
Norzelatun Rodhiah Hazmi\textsuperscript{1}, Abdullah Abrizah\textsuperscript{2}, and Yanti Idaya Aspura Mohd Khalid\textsuperscript{2}

\textsuperscript{1}Universiti Malaya
\textsuperscript{2}Affiliation not available

November 07, 2023

Abstract

Research data governance (RDG) plays a vital role in the effective management of research data within organizations. With the increasing volume and intricacy of research data, robust data sharing is necessary to facilitate advancements. However, despite its significance, RDG has not received the level of attention it deserves in comparison to other domains of research data management (RDM). Therefore, this desk study aims to fill the gap by identifying the key RDG activities implemented by top research performing organizations (RPOs) to provide insights into the key activities of governance that research organizations prioritize in their RDM policies. A content analysis of 36 policy documents was conducted, identifying 55 unique RDG activities. The findings showed that RPOs are more focused on implementing activities rather than defining or monitoring them, indicating an increasing awareness of the importance of RDG. The study identified two key activities that RPOs prioritize, highlighting a solid commitment to maintaining ethical and legal standards for RDM. As RDG practices evolve, it is crucial to identify emerging trends and technologies that impact RDG practices and explore innovative solutions to address challenges. Therefore, future research should explore innovative solutions for addressing challenges and developing more effective RDG practices.

Norzelatun Rodhiah Hazmi 1,2, A. Abrizah2,3, Yanti Idaya Aspura, M. K.2

Abstract:
Research data governance (RDG) plays a vital role in the effective management of research data within organizations. With the increasing volume and intricacy of research data, robust data sharing is necessary to facilitate advancements. However, despite its significance, RDG has not received the level of attention it deserves in comparison to other domains of research data management (RDM). Therefore, this desk study aims to fill the gap by identifying the key RDG activities implemented by top research performing organizations (RPOs) to provide insights into the key activities of governance that research organizations prioritize in their RDM policies. A content analysis of 36 policy documents was conducted, identifying 55 unique RDG activities. The findings showed that RPOs are more focused on implementing activities rather than defining or monitoring them, indicating an increasing awareness of the importance of RDG. The study identified two key activities that RPOs prioritize, highlighting a solid commitment to maintaining ethical and legal standards for RDM. As RDG practices evolve, it is crucial to identify emerging trends and technologies that impact RDG practices and explore innovative solutions to address challenges. Therefore, future research should explore innovative solutions for addressing challenges and developing more effective RDG practices.

Keywords: Research data governance; Research performing organizations; Research data management; Desk research; Content analysis

Introduction
The proper management of research data is not only essential but also paramount to ensure the responsible, ethical, and secure handling of this valuable asset, as supported by various studies (Alhassan et al., 2018; Benfeldt et al., 2020; DAMA International, 2017; Khatri & Brown, 2010; Wang et al., 2018). Researchers play an essential role in the effective management of research data because they engage in activities and practices that promote long-term research data preservation and use (Adika & Kwanya, 2020), resulting in scientific advancement and the discovery of new knowledge (Borghi, 2021; Tripathi et al., 2017). Moreover, it is crucial to emphasize the significance of research data governance (RDG). It allows research organizations to effectively handle and control research data by ensuring accurate information is accessible to the appropriate individuals (Lefebvre et al., 2018). Therefore, prioritizing RDG and research data management (RDM) becomes imperative for organizations to safeguard the integrity, quality, and transparency of the research data.

Norzelatun Rodhiah Hazmi
s2038059@siswa.um.edu.my
ORCID 0000-0001-5911-9194

A. Abrizah (corresponding author)
abrizah@um.edu.my
ORCID 0000-0002-8224-5268

Yanti Idaya Aspura Mohd Khalid
yanti@um.edu.my
ORCID 0000-0002-7642-469X

1 Dar al-Hikmah Library, International Islamic University Malaysia
Kuala Lumpur, MALAYSIA

2 Department of Library & Information Science, Faculty of Arts & Social Sciences,
Universiti Malaya, Kuala Lumpur, MALAYSIA

3 Institute for Advanced Studies,
Universiti Malaya, Kuala Lumpur, MALAYSIA
Effective RDG is a cost-effective solution and enhances collaboration and decision-making, saving valuable time and resources (Omar & Almaghthawi, 2020). Furthermore, good RDG can improve data quality, accuracy, and usability, leading to increased data trust and accessibility (Hendey et al., 2018; Jamii et al., 2020) while also ensuring data security and regulatory compliance (Lefebvre et al., 2018; Corti et al., 2019). In addition, a good RDG strategy is crucial to maximizing data value (Abraham et al., 2019; Brous & Janssen, 2020; Lis & Otto, 2020) while also minimizing data-related risks (Austin et al., 2021; Downs, 2021; Matthewson, 2019; Redkina, 2019; Troccoli, 2018). Consequently, organizations must prioritize RDG and management to guarantee research data's effectiveness, precision, and security. This, in turn, will contribute to improved decision-making processes and successful outcomes.

However, despite the crucial nature of data governance, several studies have uncovered a noticeable lack of knowledge and understanding in this domain (Abraham et al., 2019; Eke et al., 2022; Kuzio et al., 2022). Likewise, when compared to other RDM bodies of knowledge, there is a notable lack of attention paid to RDG in the library and information science (LIS) literature (Borda et al., 2020; Donaldson & Koepke, 2022; Tang & Hu, 2019; Wolski et al., 2017), indicating a scarcity of research on the topic (Al-Ruithe et al., 2019). This attention deficiency is particularly significant given the importance of RDG within LIS. Furthermore, previous literature has primarily focused on defining concepts rather than exploring their practical application (Alhassan et al., 2018).

To contribute to this knowledge gap, the authors conducted a desk study to identify the key organizational RDG activities based on the implementation and experience of selected research organizations. The concept of data governance activities, as outlined by Alhassan et al. (2018) refers to conditions or actions required to be classified in data governance. Within the framework of this study, RDG activities are the specific actions or tasks carried out to establish, maintain, and enforce effective research data governance within a research organization. The primary research question posed was as follows: What specific research data governance activities are implemented by research performing organizations? This desk study contends that a comprehensive approach to RDG is necessary to establish greater societal values, particularly in organizations that generate complex and diverse research data. Such entities, termed research performing organizations (RPOs) encompass universities and research institutions, and they are tasked with implementing efficacious strategies for governing research data. This ensures the responsible collection, management, sharing, and secure utilization of data, thereby facilitating the significant societal impact of research. Thus, the current study aims to explore RDG practices, with a particular focus on preventing data loss, improper disclosure (Solomonides, 2019), lack of trust among data users (Abduldayan et al., 2021), and data misuse (Alhassan et al., 2018; Al-Ruithe et al., 2018; Tiffin et al., 2019). The findings provide insights into the essential components of effective RDG and their potential benefits for promoting greater societal values. These insights can inform the development of evidence-based policies and guidelines for promoting responsible RDG in various organizational contexts.

Literature review

Research performing organizations

RPOs are institutions that primarily conduct research and experimental development in various disciplines (Cruz-Castro et al., 2020, p. 381). RPOs are universities, research institutes, non-profits, and government agencies with a research mandate. RPOs not only play a critical role in advancing knowledge and fostering innovation by providing opportunities for training and education to researchers and professionals (Leonelli, 2018), but they also contribute to institutional research innovation through their impact on science, public health improvement, and the economy (Vernon et al., 2021). Furthermore, RPOs frequently collaborate with industry and stakeholders to translate research findings into valuable societal applications. As a result, RPOs generate a massive amount of research data, which has grown exponentially over time (Lovakov et al., 2022). For example, it is estimated that by 2020, the total amount of data generated, recorded, duplicated, and consumed globally will exceed 64.2 zettabytes (Statista, 2021), with a projected increase to 180 zettabytes by 2025. (See, 2021). In 2015, the United States held the foremost position in research and development (R&D) expenditures, allocating more than $500 billion, which constituted 26 percent of the global total. China was close behind in second place, with an estimated $400 billion in spending (Tollefson, 2018). Figure 1 shows the significant expansion of research data repositories worldwide due to the significant growth in research data. The data for this expansion is sourced from http://www.re3data.org.
2.2 Research Data Governance

The growing volume and complexity of research data and the need for effective data sharing to promote scientific progress have underscored the increasing importance of RDG. However, the lack of standardized approaches to RDM across different fields and disciplines poses a significant challenge. Implementing effective data governance in research organizations faces barriers such as the need for a deeper understanding of professional data management complexity and the failure to adequately connect research data support with diverse projects (Lefebvre et al., 2018). To address these challenges, various frameworks and guidelines, including the widely accepted FAIR principles, have been developed to prioritize the findability, accessibility, interoperability, and reusability of data (European Commission, 2016; Wilkinson et al., 2016). Ensuring data security and privacy is also crucial, with the General Data Protection Regulation (GDPR) serving as key legislation governing data protection in the European Union (Peukert et al., 2022; Starkbaum & Felt, 2019).

On top of that, universities and research institutions have implemented data management plans (DMPs) to guide researchers in managing their data throughout the research process (Burgi et al., 2017; Nanyang Technological University, 2019). Ongoing engagement and coordination among researchers, institutions, and communities are vital for effective RDG, alongside the utilization of frameworks, guidelines, and policies that support responsible data management and usage. Moreover, it is imperative for journal publishers to offer guidance and clear instructions on data deposition for authors (Aleixandre-Benavent et al., 2019). The incorporation of specific data types shared through endorsed repositories exerts a significant influence on research data policies (Rousi & Laakso, 2020). These collective factors foster the development of RDG and contribute to the unprecedented growth of research data repositories globally.

2.3 A guiding view of the Data Governance Activities Model

The study of data governance has become increasingly important in recent years due to the growing volume and complexity of research data, as well as the need for effective data management and sharing practices. According to Alhassan et al. (2018), data governance activities are the necessary actions and requirements needed to be performed to ensure that data is effectively governed. In this study, the authors drew upon the data governance activities model put forth by Alhassan (2016), and integrated insights from the evolving data governance literature, including works by Abraham et al. (2019), Alhassan et al. (2016, 2018), and al-Ruithe et al. (2016), to provide a robust groundwork for the study. Additionally, the literature related to the emergence of data governance helps to conceptualize the areas and decision domains related to RDG. Decision domains are the areas of RDG that require control to attain organizational objectives (Sung et al., 2019, p. 6380). By incorporating these frameworks and models, the authors were able to provide a comprehensive analysis of data governance activities and their role in promoting effective RDG.
The current study relied on a research framework that grouped all data collection, analysis, and findings according to the data governance activities model proposed by Alhassan et al. (2016), which was also utilized in their later study (Alhassan et al., 2018). Since the objective of this study was to gain a deeper understanding of how RDG is executed in practice, the same approach used by Alhassan et al. (2018) was followed. The data governance activities model consists of three constructs: 'action,' 'area of data governance,' and 'decision domain of data governance.' The action construct of data governance is divided into three categories: 'define,' 'implement,' and 'monitor.' In reviewing the literature on data governance frameworks, the areas of data governance were identified and used as a preliminary guide to identify the RDG areas mentioned in the selected policy documents.

Furthermore, a combination of the frameworks proposed by Abraham et al. (2019) and Khatri and Brown (2010) were selected to present the decision domains that should be considered for RDG. Decision domains are crucial because the definition of data governance also pertains to who has decision rights and accountability regarding an enterprise's data assets. Therefore, identifying the decision domains is necessary to assign the appropriate responsibilities and duties (Alhassan et al., 2016).

Materials and Method

The current study utilized desk research to explore RDG in RPOs. Desk research involves summarizing and synthesizing previously published research (Goundar, 2012; Kabir, 2016; Woolley, 1992). This approach was often recommended as a starting point for research projects as it provides a benchmark against which primary data findings can be evaluated (Kabir, 2016).

Content analysis was deemed an appropriate method to analyze the collected data for the desk research. Content analysis is a commonly used technique for analyzing texts (Kleinheksel et al., 2020), mainly when the meaning of the text is straightforward and clear. Researchers systematically code the text to identify structures and patterns that reveal the meaning of the text (Vaismoradi et al., 2013). This approach offers researchers a flexible and pragmatic method for developing and extending knowledge (Hsieh & Shannon, 2005). The content analysis process allows for qualitative data analysis while also providing the ability to quantify the data (Grbich, 2012).

This study employs purposive sampling to select existing RDM policies as the sample. The sampling criteria for the policy documents required that they be produced by a RPO listed among the top 350 ranked institutions in the Scimago Institutions Rankings, filtered by Research Rank (Scopus, 2022), and must relate to RDG/management policy. The Scimago Institutions Rankings was chosen above other well-known rankings, such as the Leiden Ranking because it measures research activity at universities and research institutions worldwide. Meanwhile, the Leiden Ranking merely provides information on universities’ scientific performance based on bibliographic data from publications indexed in the Web of Science (CWTS, 2022), with no research institutions included.

Since there are 8084 ranked institutions listed, the scope was narrowed by limiting the list to the top 350 ranked institutions. The online tracking of the policy documents was conducted in May 2022 using the search terms ‘research data governance’ OR ‘data governance’ OR ‘research data management policy’ OR ‘data policy.’ The selected policy documents were required to be easily accessible online and downloadable, and written in English. Although the sample was not restricted to any specific publication date, all selected institutions must have a publicly available version of the document. Table 1 outlines the inclusion and exclusion criteria for sampling purposes.

<table>
<thead>
<tr>
<th>No</th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The document must be produced by a RPO.</td>
<td>The document is produced by a non-RPO.</td>
</tr>
<tr>
<td>2</td>
<td>The RPO must be one of the top 350 ranked institutions in Scimago Institutions Rankings by Research Rank.</td>
<td>The RPO is not listed in the top 350 ranked institutions in Scimago Institutions Rankings by Research Rank.</td>
</tr>
<tr>
<td>3</td>
<td>The document is either on RDG/RDM policy.</td>
<td>The document is not related to RDG/RDM policy.</td>
</tr>
<tr>
<td>4</td>
<td>The document must be easily accessible online and downloadable.</td>
<td>The document is not easily accessible online or downloadable.</td>
</tr>
<tr>
<td>5</td>
<td>The sample is restricted to documents in English.</td>
<td>The document is not in English.</td>
</tr>
<tr>
<td>6</td>
<td>The sample is not restricted to documents on any specific publication date.</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
After considering all the inclusion criteria, 36 documents were identified and included in the study. The sample included 34 research data management policies and 2 RDG frameworks from 34 universities and 2 research institutions. The purposive sampling method has allowed for selecting a sample that met the criteria needed for the research, and the Scimago Institutions Rankings provided a reliable source for identifying suitable RPOs.

Data collection procedure and data analysis

Following the methodology employed by Alhassan et al. (2018), this desk study employed eight coding steps to conduct a content analysis on selecting policy documents (Fig. 2). These sequential steps, which include data collection and coding procedures, promote clarity and transparency in the research processes.

Fig. 2 Coding steps for content analysis of selected policy documents

Step 1: Determine the level of analysis

Step 1 involved determining the level of analysis, which involved selecting the unit of language to be examined, such as a single word, a set of words, phrases, or an entire document (Carley, 1993). In this study, the level of analysis focused on words or phrases related to RDG activities. For content analysis, 36 RDG/management policy documents were selected from 34 higher educational institutions and 2 research institutions worldwide. The policy documents were identified by searching selected institutional websites.

Step 2: Identify concepts to code
To ensure a structured and rigorous approach, the authors employed directed content analysis techniques described by Hsieh and Shannon (2005), in which an initial coding scheme is developed based on an existing theoretical or conceptual model. Directed content analysis is generally more structured than inductive methods. The researchers begin the research by identifying key categories and themes as initial coding categories based on previous research (Potter & Levine-Donnerstein, 1999; Vreugdenhil et al., 2022). This deductive approach allowed for identifying critical categories based on previous research. The initial coding scheme for this study was guided by data governance frameworks from the literature, as shown in Fig. 2. Throughout the analytical process, new codes were also inductively developed while reading the texts, for instance, data citation and data custodianship. Over 100 concepts were identified and then themed as RDG activities. The transcripts were read multiple times to better understand the RDG activities. Then, other co-authors reviewed and discussed the codes and categories to ensure rigor.

Step 3: Code for a concept’s existence or occurrence

To determine the most appropriate coding strategy, researchers must consider whether to code concepts based on their mere existence or frequency after a sufficient number of concepts have emerged (Alhassan, 2018). For this study, the authors first coded for the existence of concepts, followed by coding for the frequency of themes. This approach enabled the authors to capture all potential RDG activities implemented by RPOs and gain a deeper understanding of the emerging concepts. The themes were then compared based on how frequently they appeared to identify priority themes.

Step 4: Differentiate between concepts

During this step, researchers should decide whether to code the concepts verbatim or in a modified or compressed version the concepts (Alhassan, 2018). The analysis of the documents was carried out, and excerpts were extracted. Upon observation, specific excerpts were found to possess an implicit congruence of meaning, thus consolidating them under a single concept. Following this, the identified concepts were sorted according to the three fundamental constructs of the data governance activities model: ‘action,’ ‘area,’ and ‘decision domain.’ Subsequently, the sorted concepts were organized into coherent clusters based on relevant themes.

Step 5: Create rules for coding texts

The following rules were established during the coding procedure in this research to maintain coding consistency (Alhassan, 2018):

1. Policy documents were initially read to identify RDG activities with imperative verbs indicating an action to be taken;
2. Emerged excerpts were compared to identify similarities and differences and grouped under relevant concepts; and
3. The concepts’ meaning and connection with the governance bodies were carefully examined in the context of the text to ensure accurate categorization.

Step 6: Remove “irrelevant” information

In the current study, a critical analysis stage involves determining an appropriate approach for dealing with the text that was not coded (Alhassan, 2018). All 36 policy documents underwent multiple thorough readings to identify recommended RDG activities comprehensively. A sentence must contain an imperative verb that indicates a specific action, such as "define," "set," "approve," or "provide" to qualify for coding. These actions signify the conditions or tasks that must be implemented to satisfy the requirements for RDG. Carley (1993) recommended discarding irrelevant information to streamline the content analysis procedures.

Step 7: Code the text

After establishing the appropriate approach for handling irrelevant information, researchers must follow the translation rules outlined in Step 5 to initiate the coding process (Alhassan, 2018). In this study, the authors utilized systematic coding procedures by utilizing Atlas.ti 9.1.3, a qualitative analysis software system. The coding scheme for RDG areas included data policies (e.g., standards, processes, procedures), performance measurement, contractual agreements, compliance monitoring, communication, issue, and risk management, coordination of decision-making (Abraham et al., 2019), data ownership (Brous et al., 2016), data security, data privacy, data integrity (al-Ruith et al., 2016), and data stewardship (Peng et al., 2018). Meanwhile, the interrelated decision domains are identified: data quality, data security, data architecture, data lifecycle, metadata, data storage and
infrastructure, data principles, and data access. The coded data were then entered into an MS Excel spreadsheet, facilitating further data analysis. From these codes, the researchers derived themes related to RDG activities, categorizing them into three key dimensions: Action, Area of Data Governance, and Decision Domain of Data Governance. These themes are elaborated upon in the findings section of this paper. Fig. 3 provides an illustrative representation of the research model for enhanced clarity.

![Fig. 3 A research model of the RDG Activity Dimensions](image)

**Step 8: Examine the outcomes**

After coding the data using the qualitative analysis software system Atlas.ti 9.1.3 and applying the translation rules identified in Step 5, the authors had to decide on the method of representing the RDG activities, and for this research, they opted to focus on the existence of the concepts, then later count for the frequency of the themes produced based on the existing concepts as the principal approach to review and present the results.

**Credibility and validity**

To ensure the credibility and validity of the study, the authors utilized several methods. Firstly, they aimed to collect relevant policy documents from diverse RPOs to gain a comprehensive understanding of the contexts that informed the study, as suggested by White & Marsh (2006, p. 36), that the sampling should be theoretical and purposive. Furthermore, they conducted a rigorous analysis of the documents by coding the texts. Once the initial coding was completed, the authors engaged in a collaborative review process, where they jointly reviewed the codes and discussed the interpretations to ensure the accuracy and consistency of the findings.

**Results**

**Descriptive findings of the RDG policy documents**

When analyzing the policy documents related to RDG, considering the study’s context and scope becomes crucial. The authors collected policy documents from 36 higher education institutions, including universities and research institutions from various countries across the globe as shown in Fig. 4. Notably, the United Kingdom had the
highest number of RDG-related policy documents, accounting for 47 percent (17) of the total sample. The United States is the second highest contributor, contributing 14% (5) of the sample. Then followed by Australia and Netherlands, each accounting for 11 percent (4) of the sample. Other countries, including Ireland (2), South Africa (2), Belgium (1), and Singapore (1), accounted for smaller proportions of the sample, with percentages ranging from 3 percent to 5.5 percent.

Fig. 4 Distribution of RDG policy documents across different countries in the sample (Photo from Google)

These countries have recognized the significance of research, as evidenced by their substantial investments in research and development (R&D) relative to their GDP expenditure (World Bank, 2022). Notably, Belgium allocated 3.16 percent of its GDP to R&D in 2019, while Australia invested 1.83 percent, the United Kingdom invested 1.71 percent, the United States invested 3.12 percent, and the Netherlands invested 2.18 percent. Furthermore, other countries such as Singapore, Ireland, and South Africa also marked notable investments in R&D as percentages of their GDPs.

Out of the 36 policy documents analyzed, 34 originated from universities, identified by codes U1-U34. The remaining 2 documents were from research institutions, and were designated as R1 and R2. The RPOs that contributed to the policy documents in the study are listed in Table 2. The institutions are recognized for their substantial publication presence in two prominent citation databases, namely Scopus and Web of Science (WoS).

Table 2 Research Performing Organizations contributing policy documents in this study

<table>
<thead>
<tr>
<th>Institution Code</th>
<th>Research Performing Organizations</th>
<th>Institution Code</th>
<th>Research Performing Organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>Macquarie University</td>
<td>U9</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>U2</td>
<td>University of Birmingham</td>
<td>U20</td>
<td>University of Surrey</td>
</tr>
<tr>
<td>U3</td>
<td>Boston University</td>
<td>U21</td>
<td>Delft University of Technology</td>
</tr>
<tr>
<td>U4</td>
<td>University of Cambridge</td>
<td>U22</td>
<td>University of Twente</td>
</tr>
<tr>
<td>U5</td>
<td>University of Cape Town</td>
<td>U23</td>
<td>University of Ulster</td>
</tr>
<tr>
<td>U6</td>
<td>University of Edinburgh</td>
<td>U24</td>
<td>University of New South Wales</td>
</tr>
<tr>
<td>U7</td>
<td>University of Exeter</td>
<td>U25</td>
<td>Vanderbilt University</td>
</tr>
<tr>
<td>U8</td>
<td>Ghent University</td>
<td>U26</td>
<td>University of Wollongong</td>
</tr>
<tr>
<td>U9</td>
<td>Heriot-Watt University</td>
<td>U27</td>
<td>University of Groningen</td>
</tr>
<tr>
<td>U10</td>
<td>Imperial College London</td>
<td>U28</td>
<td>University College Cork</td>
</tr>
<tr>
<td>U11</td>
<td>Leiden University</td>
<td>U29</td>
<td>University of Notre Dame</td>
</tr>
<tr>
<td>U12</td>
<td>University of Liverpool</td>
<td>U30</td>
<td>Florida State University</td>
</tr>
<tr>
<td>U13</td>
<td>London School of Hygiene &amp; Tropical Medicine</td>
<td>U31</td>
<td>University of Oxford</td>
</tr>
</tbody>
</table>
Fig. 5 illustrates the cumulative count of publications indexed in Scopus and WoS for the period spanning 2018 to 2022 across these institutions. The trend of publications indicates a greater propensity of these institutions towards publishing in WoS as opposed to Scopus. The University of Oxford has the highest number of publications, totaling 68,252 in Scopus and 86,492 in WoS, surpassing all other institutions. Notably, a study (Lancho-Barrantes & Cantu-Ortiz, 2021) emphasizes that the University of Oxford is recognized for its extensive collaborations in publication production, firming its position among the most collaborative universities. On the other hand, The Open University produced the least number of publications, which is 3,396 in Scopus and 5,768 in WoS.

These institutions are highly esteemed RPOs, as indicated by their Scimago rankings. They exhibit a commitment not only to publications but also to the proper management of research data. This commitment is evident through their well-defined policy documents about RDM, underscoring the importance of RDG within these organizations. Information on the RPOs’ Scimago Ranking and the implementation/revision year of their policy documents is presented in Table 3.
### Table 3. Scimago ranking and implementation/revision years of RPOs’ policy documents

<table>
<thead>
<tr>
<th>Item</th>
<th>Number (#/36)</th>
<th>Percentage (100%)</th>
<th>Institution Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scimago Ranking</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-50</td>
<td>5</td>
<td>14</td>
<td>U4, U10, U31, U33, R2</td>
</tr>
<tr>
<td>51-100</td>
<td>3</td>
<td>8</td>
<td>U6, U12, U34</td>
</tr>
<tr>
<td>101-150</td>
<td>8</td>
<td>22</td>
<td>U2, U3, U8, U11, U19, U21, U25, U27</td>
</tr>
<tr>
<td>151-200</td>
<td>1</td>
<td>2.8</td>
<td>U7</td>
</tr>
<tr>
<td>201-250</td>
<td>2</td>
<td>5.6</td>
<td>U13, R1</td>
</tr>
<tr>
<td>251-300</td>
<td>9</td>
<td>25</td>
<td>U1, U5, U18, U20, U22, U26, U29, U30, U32</td>
</tr>
<tr>
<td>301-350</td>
<td>6</td>
<td>17</td>
<td>U14, U15, U16, U17, U23, U28</td>
</tr>
<tr>
<td>351-400</td>
<td>1</td>
<td>2.8</td>
<td>U9</td>
</tr>
<tr>
<td>401-450</td>
<td>1</td>
<td>2.8</td>
<td>U24</td>
</tr>
<tr>
<td><strong>Implementation / Revision Year</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2023</td>
<td>1</td>
<td>3</td>
<td>R2</td>
</tr>
<tr>
<td>2022</td>
<td>5</td>
<td>14</td>
<td>U6, U10, U13, U24, U33</td>
</tr>
<tr>
<td>2021</td>
<td>5</td>
<td>14</td>
<td>U1, U4, U17, U20, U25</td>
</tr>
<tr>
<td>2020</td>
<td>5</td>
<td>14</td>
<td>U15, U16, U26, U29, U32</td>
</tr>
<tr>
<td>2019</td>
<td>4</td>
<td>11</td>
<td>U12, U14, U19, U34</td>
</tr>
<tr>
<td>2018</td>
<td>6</td>
<td>17</td>
<td>U2, U3, U5, U21, U22, U27</td>
</tr>
<tr>
<td>≤2017</td>
<td>8</td>
<td>22</td>
<td>U7, U8, U9, U11, U18, U28, U30, R1</td>
</tr>
<tr>
<td>No Date</td>
<td>2</td>
<td>5</td>
<td>U23, U31</td>
</tr>
</tbody>
</table>

### Research data governance activities

We identified nineteen (19) RDG areas based on our approach (Fig. 2), and subsequently, these RDG areas were categorized into eight RDG decision domains (Table 4), namely data principles, data architecture, data lifecycle, data storage and infrastructure, metadata, data quality, data security, and data access.

**Table 4** The pairing of research data governance areas and decision domains

<table>
<thead>
<tr>
<th>Decision Domain</th>
<th>Area of RDG</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Principles</strong></td>
<td>Data Policy</td>
<td>Data policy provides guidelines and rules for the use and management of research data, closely connected to data principle (Allela &amp; Mwai, 2019, p. 157)</td>
</tr>
<tr>
<td></td>
<td>Data Strategy</td>
<td>Organizations need focused and tangible data strategies that align with their goals to treat data as valuable assets (Alhassan et al., 2019, p. 108)</td>
</tr>
<tr>
<td></td>
<td>Compliance Monitoring</td>
<td>Compliance monitoring helps ensure that data principles are being followed and enforced (Goel et al., 2021, p. 6)</td>
</tr>
<tr>
<td></td>
<td>Data Ownership &amp; IPR</td>
<td>Defining data ownership and IPR is critical to ensuring that research data is managed and used appropriately (Boullenois, 2021, p. 3; Sung et al., 2019, p. 6380)</td>
</tr>
<tr>
<td></td>
<td>Performance Measurement</td>
<td>Performance measurement is necessary to assess the effectiveness of data governance based on strategic business goals (Fauzy et al., 2021, p. 650)</td>
</tr>
<tr>
<td></td>
<td>Decision-Making Coordination</td>
<td>Decision-making coordination is essential to ensuring that data governance decisions are made efficiently and effectively across departments (Ranathunga &amp; Wickramarachchi, 2021, p. 927)</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>Communication is necessary to ensure that everyone within an organization understands the data governance policies and procedures (Tan &amp; Lim, 2022, p. 2)</td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>Training is necessary to ensure that everyone within an organization is knowledgeable about data governance and how it</td>
</tr>
</tbody>
</table>
The 36 policy documents were analyzed to extract excerpts that contained imperative verbs related to RDG task areas. These excerpts were then subjected to explicit and intrinsic comparison to group them based on their meaning. Excerpts that conveyed similar meanings were grouped to represent a unique concept, while those with distinct meanings were classified into separate concepts. The coding procedure yielded 118 distinct concepts; i.e. unique task areas. However, an individual concept has the potential to be classified into multiple categories depending on the specific context and governance bodies. For instance, the task of an organization is to “ensure that research data are made available, wherever possible, for use by research community” can be categorized into four categories: (1) defining data strategy for data principle, (2) defining data licensing for data access, (3) monitoring data sharing for data access, and (4) monitoring data selection for data storage and infrastructure.

Using the data governance activities model as the guiding lens, these concepts were then classified into bigger categories, resulting in 226 categories. In maintaining the meaning of the categories, it was decided that each category would consist of the following three constructs: (1) action, (2) area of governance, and (3) the decision domain. Then each category was classified into bigger themes, considered a RDG activity. From 226 categories, it went down to 55 unique RDG activities. Table 5 shows the terms used in the coding procedure associated with the number of results counted after analyzing the 36 policy documents.
Table 5 Terms included in coding procedures

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>Coding example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excerpt (Document Code)</td>
<td>640</td>
<td>Ensure an efficient transmission of general data management information between the central level and the research community (U8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ensure that all relevant role-players in their departments are aware of, and adhere to, the University’s Research Data Management policy (U18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ensure that staff and students are aware of their responsibilities and obligations in effective management of Research Data and identify or promote training where gaps in these skills are identified (U19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ensuring that there is appropriate communication and organizational awareness of Research Data Management issues (U26)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Take appropriate steps to inform staff and PhD students about the policy and expected procedures (U27)</td>
</tr>
</tbody>
</table>

| Concept (Task Area)         | 118   | Ensures an efficient transmission of general research data management information between the central level and the research community         |

| Action                      | 3     | Monitor                                                                                                                                   |

| Area of governance          | 19    | Communication                                                                                                                            |

| Decision domain             | 8     | Data principle                                                                                                                           |

| Theme (Activity)            | 55    | Monitor communication for data principle                                                                                               |

Meanwhile, the study found that RDG activities under the ‘implement’ construct are the most frequently executed task among selected RPOs, comprising 119 activities. Followed by RDG activities under the ‘define’ construct with 60 activities. RDG activities under the ‘monitor’ construct are the least frequently executed task, with only 47 activities. Regarding individual activities, the findings indicate that RPOs focus on ‘Implement Compliance Monitoring for Data Principle’ and ‘Monitor Compliance Monitoring for Data Principle’ the most, with 17 and 12 activities, respectively. Followed closely by ‘Define Data Strategy for Data Principle,’ which had 15 activities. Details are outlined in Table 6.

Table 6 The code co-occurrence of RDG activities for RPOs by three main constructs

<table>
<thead>
<tr>
<th>RDG Activities</th>
<th>Define Total=60</th>
<th>Implement Total=119</th>
<th>Monitor Total=47</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication for Data Principle</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Compliance Monitoring for Data Principle</td>
<td>2</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Data Ownership &amp; IPR for Data Principle</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Data Policy for Data Principle</td>
<td>2</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Data Strategy for Data Principle</td>
<td>15</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Decision-making Coordination for Data Principle</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Performance Measurement for Data Principle</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Training for Data Principle</td>
<td>4</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Data Custodianship for Data Architecture</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Data Stewardship for Data Lifecycle</td>
<td>8</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>
Discussion

The present desk study employed a rigorous coding procedure to analyze the RDG activities of top leading RPOs. These organizations are from highly reputable RPOs, as ranked by Scimago. In line with this, a study conducted by Sarwar et al. (2021) confirms that North American and European universities generally achieve higher rankings compared to other universities.

The findings indicate that RPOs prioritize the execution of activities rather than their definition or monitoring. This aligns with the previous study (Alhassan et al., 2018), which revealed that practice-oriented organizations tend to have a greater emphasis on data governance activities related to implementation. This suggests that RPOs are increasingly recognizing the importance of RDG in maximizing the value of research data (Abraham et al., 2019; Kouper et al., 2020; Omar & Almaghthawi, 2020).

By analyzing individual RDG activities, the researchers gained a more detailed understanding of the specific actions that RPOs prioritize in their research data management practices. The findings indicate that RPOs strongly emphasize activities related to compliance monitoring for data principles, as evidenced by the high frequency of both 'Implement Compliance Monitoring for Data Principle' and 'Monitor Compliance Monitoring for Data Principle' activities. These findings reflect RPOs' commitment to maintaining ethical and legal standards for research data management (Kouper et al., 2020; Liu et al., 2020; Marlina et al., 2022).

Moreover, the results highlight the importance of developing a proactive approach toward research data management (Alhassan et al., 2019; Willaert et al., 2019), as evidenced by the significant number of activities related to 'Define Data Strategy for Data Principle.' By developing clear strategies for managing research data effectively, RPOs can better align their research data management practices with their research objectives and ensure compliance with regulatory and funding agency requirements.

Overall, the study provides valuable insights into the RDG activities of top leading research performing organizations and underscores the importance of establishing comprehensive research data management policies to comply with regulatory and funding agency requirements. The findings of this study may have implications for RPOs seeking to improve their research data management practices, as well as for policymakers seeking to develop guidelines and best practices for RDG.

Conclusion

In the present data-driven world, RDG is crucial in ensuring effective management and use of research data. This study sheds light on the key organizational RDG activities that promote equitable research data access, improve data quality, and mitigate data-related risks, which can help data practitioners and organizational stakeholders gain a deeper understanding of the RDG discourse and adopt effective RDG practices. Furthermore, effective RDG practices can promote compliance with regulatory requirements, safeguard vital data, and positively impact social change by generating more accurate and reliable research data to shape public policies. However, as technology and data evolve rapidly, it is imperative to develop best practices to keep up with these changes.

| Data Selection for Data Storage & Infrastructure | 2 | 1 | 1 |
| Data Repository for Data Storage & Infrastructure | 5 | 8 | 1 |
| Data Retention & Disposal for Data Storage & Infrastructure | 2 | 6 | 1 |
| Data Citation for Metadata | 1 | 5 | 0 |
| Data Integrity for Data Quality | 1 | 10 | 1 |
| Issue & Risk Management for Data Quality | 3 | 4 | 1 |
| Data Privacy for Data Security | 1 | 8 | 4 |
| Data Sharing for Data Access | 2 | 2 | 3 |
| Data Licensing for Data Access | 3 | 4 | 4 |
Therefore, future research in this field should focus on identifying emerging trends and technologies that may impact RDG practices, exploring innovative solutions for addressing the challenges that arise, and ultimately contributing to the development of more effective and efficient RDG practices that better serve the needs of researchers, institutions, and society as a whole.

It is important to note that this desk study has some limitations that should be discussed and must be considered when interpreting the results and their generalizability. Firstly, the study only relies on the collected policy documents, which may not provide a comprehensive view of how RDG is perceived in other RPOs. Additionally, the analysis of the policy documents may be influenced by the authors’ subjective perspectives and positions as researchers, which could impact the interpretations and conclusions drawn in this study. Finally, the specific documents examined in this study are limited to the organizations included in the sample. Thus, the findings may not be applicable beyond the scope of this research.

Funding This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability We authors declare the total availability of data from our research.

Declarations

Conflicts of interest The authors have no conflict of interest to disclose.

References


Imperial College. (2022). *Research Data Management Policy*. Imperial College. (imperial.ac.uk)


Scopus.


London School of Hygiene & Tropical Medicine. (2022). Research Data Management Policy. London School of Hygiene & Tropical Medicine.


