

Data, open data and big data: challenges and opportunities



NACIONES UNIDAS



Wouter Schallier

Chief, Hernán Santa Cruz Library

A big thank you to: Claudia Vilches & Gabriela Andaur

Content

1. Are we ready for BIG data?
2. Are we ready for (not so big) data?
3. Results of the LEARN project
(<http://www.learn-rdm.eu/>)
4. Are we ready for research data? RDM in LAC



NACIONES UNIDAS

CEPAL

Are we ready for BIG Data?



Getty Images

Conducción autónoma Law Strip

La avalancha de datos que seguiría a la generalización del coche autónomo no cabe en ningún sitio

Por Montse Hidalgo Pérez

-
- Una flota global de vehículos que generarían un giga de datos por segundo exigirá nuevas tecnologías y sistemas de gestión



NACIONES UNIDAS

CEPAL

Are we ready for BIG Data? (2)

1 autonomous car = 1 Gigabyte/sec.

2 billion cars in the world (by 2035)

average hours driven per car per year: 300-500h

assuming that only 1% of the cars are autonomous, and that each of them drives 300h per year

then these produce, per year:

21,600,000,000 Terabyte

= 21,600,000 Zettabyte

= 21,600 Yottabyte

WE'RE NOT!!!



NACIONES UNIDAS

CEPAL

Are we ready for (not so big) Data?

selection (“we keep everything”)

analysis (“oil extraction and refining”)

monetizing

metadata, linking with other data (social networks etc.),
data mining, re-use, machine actionable

centralized vs. de-centralized data storage (“edge
computing”)

(instant) access and availability



NACIONES UNIDAS

CEPAL

Are we ready for (not so big) Data? (2)

back up

long term preservation

standardization of data publication

authenticity, integrity (blockchain)

provenance

privacy, security and ethics

visualization



NACIONES UNIDAS

CEPAL

Computational archival science

An interdisciplinary field concerned with the application of computational methods and resources to large-scale records/archives processing, analysis, storage, long-term preservation, and access, with aim of improving efficiency, productivity and precision in support of appraisal, arrangement and description, preservation and access decisions, and engaging and undertaking research with archival material.

<http://dcicblog.umd.edu/cas/>



NACIONES UNIDAS

CEPAL

CEPAL and (Big) Data

<https://www.cepal.org/es/organos-subsidiarios/conferencia-estadistica-americas>

Conferencia Estadística de las Américas

Disponible en: [English](#) [Español](#) [Português](#)

Creada en el año 2000 mediante la resolución 580(XXVIII) de la CEPAL y la resolución 2000/7 del Consejo Económico y Social, contribuye al progreso de las políticas y actividades de estadística en los países de la región y promueve la cooperación internacional, regional y bilateral entre los institutos nacionales de estadística y los organismos internacionales y regionales.



The banner features the United Nations logo on the left, the CEPAL logo below it, and a stylized map of the Americas on the right. The text reads: "Seminario regional 'Potenciando el uso de los Registros Administrativos con fines Estadísticos para el seguimiento de la Agenda 2030' Conferencia Estadística de las Américas de la CEPAL".

[Antecedentes de la CEA](#)

[Antecedentes del Comité Ejecutivo](#)

[Miembros de la CEA](#)

[Grupos de trabajo de la CEA | Bienio 2014-2015](#)

[Grupos de trabajo de la CEA | Bienio 2016-2017](#)

[Grupo de coordinación estadística para la Agenda 2030 en América Latina y el Caribe](#)

[Actividades estadísticas de América Latina y el Caribe](#)

REUNIONES DE CONFERENCIAS

EVENTO | 14-16 NOV 2017 , AGUASCALIENTES, MÉXICO
Novena Reunión de la Conferencia Estadística de las Américas de la CEPAL

REUNIONES DE COMITÉS

EVENTO | 3-6 ABR 2017 , SANTIAGO, CHILE
XVI Reunión del Comité Ejecutivo de la Conferencia Estadísticas de las Américas de la CEPAL

DOCUMENTOS POR TIPO

- Documentos de trabajo
- Documento principal



NACIONES UNIDAS

CEPAL



Leaders Activating Research Networks

Líderes Activando Redes de Investigación

EU funded project under Horizon 2020
Research and Innovation Programme

24 months (1/06/2015-31/05/2017)

Grant agreement 654139

EU funding: 496,582 €

Coordinación: UCL (UK)

Otros socios: CEPAL, UVI, UB, LIBER



NACIONES UNIDAS

CEPAL

What is Research Data?

Research data, from the point of view of the institution with a responsibility for managing the data, includes:

- ***All data** which is **created by researchers** in the course of their work, and for which the institution has a curational responsibility for at least as long as the code and relevant archives/ record keeping acts require, and*
- ***Third-party data** which have originated within the institution or come from elsewhere.*

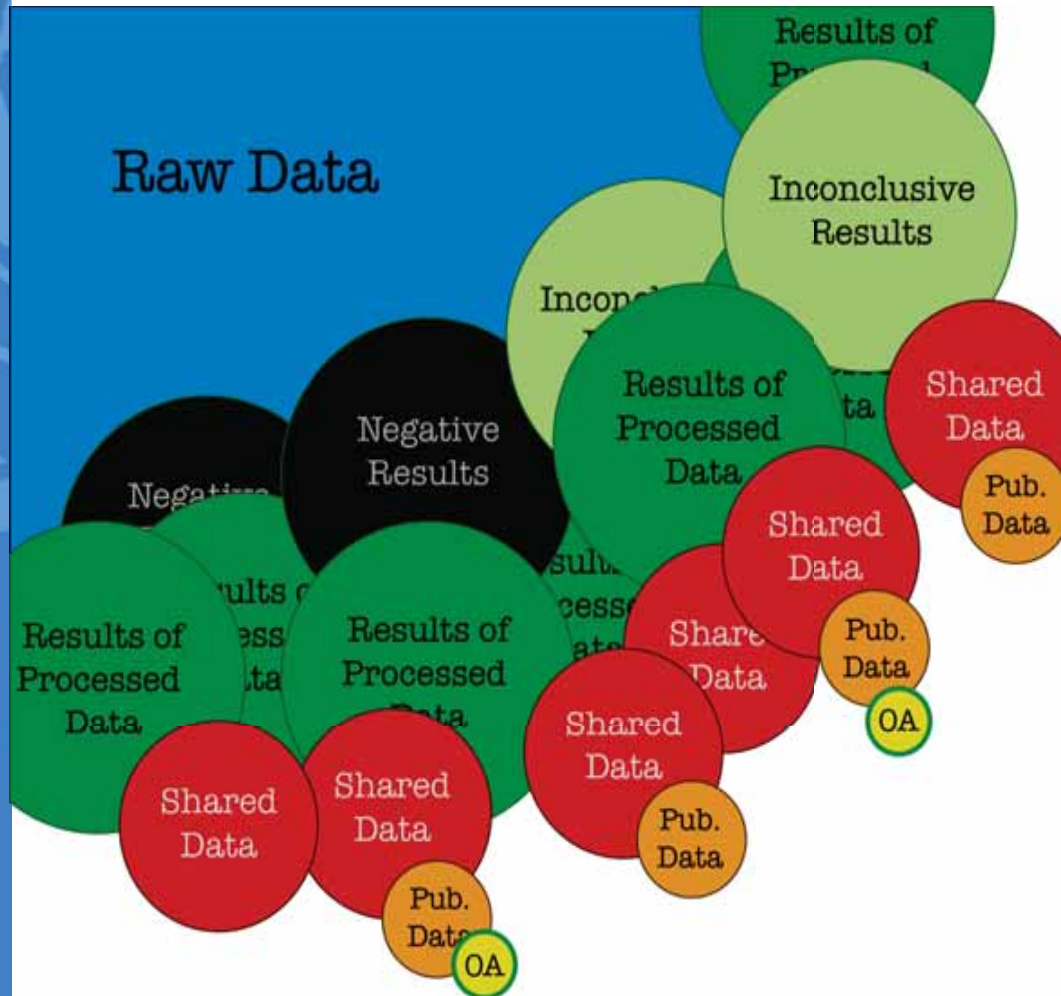
*LERU Research Data Working Group, Roadmap - Advice
Paper No. 14 – December 2014



NACIONES UNIDAS

CEPAL

Data from Research Processes: from raw data to open access published data by Raman Ganguly



Datos Primarios (Raw Data): datos directos de la medición o recolección, derivados del proceso de investigación.

Datos procesados (Processed Data): Datos derivados que han sido objeto de análisis e interpretación (limpieza o extracción de grandes set de datos). Incluye los resultados **negativos** e **inconclusos** producto del proceso de análisis.

Datos compartidos (Shared Data): datos que serán compartidos con otros

Datos publicados (Published Data): datos disponibles públicamente

Datos publicados de acceso abierto (Open Access Published Data): datos publicados bajo modalidad de acceso abierto.



NACIONES UNIDAS

CEPAL

Openness of (Research) Data

**Open whenever possible,
closed whenever needed...**

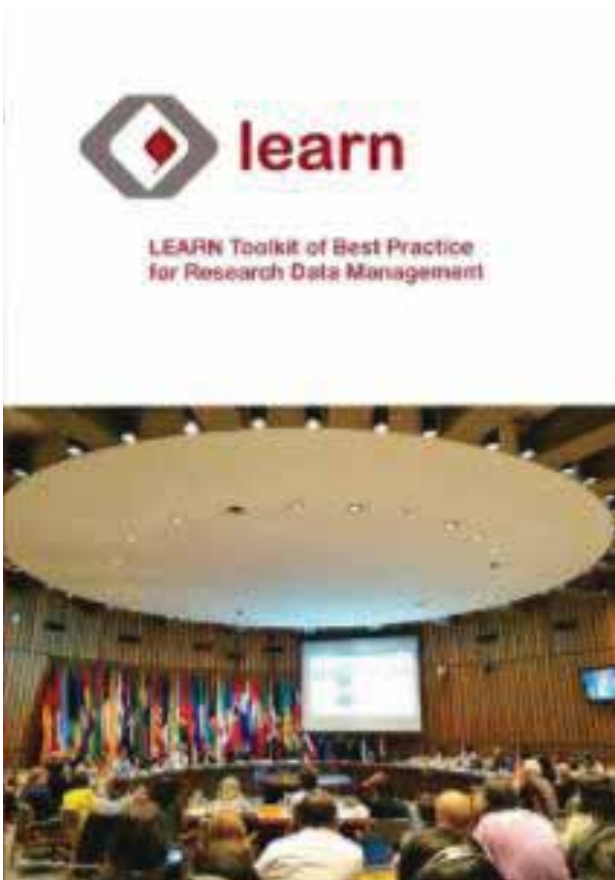


NACIONES UNIDAS

CEPAL

Outcomes of the LEARN project

<http://learn-rdm.eu/en/dissemination/>



Toolkit with 25 case studies of **good practices**

Executive summary of the LERU Roadmap in 5 languages

Model **Policy for RDM**



NACIONES UNIDAS

CEPAL

LEARN Community in LAC



Argentina, Bolivia, Brazil, Chile, Uruguay, Paraguay, Peru, Ecuador, Colombia, Venezuela, Guyana, Costa Rica, Panama, Honduras, El Salvador, Mexico, Cuba, Jamaica, Dominican Republic, Trinidad and Tobago, Barbados, Curaçao and Saint Lucia



NACIONES UNIDAS

CEPAL

National policy - Funding agency

3 instruments

Institutional
RDM policy

Road-
map

RDM
Plan

Good practices



NACIONES UNIDAS

CEPAL

Example of RDM policy



Policy on the Management of Research Data and Records

1. The University of Oxford seeks to promote the highest standards in the management of research data and records¹ as fundamental to both high quality research² and academic integrity.
2. The University recognises that accurate and retrievable research data are an essential component of any research project and necessary to verify and defend, when required, the process and outcomes of research. Research data are valuable to researchers for the duration of their research, and may well have long-term value for research, teaching and for wider exploitation for the public good, by individuals, government, business and other organisations, as a project develops and after research results have been published.
3. The University acknowledges its obligations under research funders' data-related policy statements³ and codes of practice to ensure that sound systems are in place to promote best practice, including through clear policy, guidance, supervision, training and support.
4. Researchers⁴, departments/faculties, divisions, central administrative units and service providers and, where appropriate, research sponsors and external collaborators, need to work in partnership to implement good practice and meet relevant legislative, research funder and regulatory requirements.
5. Research data and records should be:
 - a. Accurate, complete, authentic and reliable;
 - b. Identifiable, retrievable, and available when needed;
 - c. Secure and safe;
 - d. Kept in a manner that is compliant with legal obligations and, where applicable, the requirements of funding bodies and project-specific protocols approved under the University Policy on the Ethical Conduct of Research Involving Human Participants and Personal Data⁵.
 - e. Able to be made available to others in line with appropriate ethical, data sharing and open access principles.
6. Research data and records should be retained for as long as they are of continuing value to the researcher and the wider research community, and as long as specified by research funder, patent law, legislative and other regulatory requirements. The minimum retention period for research data and records is three (3) years after publication or public release of the work of the research. In many instances, researchers will resolve to retain research data and records for a longer period than the minimum requirement.

¹ Research data and records are defined as the recorded information (regardless of the form or the media in which they may exist) necessary to support or validate a research project's observations, findings or outputs.

² Research is defined as per the Frascati manual, i.e. creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications.

³ An overview of the major research funders' data policies is available at <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>. See also the BCUR Common Principles on Data Policy (2011) at

<http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

⁴ Researchers are defined as members of the University including staff and doctoral students, and those who are not members of the University but who are conducting research on University premises or using University facilities.

⁵ <http://www.admin.ox.ac.uk/curec/policystatement/>

Preámbulo (1 - 2)

Jurisdiction (3-4)

Alcance y cobertura (roles) (4)

Manejo de los datos (5)

Anexos -
Definiciones



NACIONES UNIDAS

CEPAL

Example of RDM policy (2)

7. Where research is supported by a contract with or a grant to the University that includes specific provisions regarding ownership, retention of and access to data, the provisions of that agreement will take precedence.
8. If research data and records are to be deleted or destroyed, either because the agreed period of retention has expired or for legal or ethical reasons, this should be done so in accordance with all legal, ethical, research funder and collaborator requirements and with particular concern for confidentiality and security.
9. Researchers are responsible for:
 - Managing research data and records in accordance with the principles and requirements in 5-8 above;
 - Developing and documenting clear procedures for the collection, storage, use, re-use, access and retention or destruction of the research data and records associated with their research. This shall include, where appropriate, defining protocols and responsibilities in a joint or multi-institution collaborative research project. This information should be incorporated, where appropriate, in a research data management plan;
 - Planning for the ongoing custodianship (at the University or using third-party services) of their data after the completion of the research or, in the event of their departure or retirement from the University, reaching agreement with the head of department/faculty (or his/her nominee) as to where such data will be located and how this will be stored;
 - Ensuring that any requirements in relation to research data and records management placed on their research by funding bodies or regulatory agencies or under the terms of a research contract with the University are also met.
10. The University is responsible for:
 - Providing access to services and facilities for the storage, backup, deposit and retention of research data and records that allow researchers to meet their requirements under this policy and those of the funders of their research;
 - Providing researchers with access to training, support and advice in research data and records management;
 - Providing the necessary resources to those operational units charged with the provision of these services, facilities and training.
11. The University's Research and Information Sub-Committee, a sub-committee of the University Research Committee, is responsible for guiding the development and updating of this policy.

Relationship with existing policies

12. This policy will operate in conjunction with other University policies such as:
 - Academic Integrity in Research (<http://www.admin.ox.ac.uk/personnel/cops/researchintegrity/>)
 - Policy on the ethical conduct of research involving human participants and personal data (<http://www.admin.ox.ac.uk/curec/policystatement/>)
 - Intellectual property policy (http://www.admin.ox.ac.uk/statutes/790-121.shtml#_Toc28143157)
 - Data protection policy (<http://www.admin.ox.ac.uk/dataprotection/policy/>)
 - Freedom of Information (<http://www.admin.ox.ac.uk/foi/>)
 - Information Security Policy (<http://www.it.ox.ac.uk/infosec/ispolicy/>)

Manejo de los datos
y responsabilidades
(7-8)

Roles y responsabilidades
(9-10)

Validez - Revisiones y
actualización (11)

Anexos (12)



NACIONES UNIDAS

CEPAL

Roles and responsibilities

ROL

ARCHIVE

INSTITUTION

FUNDING AGENCY

CONTENT SUPPLIER

(Proveedor de contenidos)

RESPONSIBILITY

PROTECTION

LEGAL SECURITY

SOCIAL RESPONSIBILITY

QUALITY



NACIONES UNIDAS

CEPAL

Guide of good practices

<http://biblioguias.cepal.org/gestion-de-datos-de-investigacion>



NACIONES UNIDAS

CEPAL

Bibliogúías - Bibliotecas CEPAL
Comisión Económica para América Latina y el Caribe

CEPAL Biblioteca CEPAL Repositorio Digital Pregúntenos

Biblioteca CEPAL » Bibliogúías » Gestión de datos de investigación » El ciclo de vida de los datos

Gestión de datos de investigación

Buscar en esta Guía Buscar

Inicio

Módulo 1 - Introducción a la GDI

Módulo 2 - Plan de Gestión de Datos (PGD)

2.1 Requerimientos de las agencias de financiamiento

2.2 EL PGD y el ciclo de vida de los datos

2.3 Componentes de un PGD

2.4 Herramientas para elaborar un PGD



NACIONES UNIDAS

CEPAL

Are we ready for (Research) Data?

**Test your readiness in
RDM now!**



<http://goo.gl/forms/m6PGJ34tGr>



NACIONES UNIDAS

CEPAL



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.



RDM in LAC

- Policies on institutional level are needed
- Promote a cultural change
- Large amounts of data, but doubts about accessibility (Caribbean) and usability (LA)
- Many barriers (or: too little incentives) for data sharing



NACIONES UNIDAS

CEPAL

RDM in LAC (2)

- Need for training and skills in *data science*
- Need for dialogue between different *stakeholders* (libraries, researchers, (vice)rectors, ICT, funding agencies, ministries, private sector)
- Libraries can play an important role in the promotion of RDM



NACIONES UNIDAS

CEPAL

WANTED: pilot institutions



NACIONES UNIDAS

CEPAL

Resources consulted

<http://dcicblog.umd.edu/cas/>

<https://www.informationweek.com/how-to-protect-the-big-data-archive/d/d-id/1104090>

http://dcicblog.umd.edu/cas/wp-content/uploads/sites/13/2016/05/submission_final_draft.pdf

https://interparestrust.org/assets/public/dissemination/IPT_NA08_FinalReport_1Oct2016_fordistribution_.pdf

<http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-01-2014-0008>

<http://www.emeraldinsight.com/doi/abs/10.1108/RMJ-01-2014-0010>

<https://medium.com/@puntofisso/my-report-on-best-practice-for-local-data-initiatives-313d12f83865>

https://retina.elpais.com/retina/2017/11/17/tendencias/1510920126_844738.html

<https://www.nature.com/articles/sdata201618>



NACIONES UNIDAS

CEPAL

wouter.schallier@cepal.org

www.cepal.org/biblioteca

[@bibliotecaCEPAL](#)

www.learn-rdm.eu



NACIONES UNIDAS

