

State of Research Data Management in Latin American Universities 2022

Authors: Amado Jose ¹, Olivera Dianelis ², Machado Manuel ², Vancauwenbergh Sadia ³, Limaymanta Cesar ¹, Gonzales Celso ¹, Peralta Maria Josefa², Garcia Arelexys ⁴, Ballivian Andrea ⁵, Meneses Grizly ², Chavez Georgina ⁵, Salvatierra Angel ¹, Vlietinck Hanne ³, Goovaerts Marc ³

1. Universidad Nacional Mayor de San Marcos, Lima, Peru.
2. Universidad Central "Marta Abreu" de las Villas, Santa Clara, Cuba.
3. Hasselt University, Hasselt, Belgium.
4. Universidad de Cienfuegos "Carlos Rafael Rodríguez", Cienfuegos, Cuba.
5. Universidad Católica Boliviana "San Pablo", La Paz, Bolivia.

SUMMARY

Introduction: Adequate research data management (RDM) improves the reliability of knowledge.

Objective: To describe RDM in researchers of some Latin American universities.

Methodology: Descriptive study carried out in two universities in Cuba, one in Peru and one in Bolivia. A survey was applied by means of a structured questionnaire with five dimensions (data type-format, storage-archiving, infrastructure-services, ethics-legal, accessibility-reuse). Data were coded and descriptive statistics were applied.

Results: 652 researchers from Cuba (74.5%), Peru (17.9%) and Bolivia (7.6%) participated. Type and format of data: 73.9% of researchers generate their own research data, quantitative type 81.1%, in text format 87.3%, digital format 67.2%, and spreadsheets 58.0%. Data storage and archiving: over 65% report having small to medium data volume (<100 GB), 61% store data on personal computers, 29.8% store data at their institution, and 51% report losing research data. Infrastructure and services: 60% referred that data misuse is an obstacle to data sharing, suggesting improving institutional repository support. Ethical and legal aspects: 34.8% use personal or sensitive data and 65.6% do not know the legislation on research data. Accessibility and reusability: more than 60% generate reusable data, 36.2% create passwords for their research data, sharing them via email and the cloud, but the data are accessible with restrictions.

Conclusions: Most researchers generate quantitative data in low to medium volume in digital format, being stored on personal computers, with high risk of loss and vulnerability. There is a frequent fear of misuse of research data, low awareness of legal aspects and deficiencies in institutional repositories.

Key Word: Research Data Management; Questionnaire; Diagnostic; Researcher; Library Services

Introduction

Adequate research data management (RDM) is a requirement in the current context, in addition to improving the reliability of the results and allowing the global exchange of knowledge, it produces greater recognition of the research and the authors who openly share their data have greater scientific impact (Federer et al., 2020; Sholler et al., 2019; Sudrajat et al., 2024).

The science is a public object and an universal right, which should seek to benefit society and solve major global problems. Therefore, scientific knowledge and research results should be shared with stakeholders, requiring a collective effort of researchers, policy makers and civil society organizations for this purpose (Angelozzi, 2020; Vancauwenbergh, 2021). Universities have a very important role in higher education, scientific research and social projection, allowing the strengthening of cognitive and research capacities (Amésquita et al., 2020).

Latin America has several initiatives that promote open science and tends to become a leader in non-commercial knowledge, but it has yet to develop policies and standardize criteria in the academic and scientific community (Meneses-Placeres et al., 2022; Universidad Nacional de Lanus. Argentina & Martinovich, 2021). In this context, this article seeks to describe the situation of RDM in teacher-researchers from some Latin American universities.

Methodology:

Descriptive study conducted between November and December 2022. Non-probabilistic sampling in professors who carry out research in four Latin American universities. Universidad Central "Marta Abreu" de las Villas (Santa Clara) and Universidad de Cienfuegos "Carlos Rafael Rodríguez" (Cienfuegos) in Cuba; Universidad Nacional Mayor de San Marcos (Lima) in Peru; and Universidad Católica Boliviana "San Pablo" (La Paz) in Bolivia. An area of science was assigned to each participant, according to the OECD classification of areas of science and technology .(European Commission, n.d.)

A survey (virtual and physical) was applied using a structured questionnaire based on a publication of the University of Hasselt (Haverbeke et al., 2018) and adapted by the Project Research Data Management Strategy (RDMS LatAm) team. This questionnaire consists of 42 questions distributed in 5 dimensions (data type and format 9, storage-archiving 11, infrastructure-services 5, ethics-legal 6, accessibility-use 11).

Data were coded and analyzed with SPSS 27, applying descriptive statistics and comparing responses according to research level and participating countries. Institutional authorization and informed consent were requested from the participants, preserving the confidentiality of the data.

Results:

652 researchers from universities in Cuba, Bolivia and Peru participated with low proportion of health sciences, humanities and agricultural sciences. A total of 22.7% of the participants were pursuing doctoral studies (PhD) and 28.8% were considered senior researchers.

Dimension 1. Type and format of data: 73.9% of researchers generate their own research data, being more frequently of quantitative type 81.1%, in text format 87.3% and spreadsheets 58.0%, with no difference in researchers who are pursuing doctoral studies. The majority (68.1%) refer that more than 60% of their research data are used in digital format. 67.2% reported that they document their research data and 28.1% mentioned that they always develop a data management plan.

Dimension 2. Data storage and archiving: More than 65% of participants report having a small or medium data volume (<100 GB), less than half manage multiple versions of their data. A low percentage of research data storage in the reference institution (physical or virtual) is reported, being higher in Cuba and Bolivia, most of them continue to store their data in their personal computers. About half of the researchers report having lost research data. Less than half report having all their data backed up, except Bolivia with 57%. A low proportion of researchers have backup copies of their data at their institution (Bolivia 16.3, Cuba 40.5 and Peru 8.5%); reporting a higher proportion on individual portable devices. At the end of the research, less than 30% say that the data are kept at their research center or university.

Dimension 3. Infrastructure and services: More than 80% of researchers report that scientific recognition is an incentive to improve the management of research data. About 60% of participants referred that the misuse of data is an obstacle for the adequate management of research data, considering that workshops, technical assistance and the improvement of the institutional repository are the best support measures in this aspect.

Dimension 4. Ethical and legal aspects: Almost half of the researchers reported using personal or sensitive data in their research (Bolivia 53.1, Cuba 30.3 and Peru 46.1%). A high proportion reported knowing who is the owner of the data, but more than 60% reported not knowing the legislation regarding research data management and more than 80% were never involved in legal arguments or user agreements.

Dimension 5. Accessibility and reuse of data: Most researchers generate reusable data (Bolivia 83.7, Cuba 81.5 and Peru 59.8%), sharing it most frequently by email and the cloud, but in Cuba it is done more by portable means. Data are accessible to science-related personnel and with restrictions, in greater proportion in universities in Cuba. Less than 40% of researchers create passwords for their research data. More than half of researchers do not share their data for fear of confidentiality breaches.

Table 1. Participating researchers by country and area of knowledge in the diagnosis of research data management in Latin American universities 2022.

Subject área	Cuba	Peru	Bolivia	Total
Agricultural	24	6	0	30
Humanties	43	3	1	47
Engineering	89	15	26	130
Medicine- Health	5	47	1	53
Natural	108	22	1	131
Social	217	24	20	261
Total	486	117	49	652

Figure 1. Proportion of participants per university in the diagnosis of research data management in Latin American universities 2022.

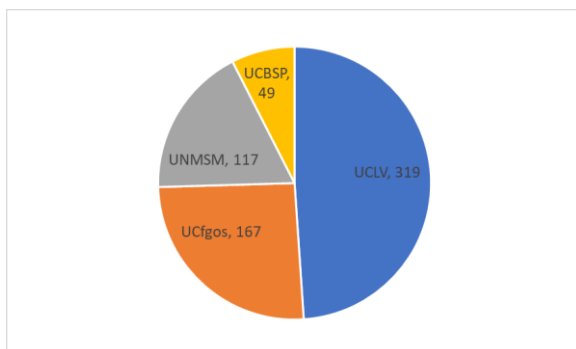


Table 2. Responses related to the dimension Type and Format of research data in Latin American universities 2022, comparing whether they have a doctoral degree with the total number of participants.

Dimension 1: Type and format of data	PhD student		Total		p
	n	%	n	%	
Source: Public organizations	55	37.2	221	33.9	0.340
Source: Commercial institutions	18	12.2	82	12.6	0.863
Source: Open and public data	64	43.2	285	43.7	0.896
Generate own data	104	70.3	483	73.9	0.249
Quantitative data type	116	70.8	529	81.1	0.330
Qualitative data type	121	81.8	507	77.8	0.184
Work with primary data	48	32.4	211	32.4	0.983
Work with processed data	57	38.5	307	49.1	0.017
Work with clinical data	13	8.8	67	10.3	0.496
Use laboratory notebooks	6	4.1	47	7.2	0.091
Text format	133	89.9	568	87.3	0.278
Spreadsheet format	91	61.5	378	58.0	0.325
Database format	24	16.2	152	23.3	0.020
Image data	76	51.9	344	52.8	0.696
Audio data	22	14.9	95	14.6	0.908
Video data	25	16.9	125	19.2	0.423
Applications and other type data	30	20.3	175	26.8	0.040
Documents the data	86	58.1	409	62.7	0.456
Always make a data plan	38	25.7	183	28.1	0.740

Table 3. Responses related to the other dimension of research data management in Latin American universities 2022, comparing the countries of origin.

Dimensions	Bolivia		Cuba		Peru		P
	n	%	n	%	n	%	
Dimension 2: Data storage and archiving							
Small to médium data volume (<100 GB)	32	65.4	216	65.0	85	72.6	0.290
Stores data in university	9	18.4	169	34.8	16	13.7	<0.001
Stores in external center	7	14.3	41	8.4	15	12.8	0.185
Stores in university cloud	19	38.8	138	28.4	22	18.8	0.021
Stores in external cloud	16	32.7	55	11.3	20	17.1	<0.001
Stores on individual devices	16	32.7	134	27.6	24	20.5	0.186
Stores in personal computer	34	69.4	333	68.5	72	61.5	0.334
All data is backed up	28	57.1	163	33.5	47	40.2	0.011
Has had data loss	23	46.9	249	51.2	60	51.3	0.689
Allways take securitu measures for the data	20	40.8	258	53.1	59	50.4	0.691
At the end, data are stored in the reseach unit	8	16.3	142	29.2	16	13.7	0.018
Dimension 3: Infrastructure and services							
Incentivized by scientific recognition	41	83.7	477	92.0	98	83.8	0.010
Incentivized by new contacts	33	67.3	249	51.2	48	41.0	0.007
Considers data misuse as an obstacle	30	61.2	266	54.7	73	62.4	0.257
Considers processes as an obstacle	11	22.4	50	10.3	19	16.2	0.017
Considers Infrastructure problems as an obstacle	8	16.3	148	30.5	10	8.5	<0.001
Workshops support	30	61.2	348	71.6	78	66.7	0.222
Technical Assistance support	35	71.4	300	61.7	83	70.9	0.095
Repositories support	30	61.2	267	54.9	47	40.2	0.008
Dimension 4: Ethical and legal aspects							
Uses personal or sensitive data	26	53.1	147	30.3	54	46.1	0.002
Uses open licenses	6	12.2	54	11.1	17	14.5	0.586

Uses restricted licenses	23	46.9	107	22.0	29	24.8	< 0.001
No knowledge of research data legislation	37	75.5	299	61.5	92	78.6	< 0.001
Was never subject to legal arguments or user agreements	41	83.7	403	82.9	92	78.6	0.787
Dimension 5: Accessibility and reuse of data							
Generate reusable data	41	83.7	396	81.5	70	59.8	< 0.001
Your data is accessible to the Public	9	18.4	68	14.0	19	16.2	0.625
Your data is accessible only to Scientists	17	34.7	315	64.8	46	39.3	< 0.001
There are restrictions to access your data	38	77.6	363	74.7	72	61.5	0.012
Create passwords for your data	13	26.5	193	39.7	30	25.6	0.009
Share data by email	39	79.6	341	70.2	76	65.0	0.168
Share data in the Cloud	32	65.3	158	32.5	55	47.0	< 0.001
Share data on portable devices	9	18.4	329	67.7	26	22.2	< 0.001
Found adequate storage space	12	24.5	117	24.1	33	28.2	0.649
Found adequate nomenclature	19	38.8	55	11.3	34	29.1	< 0.001

Discussion:

In Latin American universities a significant amount of research data is generated, but the state of research data management and planning is incipient. A good percentage of researchers state that they have a research data plan, however, in local institutions there are no specific regulations or policies on research data management (Muñoz Bravo, 2021). The present work allows us to know the situation of research data management in universities of the region and allows us to develop this topic based on the findings.

Type and format of data

Regarding the type and format of research data, the vast majority of participating researchers generated their own quantitative data in digital format, similar to other publications (Masinde et al., 2021; Tripathi et al., 2017). There is a low report of qualitative research or research that generates large volume data. Only a lower proportion of processed data or the use of applications or other formats was found among beginning researchers.

Data storage and archiving

Data are stored inadequately, with a high risk of loss and vulnerability, similar to other developing countries (Masinde et al., 2021; Tripathi et al., 2017), but different from advanced research centers (Chen & Wu, 2017). In addition, there is evidence of inadequate functioning of institutional repositories, and the role of the library and the university as manager and often funder of research

should be strengthened (Muñoz Bravo, 2021; Vancauwenbergh & Poelmans, 2019). There are also many researchers who still store their data in physical form and without adequate security measures.

Data infrastructure and services

Regarding data services, most of the participants reported that they do not use a methodology for data processing. There is a greater use of data services by researchers from universities with more experience in collaborative work and external funding. The lack of infrastructure and data services provided by the university to researchers is also evident, and work should be done to improve this aspect (Chen & Wu, 2017; Manu, 2018).

Ethical and legal aspects of the data

More than one third of the participating researchers use sensitive data, this is influenced by the area of science to which they are dedicated and in the present study there was little participation from the areas of health and humanities, so this data could be underestimated. Regardless of this, the lack of procedures to follow good research practices (such as anonymization or pseudonymization) is striking. Training activities in this aspect are necessary (Masinde et al., 2021; Tripathi et al., 2017). Another important problem is the lack of institutional and national policies regarding research data management (Huang et al., 2021), here Cuba differs, where they already have best practices in this regard and the adoption of a national policy is in progress.

Data accessibility and reuse

The vast majority of researchers reported having data that should be shared, but almost half reported insecure procedures or simply not sharing them. Among the obstacles to this, doubts about confidentiality and lack of knowledge of the processes were mentioned. This requires greater involvement of the data management system by the respective offices of the university, to improve teamwork and inter-institutional agreements (Sholler et al., 2019).

The main limitations of the study is that the virtual questionnaire was quite extensive, in few universities in each country and with little participation from some areas of science such as agriculture, humanities and health sciences. However, it is a first effort to gain a better understanding of this topic and serves as a basis for further research

Conclusions:

Most researchers generate quantitative data in low or medium volume digital format, being stored on personal computers, with high risk of loss and vulnerability. Less than one third of the researchers keep the data at their institution after completing the research. Most researchers obtain reusable data and share it by unsecured means, almost half of them generate sensitive data, but there is a lack of knowledge of legal aspects. Finally, there is a frequent fear of misuse of research data, low awareness of legal aspects and deficiencies in institutional repositories.

References

Amésquita, M. J. A., Ávila, D. P., Condori, A. F. Q., & García, M. S. (2020). La investigación científica en la universidad peruana. *Journal of the Academy*, 3, Article 3.

<https://doi.org/10.47058/joa3.3>

- Angelozzi, S. M. (2020). La gestión de datos de investigación en abierto: Introducción al rol emergente para las bibliotecas universitarias y científicas argentinas. *Palabra Clave (La Plata)*, 9(2), Article 2. <https://doi.org/10.24215/18539912e091>
- Chen, X., & Wu, M. (2017). Survey on the Needs for Chemistry Research Data Management and Sharing. *The Journal of Academic Librarianship*, 43(4), 346–353. <https://doi.org/10.1016/j.acalib.2017.06.006>
- European Commission. (n.d.). *About Field of science and technology classification | Joinup*. Retrieved 8 December 2023, from <https://joinup.ec.europa.eu/collection/eu-semantic-interoperability-catalogue/solution/field-science-and-technology-classification/about>
- Federer, L., Foster, E. D., Glusker, A., Henderson, M., Read, K., & Zhao, S. (2020). The Medical Library Association Data Services Competency: A framework for data science and open science skills development. *Journal of the Medical Library Association*, 108(2), Article 2. <https://doi.org/10.5195/jmla.2020.909>
- Haverbeke, H. V., Claeys, J., Hanus, J., Voecht, M. D., Mertens, M., Elsen, H., Vancauwenbergh, S., Amez, L., & Gaeveren, K. V. (2018). *Research data management en de Vlaamse Universiteiten: White paper*. <https://www.semanticscholar.org/paper/Research-data-management-en-de-Vlaamse-White-paper-Haverbeke-Claeys/0082781642ac9b0f29a6696461985665946c84b9>
- Huang, Y., Cox, A. M., & Saffi, L. (2021). Research data management policy and practice in Chinese university libraries. *Journal of the Association for Information Science and Technology*, 72(4), 493–506. <https://doi.org/10.1002/asi.24413>
- Manu, T. (2018). *Researchers' Perceptions on Research Data Management: A Survey*.
- Masinde, J. M., Chen, J., & Muthee, D. W. (2021). Researchers' Perceptions of Research Data Management Activities at an Academic Library in a Developing Country: *International Journal of Library and Information Services*, 10(2), 1–17. <https://doi.org/10.4018/IJLIS.20210701.0a11>
- Meneses-Placeres, G., Reinaldo, L. A. Á., & Rivero, M. O. M. (2022). *Revisión de las Prácticas de Ciencia Abierta en América Latina y el Caribe*. 3.
- Muñoz Bravo, A. (2021). *Implementación y gestión de un repositorio de datos de investigación para bibliotecas: Buenas prácticas y desafíos*. <https://doi.org/10.7764/tesisUC/ING/52766>

- Sholler, D., Ram, K., Boettiger, C., & Katz, D. S. (2019). Enforcing public data archiving policies in academic publishing: A study of ecology journals. *Big Data & Society*, 6(1), 2053951719836258. <https://doi.org/10.1177/2053951719836258>
- Sudrajat, R., Ruchjana, B. N., Abdullah, A. S., & Budiarto, R. (2024). Web-based information system framework for the digitization of historical databases and endowments. *International Journal of Data and Network Science*, 8(1), 319–328. <https://doi.org/10.5267/j.ijdns.2023.9.022>
- Tripathi, M., Chand, M., Sonkar, S. K., & Jeevan, V. K. J. (2017). A brief assessment of researchers' perceptions towards research data in India. *IFLA Journal*, 43(1), 22–39. <https://doi.org/10.1177/0340035216686984>
- Universidad Nacional de Lanus. Argentina, & Martinovich, V. (2021). Ciencia abierta en América Latina: Repensar la interdependencia dinámica entre las ciencias y la sociedad. *Espacios*, 42(24), 1–14. <https://doi.org/10.48082/espacios-a21v42n24p01>
- Vancauwenbergh, S. (2021). FAIR and Open Research Metadata as Leverage for Digital Libraries: The Flemish Case. In S. Vancauwenbergh (Ed.), *Digital Libraries—Advancing Open Science*. IntechOpen. <https://doi.org/10.5772/intechopen.97192>
- Vancauwenbergh, S., & Poelmans, H. (2019). The creation of the Flemish research discipline list, an important step forward in harmonising research information (systems). *Procedia Computer Science*, 146, 265–278. <https://doi.org/10.1016/j.procs.2019.01.075>