Providing Research Data Management Services and Practices at PDDI-LIPI: Preparedness, Roles, Challenges and Training

Hermin Triasih¹*, Rahmi², Katrin Setio Devi³
¹The Center for Scientific Data and Documentation, Indonesian Institute of Sciences
²,³Department of Library and Information Science, Universitas Indonesia

*Correspondence: hermin.triasih@lipi.go.id

Abstract

This study aims to analyse the implementation of RDM at PDDI-LIPI and to assess its staff’s understanding about RDM services. This article also discusses the challenges and obstacles PDDI faces in providing RDM services. The data was collected via an online survey from 28 July to 7 August 2020. The survey consisted of 35 questions and was shared with 36 respondents via social media. The results identified categories such as research data management services, data management planning services, data archiving services, funding, and staff competency and training needs. In addition, this article also discusses the approach and assessment of RDM services, challenges in providing RDM services, and plans for further developing RDM services at PDDI-LIPI. The results showed that the PDDI staff’s understanding of RDM services is adequate. As a new service, the implementation of RDM at PDDI-LIPI continues to develop toward optimisation. RIN is a platform used by PDDI to support this goal. The three biggest obstacles faced by PDDI-LIPI in developing RDM services are limited human resources, competence and budget. Various trainings related to RDM, both sending staff off campus and inviting trainers to campus, were carried out by PDDI to overcome these obstacles. It is recommended to conduct further research on the mapping and upskilling of staff in charge of RDM services.

ABSTRAK


Keywords: Research data management; Data services; Data training; Researchers; Librarian roles; LIPI

1. INTRODUCTION

The Center for Scientific Data and Documentation, Indonesian Institute of Sciences (PDDI-LIPI) recognises the emerging demand for providing a broader and higher-quality range of data services to meet patron needs at various points in the research process. Research data management (RDM) services and their framework have become a vital part of managing research data from the start to the completion of the research process (Whyte & Jonathan, 2011). RDM consists of a number of different activities and processes associated with the data lifecycle, involving the design and creation of data, storage, security, preservation, retrieval, sharing, and reuse, all taking into account technical capabilities, ethical considerations, legal issues and governance frameworks.
(Pinfield, Cox, & Smith 2014). Currently used as a basis for conference papers, journal articles, and books, if properly managed (documented, preserved, and made accessible), it could stand alone as a scientific product with the potential to influence future research (Williford & Henry, 2012).

Based on Regulation of the Head of LIPI No. 12 of 2016 concerning repositories and depositories, PDDI-LIPI was in charge of managing research data through the National Scientific Repository and Depository (Repositori Ilmiah Nasional/RIN). RIN is a system developed by PDDI LIPI to store scientific papers and primary data from research results that are useful as a data source for research reproduction and data reuse. Management of research data includes activities to plan, create, store, organise, regulate access rights, share, explain, publish, and curate data (Sulaiman, 2019; reference, n.d.). The RIN process allows a new role for the librarian in the digital library environment, as noted by Tammaro, et al (2017).

Changing roles from conventional classifying and cataloguing tasks to new roles in RDM certainly requires thorough adaptation and preparation, from policies, infrastructure, and types of services, to the readiness of human resources competencies. To carry out the process properly, the librarian must first be able to understand the RDM needs of each researcher. Various fields of expertise at LIPI signify the emergence of different requirements for RDM according to these fields. To develop data management infrastructure and services, RDM services and a framework that suit the researchers’ needs must be identified at an early stage.

This study aims to describe the understanding of RDM concepts, services, and framework of 36 PDDI-LIPI staff members. It also seeks to explain the following: RDM services, online data management plan services, data archiving services, staff skills and training needs, funding, RDM service outreach and assessment, RDM services challenges, future plans for RDM services (Lake et al., 2013). Although this research was conducted during a short period, with a plan to collect data in 11 days, the findings will be a starting point for forming and facilitating RDM implementation in Indonesia.

2. LITERATURE REVIEW

Currently, research data has become a concern for the scientific community. Not only the final publication, research data also needs to be managed, in addition to preservation purposes, research data management also aims to make the data reusable and to produce new data to improve the quality of future research. According to Swan & Brown (2008), scientific data is an institutional asset that needs to be archived, preserved and disseminated to the public. It is for this purpose that RDM is needed.

RDM is data management in the context of research activities. RDM services refer to the storage, access and preservation of data produced in particular investigations or research projects. These services support the full data lifecycle including data management planning, digital curation and metadata creation and conversion (Tenopir et al., 2012). According to Fearon, Gunia, & Pralle (2013), RDM is an integral part of the role libraries play in supporting their research community. The skills supplied by librarians are very similar to those required by RDM, and so libraries and librarians have the potential to advance the cause of RDM at their universities in the future. Librarians also occupy a unique niche within universities that enable them to serve a broad clientele without undue regard to the interests of finance or compliance. A further aim of the RDM services is to ensure research integrity and enable the use of existing data for future research endeavours.

Pinfield, Cox, & Smith (2014) assert “RDM is a complex issue involving multiple activities carried out by various actors addressing a range of drivers and influenced by a large set of factors”. These actors are researchers, librarians, administrators, ethics advisors and IT professionals. All of them have a vital contribution to make in ensuring that research data and related information is available, visible, understandable and usable over the mid- to long-term (Davidson et al, 2014).
Various libraries have made RDM as one of their main services, especially in academic libraries or research libraries. RDM continues to be developed so that it can provide benefits to support research progress.

Prior studies have investigated that in the development of RDM services, there are many things that need to be considered by libraries, including the development of policies, governance, resources, and skills (Pinfield, Cox, & Smith, 2014). This is also in line with the research conducted by Marlina & Purwandari (2019). The research showed the strategy that can be carried out to develop RDM services is to provide national policies and IT/IS infrastructure, as well as increasing the awareness of researchers about research data by conducting workshops or trainings. Another challenge in providing RDM services was also reported in a study conducted by Cox et al. (2017), those challenges are resourcing, working with other support services, and achieving “buy in” from researchers and senior managers.

3. METHOD

The data were collected via a questionnaire using the surveymonkey.com platform from 28th July to 7th August 2020. Thirty-six respondents who work at PDDI-LIPI filled in this questionnaire. The time estimate for filling out the questionnaire was about 15 minutes in Bahasa Indonesia. Respondents consisted of librarians, Information and Technology (IT) staff and PDDI-LIPI staff, especially those responsible for RDM services. The questionnaire was distributed through social media such as WhatsApp, Instagram, Facebook, and email. Moreover, we could not conduct direct surveys to get a large number of respondents because of the COVID-19 pandemic. Any questionnaires not returned by 7th July 2020 were deemed nonresponses. The online questionnaire was adapted from Lake et al. (2013) concerning the evaluation of research data management services. The questionnaire was divided into eight categories, in which several questions focused on the following:

- RDM services (Q7-Q11);
- online data management plan (DMP) services (Q12-Q17);
- data archiving services (Q18-Q25);
- staff skills and training needs (Q26-28);
- funding (Q29);
- RDM service outreach and assessment (Q30-Q31);
- RDM services challenges (Q32);
- future plans for RDM services (Q33-Q35).

The results of this study were evaluated with descriptive analysis and compared with previous studies similar to this one.

4. RESULTS AND DISCUSSIONS

Thirty-six of 55 questionnaires were returned, making a 65% completion rate. The gender distribution of respondents was 17 females and 19 males, and the ages of respondents were split into groups of less than or equal to 30 years old \((n = 8)\), older than 30 \((n = 27)\), and one respondent chose not to answer. The education background of the respondents ranged from undergraduate degrees \((n = 19)\), master degrees \((n = 14)\), and diploma degrees \((n = 3)\). The respondents’ affiliation was divided into two areas: 26 respondents worked in the central area, and ten respondents worked in a regional area.

4.1 Research Data Management Services

In the current era of open science, RDM is an important challenge for research institutions or providers. Although RDM is not a clearly defined concept, its scope consists of activities related to
data creation, organisation, arrangement and naming for backup, storage, conservation and sharing, and for all measures that ensure data security (Lake et al., 2013). RDM aims to ensure reliable verification of results and provide opportunities for new and innovative research to be built on existing information (Whyte & Tedds, 2011). Pinfield et al. (2014) also affirmed that RDM consists of a number of activities and processes related to the data lifecycle, which involve data design and creation, storage, security, preservation, retrieval, sharing, and reuse, taking into account technical capabilities, ethical considerations, legal issues and governance frameworks.

The success of RDM is inseparable from the roles of various parties, from researchers, research users, funders, archivists, and librarians. From year to year, the number of libraries with policies to provide RDM services has increased. From 2013 to 2016, the number of academic libraries with an RDM policy in Australia nearly doubled from 29% to 56%, while in the UK over the same time period it more than doubled from 17% to 42% (Cox et al. 2017). These figures show that libraries have an important role in the development of RDM.

In Indonesia, there are still few libraries involved in research activities. PDDI-LIPI is a work unit at LIPI that initiated the development of RDM services. The success of RDM services cannot be separated from the role of librarians and all staff who are directly involved in RDM activities at LIPI. Based on the survey results, RDM is considered very important to 26 respondents (72%) and important to 10 respondents (27%). This shows that PDDI staff is aware of the importance of this service being built at LIPI for data management and research support. Thirty-three respondents (about 92%) said that PDDI-LIPI has a data retention policy or formal research data management policy that may be associated with legal/regulatory compliance, intellectual property, technology transfer, or research administration policies (Q8). Twenty-nine respondents stated that PDDI-LIPI has an RDM policy, with various answers regarding when PDDI-LIPI initiated RDM services, such as in 2012 (n = 1), in 2015 (n = 2), in 2016 (n = 5), in 2018 (n = 1), and will be initiated in 2021 (n = 1) (Q9-Q10). This shows that the librarians’ knowledge of RDM policies at LIPI is fair. They know that PDDI has made a policy related to the services they are currently building. The results of research by Marlina and Purwandari (2019) regarding the strategy for RDM in Indonesia indicate that information technology–or information systems–related policies and infrastructure are the main factors of success. However, 26 PDDI staff did not know for sure when the initiation of RDM at PDDI-LIPI took place.

Based on the survey results, there are several reasons for developing RDM services at PDDI-LIPI, including expanding support for research progress (94%); parent agency policies (56%); institutional administrative initiatives to support research (53%); and requests from researchers to assist in data management, data sharing, and data preservation (36%). Flores et al. (2015) stated that libraries are starting to incorporate RDM into the research services offered as a result of demands from government agencies or university administration, the perceived need to remain relevant in the changing world of digital research, or because of the needs of the researcher.

### 4.2 Online Data Management Plan Services

Jones (2011) shows that a data management plan (DMP) plays a very important role and provides several benefits, including research data that can be easily found when needed; continuity if project staff leaves or new researchers join, avoiding unnecessary duplication of the collection for example, and reworking the same data; enhancing collaboration and advancing research with a variety of data; making research more visible and having a greater impact; and even the credit that can be obtained from the citation. The survey results in this study related to DMP services show that 23 respondents (64%) said that PDDI-LIPI offers training classes, sessions, or workshops on DMP preparation (Q12). This indicates that PDDI considers DMP services important from the start, as an institution that is currently developing new RDM services. This is also supported by the
results of research conducted by Matusiak et al. (2017), who argued that in the early stages of developing RDM, academic libraries usually focus on needs assessment, socialisation, training, and open-access advocacy. They mainly provide consultative services on DMP development, metadata, and data curation practices.

Based on the survey results, PDDI-LIPI used several RDM service consultation methods including via email/chat/telephone (86%) and visiting researchers at research offices or other locations (81%), researchers visiting PDDI-LIPI (53%), and other consultation methods such as technical guidance or webinars on data management, socialisation introducing RDM and offering cooperation (Q13). This shows that PDDI-LIPI wants to provide maximum DMP services in ways that researchers/users can reach. Researchers can flexibly choose DMP services according to their needs, considering that researchers spend most of their time in the field. Less than 10% of the researcher’s time is used for administrative purposes (Devi et al., 2020). Having a variety of DMP consulting service options can encourage more researchers to perform DMP before they start research. DMP consulting services have a positive impact on the use of future research results, considering the many costs incurred in a study.

Twenty-one respondents stated that PDDI-LIPI offers direct assistance or consultation, either in person or by email/chat, on DMP preparation for grant proposals, and data management planning support, whereas others said no (n = 5) and did not know (n = 10) (Q14). This shows that the PDDI staff’s understanding of the DMP services provided is adequate. More than 50% of staff are aware of the DMP service methods provided by PDDI-LIPI. DMP is necessary to research, and funding institutions frequently request a data plan as part of a grant proposal. With this DMP service, LIPI researchers can undoubtedly be significantly helped. In addition, PDDI-LIPI can also contribute to supporting the progress of research with proper research management from the start.

The challenges PDDI-LIPI has encountered in providing RDM services include limited staffing (n = 29), limited staffing competence (n = 27), funding (n = 21), technology infrastructure (n = 19), policy thrust (n = 15), marketing services (n = 12), parent institution engagement (n = 10), scoping services (n = 9), and staff training (n = 8) (Q15). In an open-ended question analysed by word counts, participants wrote that PDDI-LIPI’s strategies to face its challenges and obstacles were “increase the competence of staff” (5), “data” (5), “training” (4), “budget” (4), “development” (4), “activities” (4), “cooperation” (3), “internal” (3), “national scientific repository” (RIN) (3), “data management” (3), “outreach” (3), “researchers” (3), and “do not know” (3) (Q16).

RDM resources, consulting, or training services in PDDI-LIPI provided data storage and backup planning (n = 32), data sharing & access (n = 29), data management best practices via website resources & links to relevant literature (n = 27), helping researchers identify appropriate metadata standards (n = 23), data management best practices via workshops and direct training (n = 21), and data citation (n = 20) (Q17). The three highest barriers to the new RDM services, based on the survey results, were limited HR, HR competence and budget. This is in line with the results of Tang and Hu’s (2019) research, which states that one of the obstacles in providing RDM services is the problem of HR competence. This research result is also supported by Newton et al. (2011), who revealed that currently, libraries are involved in managing research data, and the development of RDM services is a strategic priority. In developing RDM services, they are faced with three major challenges, namely those related to skills gaps, resources and cultural changes.

The change in the roles of the librarian from traditional (e.g., cataloguing, managing printed collections, and other roles) to a role in the digital era is inevitable to provide services related to research data and demands a change in the skills of librarians. Data curation is one of the newest roles open to librarians. As stated by Tammaro et al. (2017), data curation has emerged as an innovative area of responsibility for researchers, librarians, and information professionals in digital library environments. Realising that HR competency is one of the main barriers it faces, PDDI-
LIPI is trying to overcome it by providing various trainings for all staff, especially those whose work is directly related to RDM service activities. This is a sign of PDDI’s seriousness in carrying out the mandate of the parent institution to manage research data optimally.

4.3 Data Archiving Services

Thirty-five respondents stated that PDDI-LIPI archives research data (Q18). PDDI-LIPI provides some assistance to researchers for archiving research data, such as direct assistance with depositing data at a domain repository \( n = 31 \), direct assistance with depositing data at the library’s or institution’s data archive \( n = 19 \), assistance locating solutions at existing data repositories for particular research domains (e.g., directing sociology students to the ICPSR repository) \( n = 16 \), assistance depositing data with a journal publisher \( n = 3 \), and other data archiving assistance \( n = 1 \) such as journals and books (Q19). Dataverse \( n = 36 \) is PDDI-LIPI’s principal archiving solution (Q20).

The current or anticipated sources for the data in PDDI-LIPI’s data archive include data directly associated with a particular publication \( n = 33 \); data related to a full research project (e.g., all processed data for a grant) \( n = 26 \); data from graduate research, dissertations or theses \( n = 13 \); data moved from another archive to the library data archive \( n = 12 \); and other data sources \( n = 1 \) such as journal articles (Q21). As the use of data-specific archives becomes widespread, there may be shifts in data sources that the institutional repositories cannot accommodate. The library deposits data collections for the researcher \( n = 20 \), the library assists when researchers deposit data \( n = 17 \), and participates in management, e.g., each work unit should have an administrator \( n = 1 \) (Q23). Especially for data archives, the type of access may be a technical issue, not just a policy issue.

The total overall size of research data sets currently deposited to the data archive is in terabytes \( n = 24 \) and gigabytes \( n = 12 \) (Q22). These are early days for both data-specific archives and institutional repositories, and possibly also for researcher’s awareness and adoption of these archiving options (Lake et al., 2013). Another feature generally considered important for data archives is support for persistent identifiers so that data sets can be found long-term and cited reliably in publications (including, in some cases, citing updated versions of certain collections with new data). PDDI-LIPI’s data archiving solution provides a persistent identifier for deposited datasets, such as Handle System (handle.net) \( n = 20 \), DOI \( n = 14 \), ARKs \( n = 1 \), and other identifiers \( n = 1 \) (Q24). Finally, the survey asked about the primary purpose of the data archive for controlled access \( n = 19 \) and open access \( n = 14 \) (e.g., unrestricted access for online users). Other purposes include appropriate storage of data archives, supporting researchers in storing their research data so that data loss can be minimised, making data sharing easier with complete control of the data owner, and data preservation (Q25).

4.4 Staffing: Skills and Training Needs

The depth and range of RDM services that libraries offer are, of course, directly proportional to staffing, both in the number of positions and the amount of time allotted for RDM activities when a position has other responsibilities (Lake et al., 2013). The training or experience categories listed below that are the most important to the RDM services staff at PDDI-LIPI are digital/data curation training \( n = 33 \), subject domain expertise \( n = 21 \), IT technology or services experience \( n = 25 \), library MLS/MLIS training \( n = 7 \), traditional archives training \( n = 7 \), and other training or experience \( n = 1 \) e.g., communication (Q26). Respondents indicated that PDDI-LIPI RDM staff most need additional training (see Fig. 1). Respondents also stated the methods PDDI-LIPI had used to facilitate further training for the RDM services staff, such as workshop attendance \( n = 29 \), training provided by professional organisations \( n = 27 \), independent study \( n = 24 \), conference
attendance \( (n = 23) \), local courses in a computer or digital technology \( (n = 19) \), training provided by vendors \( (n = 12) \), hiring consultants \( (n = 6) \), and other training methods \( (n = 1) \) made up PDDI-LIPI’s technological guidance (Q28).

4.5 Funding

Respondents identified RDM service funding as a key challenge. PDDI-LIPI’s current RDM services are funded by regular internal budget \( (n = 33) \), external grant funding \( (n = 1) \), and other sources of funding \( (n = 2) \), such as from the State Revenue and Expenditure Budget (APBN) (Q29). A significant budget is needed for this purpose, with a special budget allocation within PDDI-LIPI for RDM purposes, which shows PDDI’s commitment to providing these services for the advancement of research in the scope of LIPI regions in Indonesia. As a unit with a research institute library, RDM services are a top priority at PDDI-LIPI.

4.6 RDM Service Outreach and Assessment

As explained in the previous sub-section (see 4.1), RDM services at PDDI-LIPI are based on supporting research progress and meeting the needs of researchers. However, to be able to be appropriately utilised by researchers, the data repositories at PDDI must be dependable, and building researchers’ trust is not easy. PDDI uses various methods to approach this issue. Respondents indicated the most effective outreach methods that are used to encourage researchers to use RDM services: workshops or presentations to researchers or other institutions \( (n = 32) \), binding institutional policies \( (n = 29) \), library staff referrals/promotion \( (n = 26) \), website links cross-posted to other library site pages \( (n = 26) \), direct emails to researchers \( (n = 11) \), and other outreach methods \( (n = 4) \) such as incentives for researchers who share data or approaching to funding sources and journal managers to use RIN as a compulsory medium for storing research data and socialising the importance of RDM for researchers (Q30).

4.7 RDM Services Challenges

Respondents described the challenges PDDI-LIPI has encountered in providing RDM services and the measures that have been taken to overcome them (see Fig. 2). This is in line with the results of research conducted by Marlina (2019) through her literature study entitled “Strategy for Research Data Management Services in Indonesia,” in which competence figures as one of the important assets in RDM services. Moreover, research conducted by Tang & Hu (2019) asserts that to develop RDM services, institutional commitment to resources and training opportunities is very important. As an emerging profession, data librarians need to be further nurtured, mentored and trained.
4.8 Future Plans for RDM Services

Respondents listed the RDM services that PDDI-LIPI plans to offer but has not yet implemented, and when they are anticipated to begin: directly assisting researchers by depositing data in the repository \((n = 29)\); institutional repositories were used to archive research data (and used for publication) \((n = 29)\); PDDI-LIPI will host or be directly involved with archiving research data in external archives or service repositories, such as the National Scientific Repository and DataVerse \((n = 28)\); PDDI-LIPI operates a research data archive dedicated to data storage and access \((n = 21)\); helping researchers find external data storage locations \((n = 15)\), and PDDI-LIPI is involved with data archiving mainly operated by other departments or groups in the parent institution \((n = 14)\) (Q33).

In an open-ended question analysed by word counts, respondents stated that the educational background/experience needed to become RDM staff included information \((9)\), university graduates \((7)\), data \((6)\), management \((5)\), data science \((4)\), computers \((4)\), information systems \((4)\), library science \((7)\), education \((4)\), and experience \((4)\) (Q34). Respondents also freely described the role they saw PDDI-LIPI playing in supporting research data management, now and in the future, as data \((25)\), researcher \((9)\), research \((8)\), policy \((5)\), RDM \((5)\), manager \((3)\), repository system \((3)\) (Q35). RDM services, as a reasonably new service at PDDI-LIPI, continue to develop and improve through efforts to expand the types of services so that RDM can be carried out appropriately by researchers at LIPI and others. PDDI-LIPI’s implementation of RDM could serve as the basis for national policymaking as well as the progress of future research.

5. CONCLUSION

This study showed that PDDI staff understanding that RDM is an important service from LIPI that can be used for data management to support research in Indonesia. RDM services can also be used as a medium for research collaboration with various data so that research can become more widely available. The implementation of RDM services that have been carried out at LIPI is to assist researchers in making DMP. Apart from the development of RDM services such as access control, data retrieval, and data preservation, LIPI also provides direct or indirect assistance or consultation to researchers for their research data archiving activities. These various approaches show that PDDI has prepared a scheme so that research data storage services can be utilised optimally. The final results of this study found that there were three major obstacles faced by PDDI LIPI in the implementation and development of RDM services: limited human resources, staff competence and budget. The limitations of this study include the small number of samples and the short survey time (28 July to 7 August 2020). Neither has this study been equipped with a literature review. Systematic reviews can help researchers to study the successes and failures of previous
research. A literature review can also assist researchers in determining new directions to produce better research in the future. This survey-based research did not use a Likert scale because of the limited time of the study. One of the urgent factors is building the trust of researchers so they submit their research data to LIPI to manage. To encourage them, LIPI held several related activities, including presentations to researchers or related institutions, and providing incentives to researchers. The implementation of RDM at LIPI has several challenges, including the number of human resources, competence and budget. In terms of competence, RDM is new in Indonesia. Institutional commitment to training resources and opportunities is essential. As an emerging profession, data librarians need to be continuously nurtured, mentored, and trained. LIPI has conducted staff training to deepen RDM knowledge. The RDM service at PDDI-LIPI is new and continues to develop and improve through efforts to expand the types of services so that LIPI researchers and others may carry out RDM. RIN is a platform created by PDDI as a solution that can be used adequately by researchers and other potential users in managing the research data they produce. If this RDM service has been carried out optimally, then policy-making can be based on research results. Additionally, research in Indonesia will also develop. With regards to the competence of data librarians, further research is needed on the mapping and upskilling of staff assigned to RDM services.

6. ACKNOWLEDGMENT
The authors thank the anonymous reviewers for their constructive comments to improve the article. This work was supported in part by the PUTI Q1 contract NKB-1450/UN2.RST/HKP.05.00/2020, and by Kappa Sigma Kappa Indonesia (KSKI). Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsors.

REFERENCES


