Peru, the UK, Turkey, Spain, Croatia, Brazil, and Indonesia?

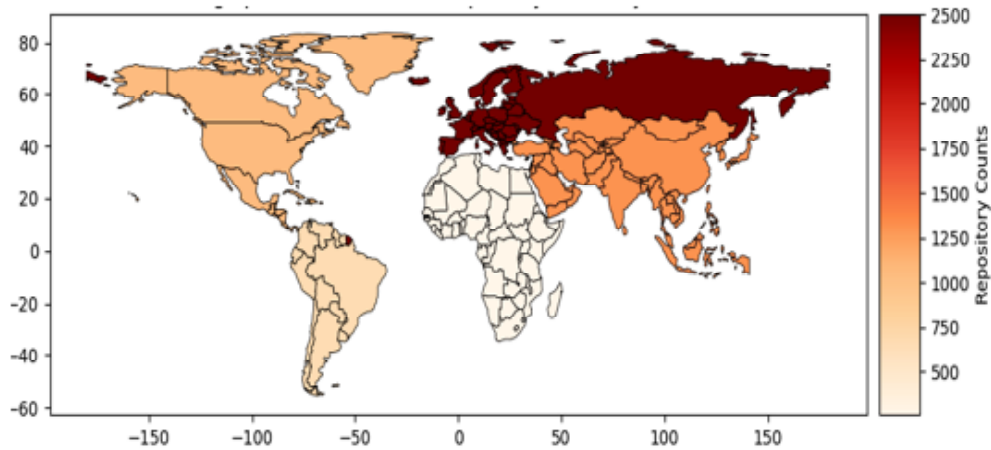


**Figure- 2: Trends in Repository counts over time for the top 10 Countries**

Figure 2 illustrates the trends in repository counts from 2005 to 2023 for the top ten countries, offering valuable insights into the evolution of digital repositories within **OpenDOAR** across diverse regions. The visualization indicates that the USA consistently maintains the highest repository counts throughout the observed period. Japan and Germany also show significant repository growth but with some fluctuations. Peru has recently shown a notable surge in repository counts, suggesting an increasing focus on digital archiving. The UK and Brazil show stable repository counts. Meanwhile, Turkey, Spain, Croatia, and Indonesia

show varying fluctuation levels. This analysis sheds light on the changing global landscape of digital repositories, revealing country-specific trends and patterns.

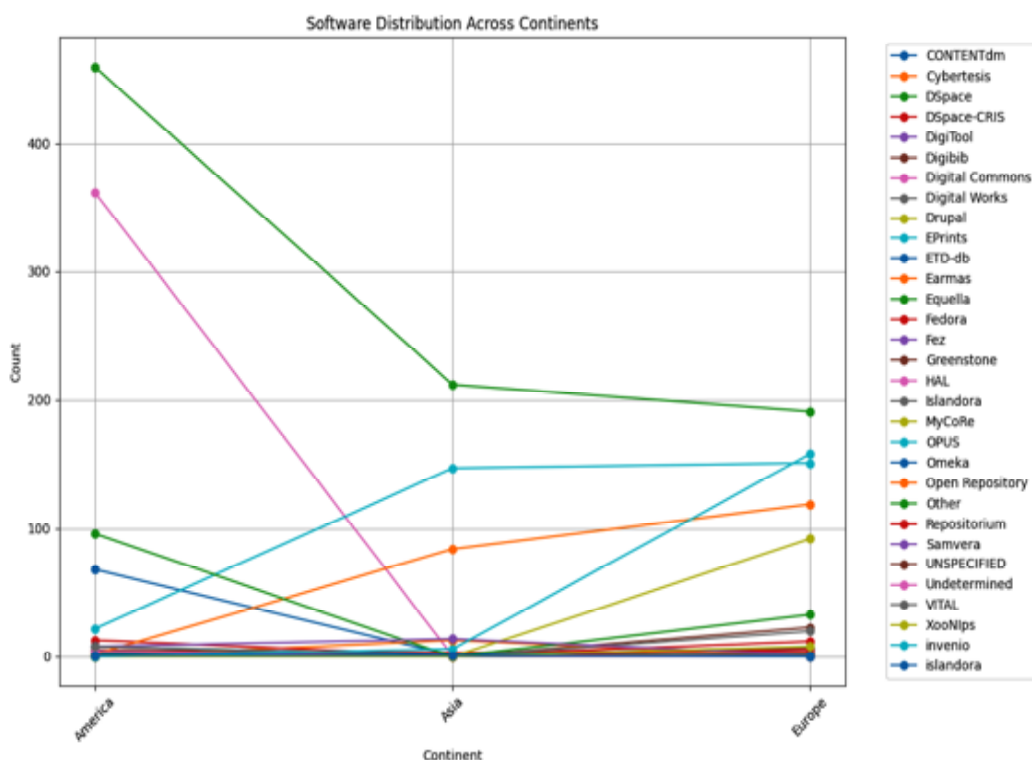
**RQ3:** How does the geographical distribution of repository counts vary across different continents, and what factors contribute to repository establishment and use disparities?



**Figure- 3: Geographical distribution of repository counts by continent**

The geographical distribution of repository counts across different continents reveals notable repository establishment and use disparities. As illustrated in the choropleth map, specific continents exhibit significantly higher repository counts than others. For example, Europe and North America show higher repository counts, suggesting a strong presence of institutional repositories and digital collections in these regions. However, continents such as Africa and South America exhibit comparatively lower repository counts, suggesting potential underdevelopment or limited adoption of digital repositories in these areas. Several factors contribute to these disparities in repository establishment and use. First, socioeconomic factors such as economic development, technological infrastructure, and educational resources play an essential role in shaping the landscape of digital repositories. Developed regions with advanced infrastructure and more significant financial resources are more likely to invest in establishing and maintaining digital repositories, thus leading to higher repository counts. The geographical distribution of repository counts reflects complex interplays between socioeconomic, cultural, and institutional factors.

**RQ4:** How does the distribution of software usage vary across different continents, and what insights can be gleaned from the disparities in software adoption and use within scholarly repository ecosystems?



**Figure 4: Software distribution across continents**

The analysis of software usage across different continents, in Figure 4, reveals notable variations in adoption and use within scholarly repository ecosystems of OpenDOAR. Several essential insights emerge after examining the distribution of software installations across America, Asia, and Europe. First, DSpace seems to be the most widely adopted software across all continents, with the highest number of installations in America, followed by Asia and Europe. This suggests a strong preference for DSpace as a repository platform globally. Second, while some software, such as Digital Commons and EPrints, exhibit consistent levels of adoption across continents, others show significant disparities. For example, MyCoRe shows minimal adoption in America and Asia but emerges as a common choice in Europe, suggesting regional preferences or institutional strategies influencing software selection. The analysis highlights certain software with negligible installations in specific continents. For example, Islandora shows minimal or zero adoption in America and Asia but gains traction in Europe, suggesting potential localized factors influencing software usage patterns. The disparities in software adoption and use across continents underscore the complex interplay of factors shaping scholarly repository ecosystems. Understanding these variations can inform strategic decision-making regarding software selection, resource allocation, and collaborative opportunities within the global scholarly community.

## 6. Findings and Conclusions

- ❖ The analysis of OpenDOAR data provides insights into the distribution of repositories across continents and top countries. Europe is the leading continent in repository count, followed by Asia, North America, South America, and Africa. The United States, Japan, and Germany are the top contributors to the repository count, with significant contributions from the United Kingdom, Peru, Spain, Turkey, Indonesia, Croatia, and Brazil. These findings reflect varying levels of research and academic activity globally.
- ❖ Trends in repository counts over time for the top ten countries reveal consistent repository growth in the USA, Japan, and Germany, while Peru has shown a notable surge in recent years. The UK and Brazil maintain stable counts, while Turkey, Spain, Croatia, and Indonesia show fluctuations. This analysis provides insights into country-specific trends and patterns in digital repository evolution.
- ❖ The geographical distribution of repository counts highlights disparities across continents, with Europe and North America showing higher counts than Africa and South America. Socioeconomic factors such as economic development, technological infrastructure, and educational resources influence repository establishment and use. These disparities reflect complex interplays between socioeconomic, cultural, and institutional factors.
- ❖ Software usage analysis across continents reveals variations in adoption and use within scholarly repository ecosystems. DSpace is the most widely adopted software globally, while other software shows regional disparities. Understanding these variations can inform strategic decision-making regarding software selection and resource allocation within the scholarly community.

## References

1. Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H. S., & Hussain, A. (2014). Open access repositories worldwide, 2005–2012: Past growth, current features, and future possibilities. *Journal of the Association for Information Science and Technology*, 65(12), 2404-2421. <https://doi.org/10.1002/asi.23131>
2. Nazim, M., Bhardwaj, R. K., Agrawal, A., & Bano, A. (2022). Open access publishing in India: Trends and policy perspectives. *Global Knowledge, Memory and Communication*, 71(1), 57-72. <https://doi.org/10.1108/GKMC-09-2021-0158>
3. Norris, M., Oppenheim, C., & Rowland, F. (2008). Finding open access articles using Google, Google Scholar, OAIster, and OpenDOAR. *Online Information Review*, 32(5), 587-600. <https://doi.org/10.1108/14684520810923881>



4. Posigha, E. B., & Eseivo, L. O. (2022). Content recruiting: Obstacle to institutional repository development in Nigeria and South Africa university libraries. *Journal of Librarianship and Information Science*, 54(3), 750-769. <https://doi.org/10.1177/02666669221105468>
5. Sofi, I. A., & Mir, A. A. (2023). Status of patent archives in the Asian continent: A vivid picture from OpenDOAR. *Global Knowledge, Memory and Communication*, 72(1), 112-127. <https://doi.org/10.1108/GKMC-07-2023-0241>
6. Wani, J. A., Sofi, T. A., Sofi, I. A., & Ganaie, S. A. (2023). The status of open access repositories in technology: Insights from OpenDOAR. *Information Discovery and Delivery*, 52(2), 164-174. <https://doi.org/10.1108/IDD-11-2022-0119>
7. Wani, J. A., Sofi, T. A., Sofi, I. A., & Ganaie, S. A. (2024). The status of open access repositories in technology: Insights from OpenDOAR. *Information Discovery and Delivery*, 52(2), 164-174. <https://doi.org/10.1108/IDD-11-2022-0119>

#### **About Authors**

**Manash Esh**, Senior Information Scientist, University of North Bengal, West Bengal

Email: [manash@nbu.ac.in](mailto:manash@nbu.ac.in)

ORCID: <https://orcid.org/0000-0003-2748-8977>

**Dr. Saptarshi Ghosh**, Professor, University of North Bengal, West Bengal

Email: [sghosh@nbu.ac.in](mailto:sghosh@nbu.ac.in)

ORCID: <https://orcid.org/0000-0001-9567-5868>