Research Data Reshaping Cultural Society: Case of the Lebanese University.

Gladys Saade[1] , Dalal Rahme[12]

[1]Faculty of Information Science, Lebanese University, Bauchriyeh, Lebanon
{dyssaade,drahmeh}@gmail.com
[2]Data Services Librarian, American University of Beirut, Beirut, Lebanon
dalal.rahme@aub.edu.lb

Abstract. Research Data Management (RDM) is a new practice in Lebanese academic institutions. The purpose of this study is to assess technological and organizational needs of one of those institutions, the Lebanese University (LU). A questionnaire was sent to academics from all faculties and branches at LU. The survey measured the demographics of survey respondents, their use of data, the degree of openness and the readiness of researchers to manage, share, and preserve datasets. Results described the concerns, challenges and level of commitment of the researchers to data management and sharing. They revealed a moderate level of awareness of information literacy and RDM practices and gave a preliminary figure of the quantity of data generated at LU allowing participants to express their needs. The paper offers guidance for developing a collection of services to support different activities in this area.

Keywords: Research data management, institutional repositories, data sharing, academic universities in Lebanon, lebanese university, open access, scholarly communication.

1. Introduction

The digital revolution has affected all sectors in the society from education, to labor market, to economy, and cultural behavior [1-3]. Technological developments, such as Web 2.0, contributed to consolidating, advancing, and transforming values such as sharing and collaborating. Research Data Management (RDM), a practice that evolved with this technological advancement, is an important way to conduct transparent, ethical, and reproducible research in academia. Since knowledge has long been considered a form of power, detaining information therefore constitutes a step closer to detaining power. RDM best practices equips the research community with this power for it makes data easily accessible to all by centralizing, documenting, and preserving it [4]. In countries that rely on research as a means to develop society, our hypothesis assumes that the application of data management faces logistical challenges including the availability of well-established infrastructures and technical

skills. Nevertheless, this issue can be solved by tailoring training and workshops. However, in the context of the Arab world and Lebanon, research rarely contributes to issues related to public life, citizenship, and democracy in a climate dominated by violence, individualism, seclusion, and fear of inevitability. The biggest challenge in this context therefore resides in the adoption of new concepts such as collaboration and knowledge sharing. Our research questions are the following: To which extent is the Lebanese University research community in Lebanon willing to apply data management best practices? How does this community perceive of RDM? What is this research community doing to adjust to the new practices in RDM and cooperation?

2. Literature Review

The available literature on research in Lebanon and the Arabic world points to a high degree to the severity of the issues facing the Arab research community. The vast majority of the available published research on this topic focuses on highlighting the problems and shortcomings in this area. This is immediately made apparent through a first look at the titles that show a recurrent use of words such as "obstacles", "desert", "problems", and "information gap" [5-7]. The available studies seem to agree that research in the Arab world suffers from a lack of funding. No Arab country spends more than 0.25 of its Gross National Product on scientific research and most of that funding is spent on salaries [8-9]. Stephan argues that researchers in Lebanon struggle to reach needed information because the financial, technical and human resources of libraries are limited and channels connecting researchers to each other are weak, hindering collaboration and cooperation efforts [10]. Unlike the trend in the scientific community, researchers in the social sciences tend to avoid collaboration. The literature suggests that social scientists are not likely to collaborate with researchers from other institutions and are even reluctant to collaborate with researchers from their own institution. The research culture seems to be generally based on competitiveness rather than cooperation. In fact, a research study conducted by Fadia Hoteit reaches the same conclusion. She discusses the lack of specialization in research in the Arab world and the lack of cooperation between researchers within one same entity and between different entities [11]. The last factor affecting research production is of historico-sociological nature. Mourtada argues that the development of thought and social structures in Europe has promoted the development of scientific activities. In that context, science therefore came to be recognized as a positive value. This is, however, not always the case in the Arab world [12].

RDM, as a concept, is pushed forward by governments and funding bodies [13-14] in many countries. Adopting it can contribute to the development of research in Lebanon where there is less collaboration among researchers across Lebanese institutions, slightly better collaboration among researchers from the same institution, and higher collaboration with western institutions [15]. RDM will maximize local collaboration among researchers from different institutions and different fields while minimizing the cost of data collection and reproduction. Considering the weakness and lack of

investment in research in the Arab world, the use of RDM and data sharing will strengthen the research production in the region. We believe that the library is well positioned in the academic institution and that specialized data librarians can play a significant role in facilitating data management along with other stakeholders [16]. They need to overcome the previous mentioned barriers as well as the concerns related to legal issues, privacy issues, and consistent procedures in the research life cycle [17].

Methodology

Our initial strategy was to target the most prominent academic institutions and research centers in Lebanon. Our attempt was unsuccessful for several reasons mentioned in the literature review. Most of the private institutions were not cooperative. They were also not familiar with the concept we are researching and conveyed a concern about the image of their institutions. This was obvious throughout the discussion we had with their communication offices where some requested to remove the question related to institutional affiliation after we explained what the survey was about. The fact is that all of those institutions work competitively not cooperatively, a common trend in the Arab world since no research policies are available [18]. The results collected were mostly (95 percent) from the Lebanese University (LU) which led us to consider only LU as our research setting.

LU, established in 1951, is the only public university in Lebanon and the largest in terms of faculties (16), academic staff (8,000), students (79,000) and libraries (62) [19]. It is also the second producer of scholarly works in Lebanon [20].

In order to understand the data management practices and needs across the LU campuses in Lebanon, the study used the survey method to collect the needed data in an anonymous way. Several case studies on data management in academic settings were examined and questions from previous surveys were borrowed and modified to fit the purpose of the study [21-22]. The survey was administered using an online platform, LimeSurvey, through which each subject received a link allowing only one submission per subject. Questions were made available in three languages: Arabic, English and French. The questions were mostly multiple choice with options to provide free text comments. Analysis of the results such as filtering on aspects of the responses to a particular question and cross tabulations were performed in LimeSurvey as a way to study the relation between the background of the faculty and their interest and engagement in our subject. The questionnaire included 20 questions divided into 5 sections: demographic information, data types and volume, data storage, data literacy, data sharing, and research data management. The survey was conducted between May 1 and May 20, 2017.

Faculty Sample. We were able to obtain 950 email addresses of LU faculty through the university's website. We included faculty members of all academic ranks (professors, associate professors, assistant professors, and instructors) from all

faculties and schools. More than 50 percent of the emails were inactive and undeliverable. The delivered emails informed the participants about the nature of the research, taking their consent before starting the survey.

3. Results

4.1 Demographics

The total number of participants was 161. Fifty two completed the survey; the remaining 109 participants stopped at different stages. Only the complete 52 submissions were taken into consideration for the purpose of this study. People completed the survey in different languages: 11 in Arabic, 20 in English, and 21 in French. As table 1 shows, respondents in the sample came from different faculties: 8 answers from the engineering school, 7 answers from the humanities, 17 answers from social sciences, 17 answers from natural sciences, 2 answers from agricultural sciences, and only 1 answer from health sciences. The study showed that 38 percent of those researchers had been involved in research for more than twenty years, 7.69 percent had sixteen to twenty years of experience in research, 19.23 percent had between eleven and fifteen years of experience, and 30.77 percent had five to ten years of experience in research. In terms of gender, female participants constituted 44.23 percent, male participants constituted 51.92 percent, and 3.85 percent of the participants refused to disclose gender information.

Table 1.Number of participants per faculty

Faculty	Number of Answers
Engineering	8
Humanities	7
Social Sciences	17
Natural Sciences	17
Agricultural Sciences	2
Health Sciences	1
Total	52

4.2 Volume and Type of Data

The volume of data generated by most of the respondents was in gigabytes (57 percent), a smaller portion of respondents (34 percent) produced data in megabytes, and fewer participants (3.85 percent), geoscientists to be more specific, produced data in terabytes. Table 2 shows the type of data used and produced. Researchers worked mostly with standard office documents (78 percent) and web-based datasets (61 percent). They rarely used encoded text related formats such as xml (7.6 percent). Similarly, participants produced largely the same formats: standard office documents

(67.31 percent), structured scientific and statistical data (40.38 percent), images (38.46 percent), internet and web-based data (25 percent), and non-digital data (21.15 percent).

Table 2. Most commonly used and produced data types by LU faculty members

Data Type	Used	Produced
Standard office documents (text, spreadsheets, presentations, etc.)	78.85 %	67.31 %
Internet and web-based data (webpages, e-mails, blogs, social network data, etc.)	61.54 %	25.00 %
Structured scientific and statistical data (e.g. SPSS, GIS, etc.)	51.92 %	40.38 %
Images (JPEG, GIF, TIFF, PNG, etc.)	51.92 %	38.46 %
Archived data (ZIP, RAR, ZAR, etc.)	28.85 %	9.62 %
Non digital data (paper, films, slides, artefacts, etc.)	28.85 %	21.15 %
Databases (e.g. in Access, Oracle, MySQL, etc.)	25.00 %	15.38 %
Software applications (modelling tools, editors, compilers, etc.)	23.08 %	13.46 %
Audio files	17.31 %	5.77 %
Raw (machine-generated) data	17.31 %	11.54 %
Configuration data (parameter settings, logs, library files, etc.)	17.31 %	11.54 %
Structured graphics (CAD, CAM, VRML, etc.)	13.46 %	9.62 %
Source code (scripting, Java, C, C++, etc.)	11.54 %	9.62 %
Encoded text (XML, SGML, etc.)	7.69 %	9.62 %

4.3 Information Literacy

Metadata. As for metadata assignments, 13 percent added technical information (such as file format, file size, and software/hardware needed to use the data), 25 percent described the data file structure, 38.5 percent applied discovery information (e.g. creator, funding body, project title, project ID, keywords, etc.), 40 percent applied administrative information (e.g. creator, date of creation, file name, access terms/restrictions, etc.) and 34 percent of the participants did not assign additional information to their research data as per table 3, below.

Table 3. Kind of metadata assigned to research datasets by LU faculty members

Type of Metadata	Percentage

Administrative information (e.g. creator, date of creation, file name, access terms/restrictions, etc.)	40.38%
Discovery information (e.g. creator, funding body, project title, project ID, keywords, etc.)	38.46%
Technical information (e.g. file format, file size, software/hardware needed to use the data, etc.)	13.46%
Description of the data file (e.g. file/data structure, field tags/descriptions, application rules, etc.)	25.00%
No, I do not assign additional information to my research data	34.62%

Administrative information was applied mostly by natural sciences departments (19.22 percent), followed by social sciences (28.83 percent), engineering (13.46 percent), and humanities (13.46 percent). For discovery information, 17.3 percent of the answers came from Natural Sciences, 9.61 percent were from social sciences, 5.76 percent from humanities, 2 percent from engineering, and 2 percent from agricultural sciences. For technical information, 9.61 percent came from natural sciences, 3.8 percent from engineering, and 2 percent from geosciences. When asked about data structure, tags, and application rules, we once again noticed that respondents who applied descriptions of the data files were mostly from natural sciences backgrounds. As for those who did not assign additional information to their research data they mainly came from social sciences backgrounds (15.3 percent), followed by natural sciences (7.69 percent), then engineering (3.84 percent), agriculture (1.92 percent) , and, 1.92 percent was, surprisingly, from the school of information sciences. Almost all (99 percent) of the participants agreed that the institution should have a predefined metadata set for uploading data into a repository once it exists. More than 52 percent of the participants expressed interest in receiving formal training on metadata best practices.

Data Citing. Almost half of the researchers (55 percent) used a standard style for citing their research data; the other half (45 percent) did not. The LU faculties provide guidelines on citation styles according to 29 percent of the participants whereas the other 71 percent of participants were not aware of the existence of such guidelines. As a result, 40 percent of the participants always cite their data using a standard style, 28 percent often following a standard style, 29 percent rarely or never following a standard citation style. Thirty two percent of the academic staff expressed they needed training in this area. Others seemed less interested in the subject.

File naming. It was found that only 9 percent of the participants were aware of the importance of the consistency in naming their files while others had rarely (17 percent) or never (46 percent) paid attention to being consistent in file naming. The majority (85 percent) agreed on the role of the university in putting in place a recommendation or a standard for file naming, others (11 percent) were not aware of its importance and were neither agreeing nor disagreeing to this suggestion and 34 percent of the participants expressed the need for training.

Version control of datasets. When asked if they use a specific tool or technique to easily recognize a specific version of the data set, 8 percent said they did it always, 20 percent did it often, 20 percent did it rarely, and the remaining 52 percent never did it. Around 5 percent of the participants had already been trained on version control and 33 percent wanted to receive training.
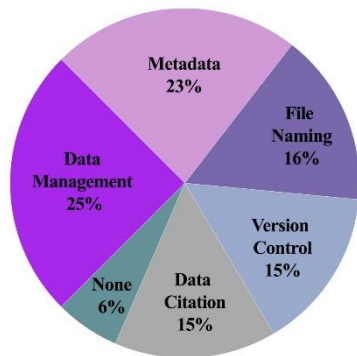

Figure 1- Overall percentage of faculty members expressing their interest in receiving formal training in different areas of RDM.

4.5 Data Storage

Most faculty surveyed (46.14 percent) generate their own data sets during research. Also, the same percentage (46.14 percent) of respondents get data from their research team at the LU. Some researchers (53.85 percent) have their own research network and connections that provide them with datasets. Despite the finding that most faculty generate and use data sets, almost all of them (88.46 percent) do not use data repositories to store their data. This lack of awareness is proven given that most faculty (98.08 percent) say they store their data on their own devices (personal computer, tablet, and external drive). A minority (17.31 percent) uses the cloud for storage, the remaining participants (11.54 percent) use either institutional servers or outside repositories and archives to store their data (table 4). In general, faculty members (77 percent) understand that data should be preserved for long future access, long after the research is completed. Some (9.6 percent) did not think about its importance and therefore were neutral (neither agreeing nor disagreeing to storing data for long terms access). Around 13.5 percent disagreed on the fact that data should be preserved beyond the lifetime of a project.

Table 4. Data Storage of most participant.

Data Storage	percent
own device	98

cloud	17.31
central servers	5.77
outside repositories	5.77
others	3.87

4.4 Data Sharing

More than 53 percent of the participants would share their data upon request, while 26 percent have their data openly accessible to everyone, 25 percent of the participants make their data available only to their research team, 15 percent refuse to share their data with anyone, and 9.6 percent allow partial access to their data. Although 44.73 percent of participants had no concerns related to sharing their data, others had several reasons that made them reluctant to provide data to others: 34.6 percent were concerned about legal and ethical issues and 26.92 percent were concerned about the lack of appropriate policies and rights protection. (Figure 2)
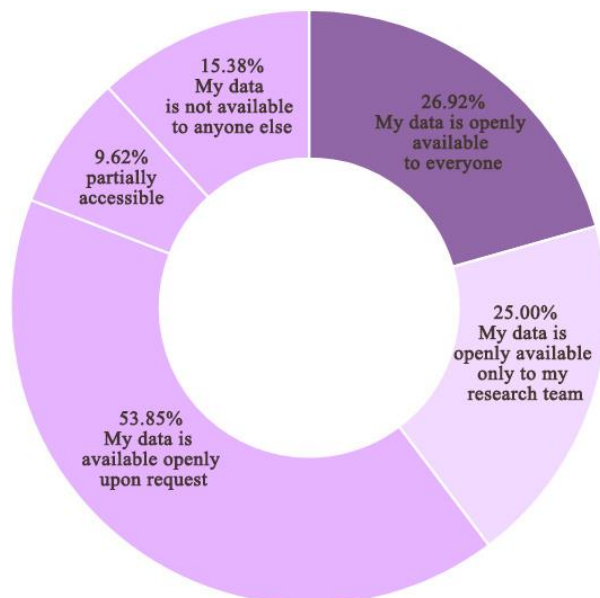

Figure 2 – Researchers sharing their data.

4.6 Data Management Plan (DMP)

According to participants (95 percent), LU does not have a data management plan (DMP) in place. Nevertheless, 5 percent of the participants were familiar with data management plans and have a DMP in place for their current research. Some of the participants (15 percent) were not sure what a DMP was. When asked if they had ever used a DMP, 75 percent of the participants confirmed they have never dealt with it.

Putting a DMP in place is essential to 94 percent of LU researchers, they stated that the institution needs to handle this process.

4. Discussion

The survey revealed the needs and gaps at LU. The fact that 52 participants out of 161 have completed the whole survey may suggest that the majority of the participants did not find the subject relevant to them. On the other hand, we can sense from the complete answers that faculty members at the LU are finding their own ways to apply data management according to the available resources they have in hand. There is certainly a lack of clear policies, infrastructures, and skilled librarians in the institution. The research shows that the LU offered almost no services of any sort to support data management through its campuses. However, participants expressed an interest in the subject of data management and a willingness to be active contributors once an infrastructure is available. Even with a low engagement of the LU libraries in efficiently introducing information literacy practices, researchers are having no major issues in acquiring the data they need from different sources. They are also dealing with almost all the types of data and different volumes. The increase in the volume of data requires good data planning and management. Nevertheless, there is a weakness in other aspects such as data documentation and the use of metadata. The latter is essential for data curation for it allows the researcher to effectively use his dataset in the future without having issues in understanding its content as well as enabling him to share it with others. In addition, respondents expressed weaknesses in citing their datasets especially that there are some guidelines in some faculties and no guidelines in others on how to cite different datatypes, particularly when those are not publically available. Furthermore, organizing information and manipulating it (file naming and version controlling) seems to be the weakest skill among the researchers. Since the university does not provide any means of storage to faculty members, researchers had to find their own way of storing their data through inadequate means. Since none of the respondents mentioned subject specific data repositories, we concluded that they were not aware of their existence. Since LU researchers had no means to disseminate their data efficiently their research data were deemed to ultimately be lost. The absence of data management planning makes researchers of the LU invisible to their peers in other institutions. They do not have the same resources that enables them to manage and share their data in order to advance their research and build significant connections with other scholars in their field.

5. Conclusion

The study results offer a preliminary figure of the management of data at the Lebanese University. It highlighted the strengths, weaknesses and needs of this scholarly community. It also sheds light on the willingness of LU academic staff to work toward enhancements in this area. Most researchers are not reluctant to share data with others. They are willing to cope with new applications of RDM, however they do not have the infrastructure, technical

support, or skills to do it. The libraries have a great opportunity to step in and work as a facilitator in this process. Support from the administration is crucial at this level and development of policies and procedures is needed.

6. Recommendations

- Implementation of an institutional repository is crucial at this level to support faculty in managing, sharing, and preserving their research data.
- Development of policies and guidelines is in order to promote a knowledge sharing culture and encourage academics to engage in the activities related to research data management since local funding agencies are requiring data management plans.
- Tailoring of special workshops and training sessions should be done in collaboration with the 62 LU libraries to reinforce information literacy and data management skills.
- Librarians should be empowered through training so they might be able to support faculty members throughout the data life cycle from creation, to documentation, sharing, and preservation of data.

References

1. Bouquillion, P., Matthews, J.: Le Web Collaboratif. Mutations des Industries de la Culture et de la Communication. Presses Universitaires de Grenoble (2010).

2. Dagnaud, M.: Le Modèle Californien: Comment l'Esprit Collaboratif Change le Monde. Odile Jacob (2016).

3. Rieffel, R.: Revolution Numerique, Revolution Culturelle? Folio actuel, Paris (2016).

4. Surkis, A., Read, K.: Research Data Management. Journal of the Medical Library Association. 103(3): 154–156 (2015).

5. Del Castillo, D.: The Arab World's Scientific Desert. The Chronical of Higher Education. (March 5, 2004).

6. A.L ALLATIF.- Mouawiqat al bahth al ilmi. Al saouddiya (2008)

7. MAADAN SH.waqeh al bahth al ilmi fi al watan al arabi fi dil al fajwa al raqmiyya. Majallat al ouloum al insaniyya.38 (2012). Retrieved from: http://revue.umc.edu.dz/index.php/h/article/view/1337

8. Zahlan.a.- al arab wa tahadiyat al asr. Beirut, markaz dirassat al wihda al Arabiya, 1990

9. Badran, A.: Les Etats Arabes. Paris: Unesco (2006). Retrieved from http://www.unesco.org/fileadmin/MULTIMEDIA/HQ/SC/pdf/sc_usr05_arab_fr.pdf

10. Stephan, M.: Recherche et Communication. Bahithat , 76-95 (1997).

11. Hotet, F.: al mouwassafat al bahthiya lilrasael al jamiiya fi elm al nafs.-Bahithat 1997 ص24-40

12. Mourtada, A.: The Conditions of Academic Research in the Arab World. Bahithat (1997).

13. Funari, M.: "Research Data and Humanities: A European Context." *Italian Journal of Library and Information Science* 5(1): 209–36(2014).

14. Surkis, A., Read, K.: "Research data management", Journal of the Medical Library Association, Vol. 103 No. 3, pp. 154-156(2015).

15. Hakim Rahme, D.:*Co-Authorship Patterns at the Medical School of the American University of Beirut between 2004 and 2014*. Paper presented at: IFLA WLIC 2016 – Columbus, OH – Connections. Collaboration. Community in Session 101 - Poster Sessions (2016).

16. Flores, J. R., Brodeur, J. J., Daniels, M. G., Nicholls, N., & Turnator, E.L. Libraries and the research data management landscape. In J. C. Maclachlan, E. A. Waraksa, & C. Williford (Eds.). *The Process of Discovery: The CLIR Postdoctoral Fellowship Program and the Future of the Academy* (pp. 82–102, 2015). Washington DC: Council on Library and Information Resources. Retrieved from www.clir.org/pubs/reports/pub167/pub167.pdf#page=88

17. Eschenfelder, K. R. and Johnson, A., "Managing the Data Commons: Controlled Sharing of Scholarly Data" (2014). *University Libraries Faculty & Staff Contributions*. 15. http://scholar.colorado.edu/libr_facpapers/15

18. Sixth Arabic Report on education and Development, http://fikrconferences.org/sites/default/files/takrir6.pdf, 2010, last accessed 2017/05/28.

19. Lebanese University Homepage, http://www.ul.edu.lb/, last accessed 2017/05/28.

20. Scopus, https://www.scopus.com, last accessed 2017/05/28.

21. Questionnaire National Research Data Survey, https://e-infrastructures.at/fileadmin/user_upload/p_e_infrastructures/PDFs/Questionnaire_e-Infra_Research_Data_Survey_Aug15.pdf, last accessed 2017/05/28.

22. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al.: Data Sharing by Scientists: Practices and Perceptions. PLoS One, 6(6), e21101 (2011).