LIBRARIANS VS. DATA CURATORS IN MANAGING RESEARCH DATA: COMPLEMENT OR REPLACE?

[1]SLAMET RIYANTO, [2]HENDRO SUBAGYO, [3]HERMIN TRIASIH, [4]DWIATRI KUSUMANINGRUM, [5]YANIASIH

[1,5]Researcher, Center for Scientific Documentation and Information, Indonesian Institute of Sciences, Indonesia
[2]Head of office, Center for Scientific Documentation and Information, Indonesian Institute of Sciences, Indonesia
[3,4]Librarian, Center for Scientific Documentation and Information, Indonesian Institute of Sciences, Indonesia
Email: [1]slamet.riyanto@lipi.go.id

Abstract - Research data must be professionally managed. Research institutions increasingly need data management but do not have a data curator profession. This paper aims to identify the stages of data management and professional roles needed. The librarian profession is compared to the data curator to see the librarian's likelihood of carrying out the role of curating data. The study used literature, observation and interview study methods. The results of the study showed that data management stages must be adjusted to the research and cultural cycles in an institution. Stages could be categorized into eight activities, namely planning, collecting, ensuring, explaining, preserving, finding, integrating and analyzing data. Analysis of the role that librarians have done compared to the need for curatorial data results in the conclusion that librarians cannot directly replace the role of data curators. Librarians must be upgraded to be able to carry out tasks at each stage of data processing. This competence is mainly related to data processing activities using advanced information technology.

Indexterms - Data curator, Data research management, Librarian, Research life cycle.

I. INTRODUCTION

The digital library is growing very rapidly. Library management and the role of librarians must change according to these developments. In terms of collections, all types of libraries have provided collections in digital form. Provision and exchange of collection data such as books, proceedings and international journals are very easy to do through the provision of this digital collection. A literature search has also been carried out independently, without the help of a librarian.

The work of librarians such as the creation of abstracts has been replaced by technology through language processing that can extract information for clustering needs, document similarities, topic analysis, and automatically create links. Especially in academic libraries, in addition to technological changes, libraries are also required to manage research data. Scientific data is an asset of academic and research institutions that need to be archived, preserved and disseminated to the public.

Aside from being evidence of the results of research activities, storing data allows data to be reused and generate new data. Through good management of research data, it is expected that the quality of research will improve and its management will be more efficient. The new role of librarians in managing digital libraries and research data raises new types of work related to digital curation and data curation. The term "digital curation" was first used in 2001 as the title of a seminar on digital archives, libraries and e-Science where various communities gather to discuss pressing challenges to improve long-term management and preserve access to digital information in libraries[1]. The data curation has emerged as a new area of responsibility for researchers, librarians, and information professionals in the digital library environment[2]. Data curation allows discovery and retrieval of data, maintain data quality, adds value, and provides data for reuse over time through activities including authentication, archiving, management, preservation, and representation[3].

Are there different tasks and competency needs between librarians and data curators? This paper describes the comparison of the work of data curators and librarians in managing research data in a repository system. The purpose of the study was to describe the type of work in the research data management; identify the role of librarian and curator of data in each stage of research data management; and recommendations regarding additional competencies needed by librarians to act as data curators.

II. METHODS

The methods used in this study are a literature review, observation and interviews. The literature studied is about repositories, data curators and research data management. Observations were made on the librarian activity process at the Center for Scientific Documentation and Information (PDII) - Indonesian Institute of Sciences (LIPI).
The interviews were conducted with librarians and data managers at PDII and in an international research centre in Indonesia. The interview topic is related to their knowledge and readiness to manage research data.

III. RESULT AND DISCUSSION

a) The current condition of librarians and data curators in research institutions in Indonesia

The librarian profession was facing major challenges in the current era of digital technology. Hoon [4]estimates that eight jobs would be lost in 2030 including librarians, mailers, cashiers, receptionists, telemarketers, travel agents, word processors, and social media managers. In fact, in the past, until the digital era emerged, a librarian became a servant and educator who was highly trusted to manage scientific work which is a source of public knowledge. So far, the public need librarians and the library offer tremendous value to the communities they serve.

The role of librarians has evolved following the demands and development of information technology. Making book and listing labels,for example, is rarely done in some libraries. With interoperability technology, it is possible to exchange data between library collections. In addition to enriching collections, this exchange can also reduce errors when entering listings. Data obtained from other agencies must have gone through a verification and validation process. Bracke[5]stated that the role of librarians would develop from existing roles. Roles will continue to change,and new roles emerge without any sign of certainty in the future. Data curators may be one of the roles assumed to be future librarians.

In Indonesia, the role of librarians is still focused on managing and servicing publications, even though the management of research data is also important and urgent. PDII as a kind of library function unit in the largest research institution also still manages publications and only recently has it been moving to the processing of research data. In the future, the role of research libraries is needed in the management of research data, including LIPI as an institution tasked with conducting research and development of Science and Technology, fostering the development of science and technology, providing science and technology services, providing means to the government on science and technology national policy (Presidential Decree No. 1/1986). PDII will evolve into a research library with services focusing on data management and analysis to support national research policies (Head of LIPI Regulation No. 12 of 2016).

PDII develops data repositories to carry out scientific data management. Repositories (also referred to as data repositories or digital data repositories) are interfacing entities that can be searched and queryable that are able to store, manage, maintain, and curate digital data / objects[3]. In Indonesia, the repository system is limited to managing scientific works in the form of proceedings, scientific journals, theses, dissertations, and books, while the management of research data has not been carried out.

In general, the management of research data includes the activities of planning, creating, storing, organizing, regulating access rights, dividing, explaining, publishing and curating data [6]. Of all these processes, data curation activities are a major concern in the management of research data. The data curation process involves managing data throughout its life cycle, starting when planning activities where data will be created, stored until data becomes obsolete and deleted. The main purpose of data curation is to ensure that data is validated and verified so that the data becomes useful, informative, and quality. The quality research data can be reused for research purposes or generate new data in the future. Lord and Macdonald[7] defined data curation as an activity of managing and promoting the use of data from the point of creation, to ensure it is consistent with contemporary goals, and available for discovery and reuse. For dynamic datasets, this can mean continuous enrichment or renewal to keep it in line with its objectives. A higher level of curation will also maintain links with annotations and other published material.

b) Roles and competencies for research data management: librarians vs. data curators

Data curator profession has not developed as a functional position in Indonesia. Functions that have the ability to approach the needs of data curators are librarians. Based on the technical instructions of the librarian's functional position and credit figures, librarian activities are to manage both printed and digital library materials including collections in the form of video, audio, and images. Management of library collections in digital form in addition to text documents has not been much regulated in the regulation. Whereas in the future, collections in forms other than text will increasingly dominate. Management of digital documents including research data needs to be regulated so that action items can be claimed in the librarian functional position credit numbers.

In general, the management of research data in the repository can be done by a librarian, but the data curation process requires special competence. There are several jobs that can be handled by librarians in data curation. This is because these jobs are almost the same as the work done by librarians. However, there are several points of work in the data curation process that is impossible for librarians because the work requires special skills and has not been accommodated in the action items in the librarian's functional, technical instructions.

Management of digital documents such as scientific data requires special competence. Although the

management of digital documents is basically the same as managing printed documents, managing research data is not easy. Research data is specific to each field of science. Multidisciplinary data management requires competence in their fields. For example, species data needs to refer to the standard metadata of the biological field, Darwin Core. There are several fields that need to be filled in so that the data becomes quality and valuable. Scientific data can be in the form of text, numbers, images, audio, and video, therefore human resources are needed that have the ability to represent data so that in managing the data repository data curation processes are needed.

The following is an analysis of the work done by a data curator in managing research data on the repository system and the opportunity of a librarian to do the work. Curator datajob descriptions are modified from the Digital Curation Center (http://dcc.ac.uk) and Data Observation Network for Earth (https://www.dataone.org/).

1) Planning stage: explain the data to be collected and how the data will be managed and accessed throughout its life cycle.

No	Job description	L	DC
1	Create and document data backup policies	√	√
2	Create, manage, and document data storage systems	√	√
3	Determine the expected output and data type	-	√
4	Determine the role and give responsibility for data management	-	√
5	Determine the data model	-	√
6	Identify data sensitivity	√	√
7	Identify the appropriate repository for data	√	√
8	Incentives, challenges and obstacles: investigate social, institutional and economic reasons for sharing data	√	√
9	Planning data management in the project from the start	-	√
10	Planning effective multimedia management	√	√
11	Provide budget information for data management plans	√	√
12	Review data management plans throughout the project lifecycle	-	√

*L=librarian, DC=data curator

2) Collecting stage: observations are made either by hand or by sensors or other instruments,and the data is placed in digital format.

No	Job description	L	DC
1	Participate with creators and data users to foster a culture of sharing and reusing data	√	√
2	How to collect data: manage data for long-term usage and data reuse	√	√
3	Incentives, challenges and obstacles: investigate social, institutional and economic reasons for sharing data	√	√
4	Preserve information: keep raw data	√	√
5	Data floods: implications for data management and culture finding	-	√
6	Use the appropriate field delimiter	-	√
7	Use code consistently	√	√

*L=librarian, DC=data curator

3) Assuring stage: data quality is ensured through checking

No	Job description	L	DC
1	Ensure the quality of the data	√	√
2	Communicate data quality	-	√
3	Confirm the match between data and description in metadata	√	√
4	Consider compatibility in data integration	-	√
5	Developing quality assurance and quality control plans	√	√
6	Double check the incoming data	√	√
7	Ensure quality control	-	√
8	Ensure that the dataset used can be reproduced	-	√
9	Identify missing values and determine the missing value code	-	√
10	Identifying outliers	-	√
11	Identify the calculated value	-	√
12	Mark data with quality control flags	-	√
13	Prove: facilitate knowledge that can be reproduced	-	√
14	Provides a version of information to be used and found	-	√

*L=librarian, DC=data curator

4) Describing stage: data is described accurately and thoroughly using appropriate metadata standards.

No	Job description	L	DC
1	Set descriptive file names	√	√
2	Select and use terminology standards to allow data to be found	√	√
3	Confirm the match between data and description in metadata	√	√
4	Create a data dictionary	√	√
5	Determine the data model	-	√
6	Determine parameters	√	√
7	Describe the format for spatial location	-	√
8	Describe the format for time and	-	√

	date		
9	Describe methods for creating data derivatives	-	√
10	Explain the contents of the data file	√	√
11	Describes the overall organization of the dataset	-	√
12	Describes the expansion and spatial resolution of datasets	-	√
13	Explain the unit of measurement in each observation	-	√
14	Document taxonomic information	-	√
15	Document the stages used in data processing	-	√
16	Document data organizing strategies	-	√

*L=librarian, DC=data curator

5) Preserving stage

No	Job description	L	DC
1	Ensure the quality of the data	√	√
2	Communicate data quality	-	√
3	Confirm the match between data and description in metadata	√	√
4	Consider compatibility in data integration	-	√
5	Developing quality assurance and quality control plans	√	√
6	Double check the incoming data	√	√
7	Ensure quality control	-	√
8	Ensure that the dataset used can be reproduced	-	√
9	Identify missing values and determine the missing value code	-	√
10	Identifying outliers	-	√
11	Identify the calculated value	-	√
12	Mark data with quality control flags	-	√
13	Prove: facilitate knowledge that can be reproduced	-	√
14	Provides a version of information to be used and found	-	√

*L=librarian, DC=data curator

6) Discovering stage

No	Job description	L	DC
1	Shows data using datacasting tools	√	√
2	Provide file name descriptors	-	√
3	Check data and other outputs for print and web accessibility	-	√
4	Participate with creators and data users to foster a culture of sharing and reusing data	√	√
5	Ensure that the dataset used can be reproduced	-	√
6	How and why researchers share data	-	√
7	Make calculated data: assess data usage	-	√
8	Open data and science: optimization of the research process	-	√

*L=librarian, DC=data curator

7) Integrating stage: data from different sources are combined to form a homogeneous data set that is ready to be analyzed.

No	Job description	L	DC
1	Consider data integration compatibility	-	√
2	Document the stages used in data processing	-	√
3	Document the integration of many datasets	-	√
4	Make calculated data: assess data usage	-	√
5	Understand the geospatial parameters of many data sources	-	√

*L=librarian, DC=data curator

8) Analyzing stage

No	Job description	L	DC
1	Consider compatibility in data integration	-	√
2	Describe methods for creating data derivatives	-	√
3	Document the stages used in data processing	-	√
4	Ensure that the dataset used can be reproduced	-	√
5	Identify the most suitable software	-	√

*L=librarian, DC=data curator

Of the 48 activities in the stages of the research data management above, there are 28 jobs that can be done by both librarians and data curators (58%), and 20 jobs that cannot be done by librarians (42%). The type of work that can be done by librarians generally in terms of process is the same as processing library materials; only the object is data. Although the process is the same, the activity cannot be claimed for the librarian's functional position because the object is not library material. Data curation activities that librarians cannot employ are at each stage of the process. Even in the stages of integrating and analyzing data, none of the jobs is suitable to be carried out by librarians.

The ability of librarians needs to be improved to be able to carry out data curation tasks. Some important competencies possessed by librarians are related to data management using information technology. Librarians must also understand the characteristics of

the data in accordance with the field of science. In this context, a subject specialist librarian is very important in his role in managing scientific data.

CONCLUSION

Research data management consists of several stages following the stages of the research cycle. These stages include the stages of planning, collecting, ensuring, explaining, preserving, finding, integrating and analyzing data. Activities in all of these stages can be carried out by data curators. But this profession has not developed yet in Indonesia. Analysis of competencies and data management processes shows the librarian profession can play a role in the processing of scientific data. Librarian competence has not been able to carry out all activities in the data management process, although there are some activities that are similar to the work of librarians in managing data can be done by librarians. In order to be able to carry out the role of data curators, librarians must increase their competence. The main competencies that must be developed by librarians are related to the analysis of data content that is multidisciplinary and the use of advanced information technology.

REFERENCES

[1] J. Kim, E. Warga, and W. E. Moen, "Competencies Required for Digital Curation : An Analysis of Job Advertisements," Int. J. Digit. Curation, vol. 8, no. 1, pp. 66–83, 2013.

[2] M. Tammaro, K. K. Matusiak, F. A. Sposito, and A. Pervan, "Understanding Roles and Responsibilities of Data Curators : An International Perspective," Libellarium, vol. 9, no. 2, pp. 39–47, 2016.

[3] L. R. Johnston, Curating Research Data: A Handbook of Current Practice. Chicago, Illinois: Association of College and Research Libraries A division of the American Library Association, 2017.

[4] M. Hoon, "Careers: 8 jobs that won't exist in 2030," 2017. [Online]. Available: https://www.usatoday.com/story/money/careers/2017/10/13/8-jobs-that-wont-exist-in-2030/104219994/.

[5] M. S. Bracke, "Emerging Data Curation Roles for Librarians : A Case Study of Agricultural Data," J. Agric. Food Information, no. March 2010, pp. 65–74, 2011.

[6] Curtin University, "Research data management: Data management plan," 2018. [Online]. Available: http://libguides.library.curtin.edu.au/c.php?g=202401&p=1333108.

[7] P. Lord and A. Macdonald, "e-Science Curation Report Data curation for e-Science in the UK : an audit to establish requirements for future curation and provision," 2003.

★★★