

Development of a culture on Research Data Management: experiences from the international RDMS-LatAm Project

Authors:

PhD. Manuel Osvaldo Machado Rivero¹, PhD. Sadia Vancauwenberg², M.Sc. Dianelis Olivera Batista³, M.Sc. César Halley Limaymanta Álvarez⁴, M.Sc. Andrea Ballivian Blanco,⁵ M.Sc. Arelexys Antonia García Cartaya⁶, PhD. María Josefa Peralta González⁷, PhD. Georgina Aurelia Chávez⁸, M.Sc. Celso Morelan Gonzales Cam⁹, M.Sc. José Percy Amado Tineo¹⁰, PhD. Angel Salvatierra Melgar¹¹, Bs. Marc Goovaerts¹², Hanne Vlietinck¹³, PhD. Grizly Meneses Placeres¹⁴

1. Universidad Central “Marta Abreu” de Las Villas. Cuba. Email: mosvaldo@uclv.edu.cu , <https://orcid.org/0000-0001-5095-8389>
2. Hasselt University. Belgium. Email: sadia.vancauwenbergh@uhasselt.be
3. Universidad Central “Marta Abreu” de Las Villas. Cuba. Email: dolivera@uclv.cu , <https://orcid.org/0000-0002-7081-9092>
4. Universidad Nacional Mayor de San Marcos. Peru. Email: climaymantaa@unmsm.edu.pe , <https://orcid.org/0000-0002-8797-4275>
5. Universidad Católica Bolivina “San Pablo”. Bolivia. Email: aballivian@ucb.edu.bo , <https://orcid.org/0000-0002-7084-8125>
6. Universidad de Cienfuegos “Carlos Rafael Rodríguez”. Cuba. Email: aagacrcia@ucf.edu , <https://orcid.org/0000-0003-3885-6907>
7. Universidad Central “Marta Abreu” de Las Villas. Cuba. Email: mjosefa@uclv.edu.cu , <https://orcid.org/0000-0002-5188-2328>
8. Universidad Católica Bolivina “San Pablo”. Bolivia. Email: gchavez@ucb.edu.bo , <https://orcid.org/0000-0002-8414-9683>
9. Universidad Nacional Mayor de San Marcos. Peru. Email: cgonzalesca@unmsm.edu.pe , <https://orcid.org/0000-0002-7968-8236>
10. Universidad Nacional Mayor de San Marcos. Peru. Email: jamadot@unmsm.edu.pe , <https://orcid.org/0000-0002-3286-4650>
11. Universidad Nacional Mayor de San Marcos. Peru. Email: asalvatierram@unmsm.edu.pe , <https://orcid.org/0000-0003-2817-630X>
12. Hasselt University. Belgium. Email: marc.goovaerts@uhasselt.be , <https://orcid.org/0000-0003-1809-4159>
13. Hasselt University. Belgium. Email: hanne.vlietinck@uhasselt.be , <https://orcid.org/0000-0003-3995-4277>
14. Universidad Central “Marta Abreu” de Las Villas. Cuba. Email: grizly@uclv.edu.cu , <https://orcid.org/0000-0003-3587-5061>

Abstract

Research Data Management is based on the so-called FAIR Principles of data, whose fundamental objective is to make data freely findable, accessible, interoperable and reusable, thus encouraging the adoption of these principles by institutions has become a standard of Best Practices. Universities, in their role as teaching and research centers, should embrace such principles in order to develop science of higher quality, accessibility, visibility and impact.

RDMS-LatAm identifies the project entitled: "Development of a Strategy for Research Data Management in Higher Education and Scientific Institutions within the Latin American context". An initiative that has brought together the Universidad Central "Marta Abreu" de las Villas (CUBA) as coordinator, Universidad Nacional Mayor de San Marcos (Peru), Universidad Católica de Bolivia "San Pablo" and the University of Hasselt (Belgium), with the support of VLIR-UOS (Belgium). The objective of this project is: "To develop capacities to adopt good practices on Research Data Management in the Latin American context".

So far, RDMS-LatAm's actions have allowed the development of diagnostics on data management in member universities, the elaboration of policy proposals, the training of teachers, researchers, librarians and undergraduate and graduate students, as well as the development of a new instrument to identify research data management practices in the institutions. In addition, collaborative relationships have been established with other centers that can benefit from the results achieved.

Keywords: Research Data Management , Open Science, RDMS-LatAm, Cuba, Peru, Bolivia

Acknowledgements: This research is part of the actions of the international project RDMS-LatAm, sponsored by VLIR-UOS.

Introduction

Scientific activity is currently in the so-called data-driven paradigm. In it, research data (RD) play a key role throughout the scientific process, from the planning phase to the publication of research results. Among the reasons that have guided data sharing highlights the scientific community's need to maximize data transparency and reliability; enable verification of research results; facilitate collaboration between data users, data creators and reusers, among others (Maredata, 2018).

As data are essential in processes to ensure reproducibility, innovation and decision making, they have become an important and invaluable intellectual resource within the research ecosystem (Mons, 2018).

The concepts of Research Data Management (RDM), Open Data (OD) and Open Science (OS) are relatively new around the world . (Angelozzi, 2020; Andrikopoulou et al., 2022). In Latin America and the Caribbean they have been introduced by organizations such as the Economic Commission for Latin America and the Caribbean (ECLAC), the Confederation of Open Access Repositories (COAR) and the Federated

Network of Institutional Repositories of Scientific Publications of Latin America (La Referencia). In 2020 UNESCO made a "Recommendation on Open Science", with regional offices (e.g. Havana) as implementers in the Latin American and Caribbean region. Different countries, including Peru, have developed legislation on open access and repositories (La Referencia, 2018; UNESCO 2020).

Research Data Management is based on the so-called "FAIR Principles" of data, whose fundamental objective is to make the latter freely findable, accessible, interoperable and reusable, thus promoting the adoption of these principles by institutions as standards of good practices. Universities, in their role as teaching and research centers, must embrace such principles in order to develop science of higher quality, accessibility, visibility and impact. The adoption of these principles for data management in institutions has become what is currently considered as RDM Best Practices (Campos et al., 2023).

RDMS-LatAm identifies the project entitled: "Development of a Strategy for Research Data Management in Higher Education and Scientific Institutions within the Latin American context". An initiative that has brought together the Universidad Central "Marta Abreu" de las Villas (CUBA) as coordinator, Universidad Nacional Mayor de San Marcos (Peru), Universidad Católica de Bolivia "San Pablo" and the University of Hasselt (Belgium), with the support of VLIR-UOS (Belgium). The objective of this project is: "To develop capacities to adopt good practices on Research Data Management in the Latin American context".

So far, RDMS-LatAm actions have allowed the development of diagnostics on data management in member universities, the elaboration of policy proposals, the training of teachers, researchers, librarians and undergraduate and graduate students, as well as the development of a new instrument to identify practices on research data management in institutions. In addition, collaborative relationships have been established with other centers that can benefit from the results achieved.

Methodology

This research is descriptive in nature as it reviews the characteristics and results achieved by the project: "Development of a research data management strategy in Higher Education and Research Institutes in a Latin American context" (RDMS-LatAm). Documentary analysis was used as a research method to investigate and interpret, as well as to review different existing documentary sources to understand the RDM phenomenon.

So far, the actions of RDMS-LatAm have allowed the development of 4 lines of work:

1-Diagnosis on RDM practices

2-Policies for the RDM development

3-Training of teachers, researchers, librarians and managers in relation to RDM

4-Collaboration Networks

Results

1-Diagnosis on RDM practices

A sample of professors and researchers and another sample of managers from the four Latin American universities that are members of RDMS-LatAm (table 1) were selected and administered an online questionnaire with the aim to identifying the practices developed in these institutions to manage research data.

The questionnaire used constituted an adaptation of the one developed and applied at the University of Hasselt (Belgium) during a study similar to the present one (Vanhaverbeke et al., 2018). The adaptation was based on a translation into the Spanish language, taking into account the appropriate and contextualized use of the different terms. It was composed of 42 questions grouped into 5 dimensions:

- Data types and formats
- Data storage, archiving, backup and loss of data
- Infrastructure and services
- Ethical and legal aspects
- Accessibility and reuse
- La aplicación del cuestionario se realizó en línea.

Table 1. Distribution of the samples by universities

Universities	Country	PR	DM	Total
Universidad Central “Marta Abreu” de Las Villas	CUB	320	112	432
Universidad Nacional Mayor de San Marcos	PER	119	23	142
Universidad Católica Boliviana “San Pablo”	BOL	49	25	74
Universidad de Cienfuegos “Carlos Rafael Rodríguez”	CUB	169	25	194
TOTAL		657	185	842

Leyenda:

CUB: Cuba

PER: Peru

BOL: Bolivia

PR: Professor and Researchers

DM: Decision Makers

Among the PRs, the percentage of subjects using data generated in their own research was higher (45.05%), while for the DMs it was data from public organizations (82.07%). In both cases the lowest percentages were for data obtained through commercial providers. There were no major differences between the use of qualitative or quantitative data by the two groups.

The most used digital data formats coincided for both PR and DM, being text documents (PR 27.19 % and DM 84.78 %), spreadsheets (PR 18.09 % and DM 72.28 %) and Graphics/Images (PR 16.42 % and DM 60.33 %).

The personal computer turned out to be the means of storage of research data most frequently used by the RPs (37.53%), while 77.72% of the DMs recognize that their researchers tend to prefer these personal devices to others, even over institutional storage services. These devices also turned out to be the place where backups are most frequently deposited.

A 17.05 % of the RPs do not back up their data with backup copies, 44.90 % do so only partially and 35.92% of them acknowledged that they have ever suffered massive loss of research data.

Some 73.37 % of the DMs recognized that in their institutions each researcher is responsible for archiving their own data, while 46.2 % stated that in their universities young researchers do not receive any preparation about how to handle RD.

The main channels for data exchange among researchers were e-mail 38,03% and devices such as USB sticks and external disks 30,28 %. Data exchange was developed mainly between researchers from the national and international scientific community.

Researchers expect from their institutions mainly the development of workshops where they can be trained in data management.

A 30.98% of the DMs responded that in their universities there is no policy for Research Data Management, 17.93 % do not know whether or not a policy exists, 33.15 % stated that it does exist (although in contact with such universities it was learned that such policies are not really in place).

2-Policies for the RDM development

After an analysis of experiences on the continent on OC practices and especially on the management of research data, the four member institutions have developed policy proposals for RDM. They are being analyzed by administrative directives for approval and implementation. The policy proposals have been structured with the following sections:

- Introduction: The basic principles and general guidelines are established, as well as the institution's recognition of the need and impact of RDM as a practice that improve visibility, reliability, reproducibility and impact of the scientific activity developed by its researchers.
- Definitions: Concepts most frequently used in the RDM are listed.
- Scope: institutional, covering all research to be developed by each university nationally and internationally.
- Purpose: To develop best practices for the optimal management of research data; to implement the principles for RDM to help ensure high quality research.

- FAIR Principles of Research Data: FAIR principles are a mandatory requirement. Making every dataset FAIR should be a goal of every research project.
- Research Data Management Plans: Research Data Management Plan will be mandatory for every research project, to plan and check the actions to be executed with data sets during and after the conclusion of each research project.
- Metadata and Data Documentation: Metadata are essential to achieve interoperability and data exchange, standardization of the metadata schemas used in each investigation is required.
- Roles and Responsibilities: Roles to be played by researchers and various administrative departments of each university.
- Intellectual property and Author's rights: The situations in which each institution will exercise the ownership (exclusive or shared) of the data generated in the different research projects are established, as well as the moral rights corresponding to the authors (researchers), the use of licenses in accordance with the degree of socialization and use that the data published by the university may have, and the use of licenses in accordance with the degree of socialization and use that the data published by the university may have.
- Others Documents and Related Regulations

Recently, the Ministry of Higher Education of Cuba approved a regulation for the development of digital repositories, which includes basic elements for the management of research data, serving as a background and basis for the work that the project has been developing.

3-Training of teachers, researchers, librarians and managers regarding RDM.

Despite the strategic value of data in decision-making processes being recognized and considered an important raw material in the digital economy, there has been an unfortunate lag in training the right skill set to use this information profitably and comprehensively delay in training the right set of skills to use this information profitably and comprehensively. Today, not everyone needs to be a data analyst or data scientist, but we do need to be data literate, especially researchers and librarians (Martín-González, 2022).

During the first year of work, 10 members of the project (5 CUB, 3 PER and 2 BOL) received a training on RDM with specialists from 3 universities in Belgium, the latter with palpable results in this regard within the Consortium of Flemish universities. The training included exchanges with data stewards from the universities of Hasselt, Ghent, Leuven and Antwerp. Also, this training included exchanges with Data Stewards from the universities of Hasselt, Ghent, Leuven and Antwerp. As well as exchanges about the experiences in Belgium in relation to the requirements established as mandatory by the European Community and VLIR on RDM.

Research Data Management was included in the undergraduate degree of Information Sciences and in the postgraduate course as a master's program "Information and Knowledge for Development" in the UCLV, in the latter, in addition to the students,

also they included librarians from said institution. The Universidad Nacional Mayor de San Marcos also included RDM as part of the Doctoral training in the Health Sciences program.

Librarians have been included in the postgraduate courses, as libraries have been assuming an increasingly important role within the RDM (Marín-Arraiza et al., 2019; Angelozzi, 2020).

In each of the participating Latin American institutions, leading researchers were selected, who have been trained through online workshops on topics related to RDM such as:

- Development of policies for RDM
- Organization of Research Data
- Management Plans of the Research Data
- Data Storage
- Protection of Research Data
- Research Data Repositories
- Research Dataset Citation

These “Data Management Champions”, after completing their training, have joined to the project members and are transferring their knowledge to other researchers through personal mentoring and workshops.

4-Collaboration Networks

In addition to the relationships developed between the 5 member universities of the project, collaborations have been established with other institutions interested in the results of RDMS-LatAm:

- a. Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (Perú)
- b. Universidad de Ciencias Médicas de Villa Clara
- c. Dirección de Ciencia del CITMA
- d. Dirección de Ciencia del MES
- e. Instituto de documentación e información Científico Técnica
- f. Universidad de Pinar del Río “Hermanos Sainz”
- g. Universidad de Camagüey “Ignacio Agramonte”
- h. Instituto Nacional de Deporte y Recreación (INDER)
- i. Direcciones de Investigación, Ciencia y Tecnología del Sistema de Universidades de Bolivia.

Through these partnerships, the likelihood of generalizing results and increasing the impact of project outcomes is increased. Through these alliances, the chances of generalizing the results and increasing the impact of the project results are increased. The emergence of these networks facilitates the development of a more favorable environment for the implementation of open science and especially research data management, enhancing the academic and scientific exchange of data, information and knowledge.

This collaboration has allowed the participation of specialists from other institutions in RDM trainings, as well as the development of new national project proposals to further develop RDM.

In recent months, the Ministry of Science, Technology and Environment known as CITMA (the regulatory body for scientific activity in Cuba), approved the project "Research Data Management in Cuban Higher Education and Scientific Institutions". An initiative presented by the Cuban universities that are members of RDMS-LatAm and associated institutions, with the aim of continuing to strengthen the RDM in Cuba and applying the results obtained by RDMS-LatAm.

Conclusions

Research Data Management has been established as one of the pillars for the development of Open Science, sustainability, reproducibility and verifiability of research activity in general. The preservation and socialization of DIs drives the genesis of new projects and scientific collaboration.

International collaborative projects are a viable alternative for the development of open science initiatives and especially for those oriented towards research data management. They enable access to funding and the creation of collaborative networks.

RDMS-LatAm is the first initiative of its kind in Cuba, Bolivia and Peru, contributing to the development of a culture of IDM among the scientific community of these nations.

References

- Angeliki Andrikopoulou, J. R. & Geoff Walton. (2022). Research Data Management (RDM) and the Evolving Identity of Academic Libraries and Librarians: A Literature Review, *New Review of Academic Librarianship*, 28:4, 349-365, DOI: 10.1080/13614533.2021.1964549
- 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* [on line]. , vol.9, n.2 [consulted: june-01-2023], pp.91-91. http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1853-99122020000100091&lng=es&nrm=iso&tlng=es
- Campos, A.F., Días, G.A., Sousa, M.R.F. de..(2023). FAIR principles, information findability and data: towards a theoretical dialogue and a practical application. *Encontros Bibli*, 28, e88060. <https://doi.org/10.5007/1518-2924.2023.e88060>
- La Referencia. (2018). Políticas para la Ciencia Abierta y los Datos Científicos en América Latina. [consulted: june-01-2023]. <http://www.lareferencia.info/pt/recursos/ciencia-abierta-documentos-externos/77-la-referencia-politicas-para-la-ciencia-abierta-y-los-datos-cientificos-en-america-latina-es>
- Maredata (2018). Recomendaciones para la gestión de datos de investigación dirigidas a investigadores. [https:// digital.csic.es/handle/10261/173801](https://digital.csic.es/handle/10261/173801)

- Marín-Arraiza, Paloma; Puerta-Díaz, Mirelys; Vidotti, Silvana Gregorio. (2019). Gestión de datos de investigación y bibliotecas: preservando los nuevos bienes científicos. *Hipertext.net*, n.º 19, pp. 13-31, <https://doi.org/10.31009/hipertext.net.2019.i19.02>.
- Martín-González, Y; Iglesias-Rodríguez, A. (2022). Alfabetización en Datos en las bibliotecas-CRAI españolas: Análisis descriptivo y propositivo. *Revista Española de Documentación Científica*, 45 (2), e322. <https://doi.org/10.3989/redc.2022.2.1857>
- Mons, B. (2018). Data Stewardship for Open Science: Implementing FAIR Principles (1.a ed.). <https://doi.org/10.1201/9781315380711>
- UNESCO. (2021). Recomendación de la UNESCO sobre la Ciencia Abierta. UNESCO. [consulted: june-01-2023]. https://unesdoc.unesco.org/ark:/48223/pf00000379949_spa
- Vanhaverbeke, H., Claeys, J., Hanus, J., De Voecht, M., Mertens, M., Elsen, H., Vancauwenbergh, S., Amez, L., Van Gaeveren, K., (2018). VLIRWG RDM & OS, Research DataManagement en de Vlaamse universiteiten: White Paper, Oktober, 26p. <https://biblio.ugent.be/publication/8555410/file/8555416>